The role of language and sensorimotor information in memory for concepts

Agata Dymarska

This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

May 2021

Department of Psychology

Lancaster University

Declaration

The thesis contains original work completed solely by the author under the supervision of Professor Louise Connell and Dr Briony Banks, and has not been submitted in the same form for the award of a higher degree at this institution or elsewhere.

This research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 682848).

Name: Agata Dymarska
Signature:
Date: 11/05/2021

Acknowledgements

Reading other people's thesis acknowledgments has been my favourite pastime throughout this project, so thank you to everyone who has written a thesis before for reassuring me that such an endeavour can indeed be completed.

I would like to thank Professor Louise Connell for her guidance and support throughout the PhD. With her commitment to her work, brilliant ideas, and willingness to share her expertise, she has been the best supervisor I could ask for. I have learned a lot and have been inspired a lot, and I can only hope to one day become as great a researcher and teacher as she is.

Thank you also to Dr Briony Banks for her support and advice, for helping me improve my writing, and making me feel proud of all the milestones along the way.

A great perk of the PhD was being a part of a fantastic research group: Thank you to everyone in the Embodied Cognition Lab; to Dr Dermot Lynott, who continued to introduce me to the world of science and academia, and all the postdocs and students who offered feedback and encouragement, as well as stimulating discussions.

I was lucky to have done most of my PhD in the pre-pandemic world and work in a real-life office. I'm grateful to Shal, Rachael, and Rens for making it feel homely, for supporting me in moments of doubt, and for all the tips and information we exchanged to help each other out.

Thank you to all other staff and PhD students in the Lancaster Psychology department who were there before me and who came after me, who were there for a chat and reassurance and created a welcoming atmosphere.

I cannot omit crediting the Polonium Foundation, in particular Justyna, Marta, Kasia, Olga, and Magda, and every single person who contributed to it, for giving me something useful to do while I procrastinated on the PhD, bringing endless inspiration and motivation, pushing me to expand my comfort zone, and making me feel at home, as part of a community, no matter where I was in the world.

Finally, I'm grateful for my parents and my sister's support of my academic pursuits, despite the uncertainties that it entails, and to Tabish, for always believing in me and supporting me during this journey.

Abstract

The linguistic-simulation approach to conceptual representations has been investigated for some time, but the role of language and sensorimotor information in memory for objects and words, both short- and long-term, has not been examined in detail. In the present thesis, I look at the interplay of sensorimotor and linguistic information in conceptual knowledge and examine which aspects of concepts are represented in memory tasks. I also aim to establish the role of consciously accessing conceptual information in word recognition and memory. The thesis includes three self-contained papers which show that the conceptual system relies on linguistic or sensorimotor information according to task demands.

In the paper in Chapter 4, I examined the linguistic bootstrapping hypothesis, which postulates that linguistic labels can serve as placeholders for complex sensorimotor representations. I tested the capacity of working memory for object concepts using an articulatory suppression task to block access to language. I found that working memory capacity for contextually related object concepts when relying on sensorimotor information is higher than the traditionally reported capacity of 3-4 for simple shapes or colours. Additionally, when linguistic labels are available to deputise for complex sensorimotor information, the capacity further increases by up to two object concepts.

In Chapters 5 and 6, I examined the content of conceptual information stored in long-term memory, and the role of sensorimotor simulation and consciously available information in word recognition and word memory. The studies revealed that consciously generated imagery is not reliably measured, and moreover, it does not contribute to word recognition in a consistent manner. Some of the effects of imageability found in the literature can be explained or subsumed by sensorimotor information, which is not fully available through conscious awareness. However, conscious imagery may be a useful strategy to support word memory when trying to explicitly remember words.

The thesis demonstrates that both linguistic labels and sensorimotor information contribute to memory for concepts. The way a concept is represented in different tasks varies depending on task demands. Linguistic information is used to circumvent resource capacity limits, while sensorimotor information guides behaviour by providing more detailed information about the meaning of concepts, and our previous experience with them.

Word count: 57148

Statement of Authorship

Here is a breakdown of contribution made by Agata Dymarska (the student), and Louise Connell and Briony Banks (the supervisors) to each chapter. The order in which the names appear roughly indicates the proportion of contribution in decreasing order.

Chapters 1-3 (literature review)

- Principle author: Agata Dymarska
- Comments: Louise Connell and Briony Banks

Chapter 4 (submitted paper)

- Conception and creation of study: Louise Connell, Agata Dymarska
- Data collection: Agata Dymarska supervised by Louise Connell
- Writing manuscript: Agata Dymarska
- Revision: Agata Dymarska, Louise Connell, Briony Banks

Chapter 5 (empirical paper in preparation)

- Conception and creation of study: Louise Connell, Agata Dymarska
- Data collection: NA
- Writing manuscript: Agata Dymarska
- Revision: Agata Dymarska, Louise Connell
- Comments: Briony Banks

Chapter 6 (empirical paper in preparation)

- Conception and creation of study: Louise Connell, Agata Dymarska
- Data collection: Agata Dymarska supervised by Louise Connell
- Writing manuscript: Agata Dymarska
- Revision: Agata Dymarska, Louise Connell
- Comments: Briony Banks

Chapter 7 (general discussion)

- Principle author: Agata Dymarska
- Comments: Louise Connell and Briony Banks

Supervisor’s name: Louise Connell Signature:

Supervisor’s name: Briony Banks Signature:

Date: 11/05/2021

Table of Contents

1 Introduction

Our knowledge about the world is acquired and stored using our sensory and motor experience. Knowing what a dog looks like, what it sounds like when it barks, the feeling of touching its fur or the motion when we take it for a walk, means understanding what a dog is. This information allows us to predict, plan and carry out future actions and behaviours. For example, when we see a piece of chocolate cake, we remember the delicious taste it had the last time we ate it, the feeling of guilt after eating too much, or even the happiness of sharing it with our loved ones – this is all thanks to our perceptual, motor, and affective systems simulating past experience. While a lot of the conceptual processing happens automatically (Connell & Lynott, 2016; Pecher et al., 2009; Sidhu & Pexman, 2016), some information can be reinstated consciously (Connell & Lynott, 2016; Rasmussen & Berntsen, 2014), and we are able to visualise concepts, in particular objects that we can interact with, such as tools, animals etc. This is thought to contribute to the process of understanding and remembering words and concepts, but the role of unconscious sensorimotor simulations in word recognition and memory is not as clear (Connell & Lynott, 2016; Juhasz et al., 2011; Sidhu & Pexman, 2016).

We also use language to store and communicate ideas. Seeing a cake will activate its associated label, and seeing or hearing the word "cake" will activate the sensory, motor, emotional and social experience we have with cake, as well as other, semantically or situationally related words. Taken together, this information is what allows us to process and interact with the world on a daily basis. Since the linguistic label, such as "dog", "cake" or "running" is less complex than the full collection of experiences of these concepts with all our senses, the label gets activated faster and can be used more efficiently. Because of that, language can be considered a "shortcut" which is used in place of sensorimotor information to reduce the processing load in our brain, or to facilitate communication (Connell & Lynott, 2014b).

The activation of the sensorimotor and linguistic information in order to recognise the object in front of us as cake, or to remember the ingredients while making cake, takes place in a system called working memory (Baddeley, 2000). This is a limited capacity system, which contains information we need to attend to, manipulate, or otherwise use in order to engage in a non-automatic task (Unsworth & Engle, 2007), such as following a new recipe, building a birdhouse, or delegating and carrying out household chores. Research findings, which have permeated into common knowledge, tell us that we can remember a maximum of 3-4 different shapes or colours on the screen (Vogel et al., 2001), or 3-5 chunks of items (Cowan, 2001; 2010), while others suggest the limit to be 7 (Miller, 1956). However, these studies mostly focus on individual stimuli presented without context, and do not correspond to all types of information we have to remember every day in real life. Even when chunking is considered, chunks are created based on the order in which items are presented (e.g., the first 3 digits of a phone number), rather than their meaning. On the other hand, an individual, complex conceptual representation combines information from multiple sources of perceptual and action experience, as well as connections to other objects and actions which occur in a similar context (Barsalou, 1999). For example, how many items of a shopping list can we remember? Given that a representation of, say, bread, will contain its colour, shape, smell, taste and other information about how to store it or what to put on it, holding the representation of bread in mind using sensorimotor information seems to require a lot of resources. How many objects can we represent that way? Is seven the magic number, or would it be less, given the complexity of the representations? Or more, because in real life objects are embedded in a context, like finding bread in a particular place in the store, or making up a recipe for a sandwich, which makes contextually related things (e.g., sandwich ingredients) easier to remember? And since we can simply use the label “bread” instead of relying on complex sensorimotor representations, does the use of language allow us to remember more objects? If so, how many more?

This thesis will focus on what the representation of our knowledge consists of, and how sensorimotor information and language are used to support our interaction with objects, words and situations occurring around us. It will address 3 main aims:

1. To examine whether the linguistic component of the conceptual system could bootstrap the capacity of working memory.
2. To study whether consciously available information about concepts supports word recognition.
3. To investigate whether sensorimotor information facilitates word memory.

The thesis comprises two chapters of literature review (Chapters 2 and 3), three empirical chapters containing 11 studies (Chapters 4-6), and a general discussion of the whole thesis (Chapter 7).

2 Theories of conceptual representations

Concepts are mental representations of our knowledge and experience of the world. They consist of perceptual and functional characteristics of objects, and are used to process perceptual input in an informative way to help us navigate the world. Theories on the exact nature of concepts used to focus on the computational abilities of the brain, and the abstract symbols that are manipulated to achieve an understanding of what a specific concept refers to (e.g., Collins & Loftus, 1975; Fodor, 1983/2008; Johnson-Laird, 1980). However, defining abstract representations in terms of their relations with other abstract symbols does not give them enough meaning to account for human cognition, as demonstrated by the Chinese room problem (Searle, 1980), where putting together symbols according to a rule book may result in a correct message, but does not mean that the sender understands the language. The symbols have to be immersed in perceptual and action experience. The objects and actions in the environment are thought to be transduced (transformed from signal to symbol) to the amodal symbols stored in the brain in order to be manipulated in terms of the abstract symbols. For example, perceiving a chair should somehow lead to activation of lists of its features (e.g., “has back, legs, seat”), or sentences where chair occurs, but the resulting simulation leads to a representation which is removed from the actual experience of a chair, and does not provide a satisfactory explanation of cognitive processes (Barsalou, 1999; Barsalou et al., 2003). Hence, relying on purely symbolic or purely sensory representations themselves does not provide information about meaning, beyond its reference to other symbols or icons. Harnad (1990) proposed that the solution to the grounding problem is considering perceptual information and linguistic labels as a hybrid system, where they are connected and inform one another as to the form and the meaning of the input from the environment.

2.1 Grounded theories of concepts

2.1.1 Simulation in conceptual processing

Over the past few decades, theories of conceptual representation have revisited the idea that the conceptual system is grounded in sensorimotor simulations. The idea dates back to Aristotle and early empiricists like Locke and Hume (Hume, 1739/1978; cited by Kiefer & Pulvermuller, 2012), but was re-introduced into modern cognitive science by Barsalou (Perceptual Symbol Systems; 1999) and Glenberg (1997). Simulation theories postulate that all perceptual modalities are involved in conceptual processing (Barsalou, 1999; 2008; Connell et al., 2012), that their activation is not necessarily conscious (Connell & Lynott, 2016; Pecher et al., 2009), and it occurs as a prerequisite to semantic understanding at a neural level (Garcia et al., 2019; Hauk et al., 2008). That is, as we perceive the world with our senses, the pattern of activation of the neural pathways responsible for perception across distributed areas is stored in long-term memory. Through multiple encounters of similar objects and experiences, we form representations which consist of the sensory properties of an object: how it looks, smells, tastes, or sounds; and motor properties: how it moves, whether it can be lifted, or turned around. The neural pathways responsible for perception are then partially reactivated when a similar word or its referent is encountered, and the experience is simulated in the sensorimotor systems in order to understand the concept.

Evidence for the involvement of modality-specific information in conceptual processing comes from the findings of behavioural experiments where the presence of sensorimotor information differentially affects performance. For example, the contribution of our experience of interacting with an object to mental representations can be tested in a mental rotation task. Suggate et al. (2019) asked participants to judge whether two images of an object, presented on the screen at different angles, were identical or mirror-images of each other, which required mentally rotating one of the objects to match the angle of the other. Response time was faster when the images represented a concept rated high on the extent to which a human body can interact with it, suggesting that the simulation of handling a manipulable object, which presumably involved perceptual and motor brain activity, was part of a representation of the concept. Sensorimotor representations also play a role in action planning. For example, when we hear the word “apple”, or are given one, we know what to do with it, thanks to our previous experience of seeing an apple on a tree or in a fruit basket, holding an apple in our hands, or biting into it, hearing the crunch and feeling the taste. The sensorimotor system allows us to simulate potential actions and their consequences, for example what an apple looks like inside once cut or where it is going to fall when you shake the tree. This was demonstrated in a behavioural experiment, where participants simulated which hand they would use to interact with an object based on its orientation, for example, a cup with a handle facing left or right led to performing the response action with the corresponding left or right hand (Tucker & Ellis, 1998). Similarly, seeing a climbing wall route activated representations of previous experiences in experienced climbers, and facilitated later recall of the steps involved in covering the route (Pezzulo et al., 2010). Thus, sensory and motor information about a concept contributes to our understanding, processing and interaction with it.

While a representation of, for example, “apple” consists of information from different senses, the multimodal representation does not always need to be fully engaged during processing, and the focus of the simulation can be modality-specific, for most efficient processing. Depending on the situation, a particular sensory or motor modality is recruited to represent an object’s most relevant feature. This is either because the representation is dominated by one feature (e.g., apples are mostly seen or tasted, flowers are smelled, dogs are seen and touched; Lynott, & Connell, 2009; Lynott et al., 2019), or because attention was directed to a particular property (Connell & Lynott, 2012a). Once a specific modality is activated, it is easier to continue processing subsequent stimuli within the same modality. For example, in a visual word recognition task, words associated with vision such as “handsome” are faster to recognise, while words associated with hearing, such as “rustling” are faster to read out loud (Connell & Lynott, 2014a), supporting the idea that conceptual representations and modality-specific activity share some of the same neural pathways.

Further, once sensorimotor simulation of a concept takes place in a specific modality, there is a cost to shifting to a different one, because it takes time to direct activation from one part of the brain to another. A property verification paradigm is often used to investigate this. In this paradigm, a word and its property are presented on the screen to be verified as true or false, for example “apple-shiny” or “apple-crunchy”. Reading these word pairs is thought to activate modality specific areas of the brain, according to the property, for example, vision when reading “shiny” or hearing when reading “crunchy”. This should elicit a priming effect on subsequent properties, leading to faster processing times. If modality-specific brain areas are recruited for modality-specific processing, then a visual-property verification (e.g., “apple-green”) should be faster following another visual-property verification (e.g., “apple-shiny”) than when it follows verification of a property in a different modality (e.g., “apple-crunchy”). On the other hand, if there is no systematic difference in processing times, it would suggest that information from different modalities is processed in the same, possibly amodal system. Indeed, a number of studies employing the property verification paradigm found that verification of whether a “shiny” apple is also “crunchy”, is slower because the word “crunchy” needs to be processed in the auditory system (Ambrosi et al., 2011; Pecher et al. 2004). The same principle applies to simulating different objects within and across modalities: a sentence describing something in the visual modality (e.g., “spinach is green”) is verified as true faster when preceded by another visual sentence, than when the preceding sentence described something in an auditory modality (e.g., “sound is echoing”), and the focus needs to change to another modality where the simulation will take place (Collins et al., 2011; Scerrati et al., 2015).

Simulating sensorimotor information during conceptual processing is not limited to object perception. Effector-specific motor action associated with concepts, such as holding a cup, kicking a football, or biting a sandwich, is activated in regions specific to each motor dimension (Hauk et al., 2004). As with perception, processing concepts in one motor dimension should incur a processing cost when switching to a different one. In a lexical decision task, recognising a stimulus as a real English word was faster when the target word (e.g., “typewriter”) was preceded by a prime which shared its manipulation features (e.g., “piano”), than when it was preceded by an unrelated word (e.g., “blanket”) (Myung et al., 2006). There is also some evidence that during sentence processing participants simulate the action or object being described, and therefore congruent concepts are faster to process. Participants are faster at picture verification of an eagle with outstretched wings if they had just read a sentence that describes the picture in the same orientation (e.g., “the ranger saw an eagle in the sky”; Stanfield & Zwaan, 2001; Zwaan et al., 2002). Similarly, the action of moving away from the body or looking up is faster when it does not require a separate simulation of the opposite movement. When reading about moving an object away from the body (e.g., “you gave pizza to Luke”), participants were faster to make a sensibility judgment when the response button was placed far from their body and required a reaching movement. On the contrary, when the sentence implied movement towards the body (e.g., “Anna told you the story”), participants where faster to respond when the response button was close to their body. Similarly, when the sentence referred to a concept being oriented upwards (e.g., “the car in front of you has a roof”) participants were faster to make a true/false judgement with a button positioned higher, than when the sentence referred to something oriented downwards (e.g., “the car in front of you has wheels”) (Borghi et al., 2004). The idea that performance is faster when participants are asked to perform an action which is congruent with an action sentence is called the “Action-Sentence Compatibility Effect” (ACE, Borreggine & Kaschak, 2006; Glenberg & Kaschak, 2002). However, there is plenty of evidence against this effect (based on equivalence testing and Bayesian evidence for the null hypothesis), that is, more recent studies found no difference in response time between trials where the action sentence and the response movement were congruent versus incongruent across a number of experiments (Morey et al., 2021; Papesh, 2015). It is possible that the action of reaching to press a button or push a lever, and the action of reaching to hand someone a pizza, are not represented using similar enough patterns of neurons, and therefore the assumption that simulating a reaching action and performing a reaching action must affect the response in a conceptual processing task, even when they pertain to different contexts, is incorrect. Instead, simulating action which facilitates sentence comprehension involves more than simulating "far" or "away" – it includes the agents, the object being passed, and perhaps associated information, so the direction of the movement with respect to the participant may not be the most critical part of the simulation. Additionally, performing a reaching action, which requires a choice between two locations on the keyboard or the screen, does not necessarily rely on the same neural pathways as the complex action of passing a pizza to a friend or catching a baseball. Nonetheless, the lack of a reliable effect does not eliminate the possibility of involvement of motor information in conceptual processing of action information, rather it indicates that the ACE relied on generalisations about the motor simulations and their role in conceptual processing, which is in fact more nuanced, and may require a more detailed investigation.

Further compelling evidence for modality-specific involvement in representing a concept comes from interference studies, where attentional demands are different than in modality-specific facilitation effects (Connell & Lynott, 2012b). When we are holding a cup of coffee, we cannot simultaneously use that hand to write a letter, and when we are reading a book, we cannot be watching a movie at the same time. Just as we cannot perform two different tasks within the same modality, we cannot always perform an efficient simulation in a modality which is already occupied with a perceptual or motor task. The simulation and a subsequent response will be slower, which will impair processing a concept, whether a hand response is required to a sentence like “he cut the fabric” (a hand-action sentence; Buccino et al., 2005), or a tool name needs to be recalled when the tool is facing a hand occupied with squeezing a rubber ball (Witt et al., 2010). This supports the hypothesis that knowledge about concepts is simulated in sensory and motor brain areas. Even when an action is in progress, processing action-related words requires simulation in the motor cortex, and can disrupt processing: participants who were asked to grasp an object in front of them were slower to perform the action if they were presented with an action verb 50ms or 200ms after action onset (Nazir et al., 2008). Modality-specific processing is also disrupted when a modality is unable to perform an action and simulate experience due to a brain injury or a neurogenerative disease. Damage to auditory cortex is linked to impairment in processing sound-related concepts (Trumpp et al., 2013), while Parkinson’s Disease patients, who have difficulty controlling their motor movements, show impairment in making judgments about fast action verbs (e.g., “run”; Speed et al., 2017), lexical decision and semantic judgment of action verbs (Fernandino et al., 2013); or action verb priming, in patients who do not take Levodopamine medication (Boulenger et al., 2008). Interestingly, patients with hand tremors show a decrease in the tremor intensity when reading hand-action words, which again demonstrates a connection between hand movement action (albeit unintentional), and hand movement simulation (Nisticò et al., 2019). Action word processing is also impaired in patients with apraxia, who have difficulty coordinating their movements (Buxbaum & Saffran, 2002).

In addition to behavioural evidence and the causal relationship demonstrated in neuropsychological work, neuroimaging studies show that brain areas which are activated when performing or watching an action, such as the premotor cortex, are also activated when reading or listening to action words (Hauk et al., 2004; Pulvermuller et al., 2005; Tettamanti et al., 2004) or phrases describing actions (Aziz-Zadeh et al., 2006). Similarly, brain areas associated with sensory perception are activated when processing words specific to the sensory modality, such as sound or smell (Gonzalez et al., 2006; Kiefer et al., 2008). For example, when verifying whether a pear is green or sweet, brain areas activated by visual and gustatory experience are active (Barros-Loscertales et al., 2011, Goldberg et al., 2006; Simmons et al., 2007). Action words, as well as words related to tool use, typically activate left motor cortex in right-handed individuals, but the right motor cortex in left-handed individuals (Lewis et al., 2006; Willems et al., 2010), indicating that semantic processing occurs in the brain region responsible for performing the action.

Nonetheless, some types of concepts pose a challenge to the idea that concepts are grounded in sensorimotor simulation. While it is easy to demonstrate how object or action representations rely on perceptual and action experience, it is not obvious how concepts considered to be abstract, such as “quality”, or “justice”, or “odd number” can be represented using neural pathways associated with sensorimotor experience, despite not being directly linked to a physical entity. Indeed, some studies have found differential brain activation for concrete vs abstract concepts (Binder et al., 2005; Skipper & Olson, 2014), as well as differences in lexical task performance (Balota et al., 2004; Cortese & Schock, 2013; Reilly & Desai, 2017), but the distinction appears to require a more nuanced approach (Barsalou et al., 2018; Desai et al., 2018; Pexman et al., 2007). While abstract concepts refer to things that cannot directly be touched or seen, they can be experienced through emotional (Vigliocco et al., 2009; Zdrazilova & Pexman, 2013) situational (Barsalou & Wiemer-Hastings, 2005), social (Villani et al., 2019) or internal bodily experience (Connell et al., 2018), as well as sensorimotor experience distributed across many perceptual modalities and action effectors (Connell & Lynott, 2012a; Lynott et al., 2019), and can be simulated using these types of information. This could be because the situation in which the concept was encountered involved some sensory or motor experience, which forms part of a concept itself. For example, our experiences of “justice” involve situations like a court building, a person who is a judge, a protest, or a set of written rules, all of which can be experienced through the senses, and therefore when thinking of or hearing the word “justice” we are able to understand it by simulating these experiences. Abstract words can also be grounded in emotions (Kousta et al., 2009). They are often given higher ratings of valence and arousal (Vigliocco et al., 2013), and activate brain areas associated with emotional processing (Desai et al., 2018; Vigliocco et al., 2009). While emotional experience appears to be an important factor in grounding abstract words, emotion-related words, such as “fear” also activate brain areas associated with arm and face movements in the primary motor cortex (Moseley et al., 2012). In addition, many abstract concepts can be situated in accompanying internal experience; concepts such as “hunger”, “thinking”, “pain”, or emotion-related words are highly associated with interoception, which refers to the perception of internal bodily sensations (Connell et al., 2018). Indeed, speed of processing of words defined as abstract can be predicted by perceptual strength (Connell & Lynott, 2012a), and looking at pictures of situational scenes activates representations of abstract concepts, such as “sharing” (McRae et al., 2018).

Many other types of concepts which are included under the umbrella term “abstract” have been found to be grounded in sensory and motor experience. Desai et al. (2018) found that concepts related to moral judgment (e.g., judging whether it is moral to blame a co-worker for a mistake that someone else made) as well as to theory of mind (such as learning that another person likes flowers), which require inferential reasoning, activated brain areas associated with processing emotions, as well as areas associated with processing concrete words, while words associated with “mental processes” (e.g., “thought”, “logic”) activated face motor areas (Dreyer & Pulvermuller, 2018). Numerical concepts are also a typical example of abstract concepts, but studies show they are nonetheless acquired using visuo-spatial reasoning, including finger counting (de Hevia et al., 2008; Di Luca & Pesenti, 2011), that they are processed using visualisation and finger movement (Di Luca & Pesenti, 2011; Scribner, 1984) and activate brain areas associated with body representation and finger movement (Desai et al., 2018). Similarly, abstract concepts associated with the physical world can activate modality-specific brain regions, for example the word “frequency” activates the same brain area as the action of performing a rhythmic movement (Mason & Just, 2016), and sensorimotor simulation can contribute to understanding of STEM-related concepts (Hayes & Kreamer, 2017). In summary, abstract words can indeed activate the same sensorimotor systems in the brain as concrete words (Desai et al., 2018; Hartpaintner et al., 2020; Pexman et al., 2007). There are many types of concepts, which are encoded and processed using a variety of information, but all can have a sensorimotor basis and be grounded in experience.

2.1.2 Linguistic-simulation theories

Nonetheless, sensorimotor simulations alone do not answer all the questions about how concepts are acquired, stored, and represented, and current linguistic-simulation theories of grounded representations take that into account (Barsalou et al., 2008; Connell & Lynott, 2013; 2014; Louwerse & Jeniaux, 2008; 2010). Words also carry information about sensorimotor experience associated with a concept, and can be used to activate it (Lupyan & Ward, 2013). When we see a dog coming from around the corner, we associate the experience of its action and our feeling towards it, such as approaching to pat it or running away, with the label of “dog”. Later, when we hear the word “dog” from a neighbour who just adopted one, a simulation of these types of experiences takes place offline, allowing us to fully engage with the concept even when it is absent from the environment (cf. Wilson, 2002; Connell & Lynott, 2014b). Using linguistic labels allows us to detect an object faster and more accurately (Lupyan & Ward, 2013; Ostarek & Huettig, 2017), and facilitates recognition of objects which share some of the sensorimotor features (e.g., shape) with the word’s referent (Noorman et al., 2018). To quantify the amount and type of sensorimotor experience associated with a word, they can be rated for how much they are associated with different sensory modalities (sensorimotor strength ratings:

Chedid et al., 2019; Lynott & Connell, 2009; Speed & Brysbaert, 2020; Vergallito et al., 2020), and different action effectors (Lynott et al., 2019), which modality dominates the experience (modality exclusivity ratings: Chen et al., 2019; Lynott & Connell, 2009; 2013; Morucci et al., 2019; Speed & Majid, 2017; van Dantzig et al., 2011), whether an object is experienced with the body (body-object interaction ratings; Bennett et al., 2011; Pexman et al., 2019; Sidhu et al., 2018), or with the senses (sensory experience ratings; Juhasz et al., 2011). It turns out that words with higher sensorimotor ratings or richer sensory and bodily experience are processed faster in a lexical decision task (Connell & Lynott, 2012; Juhasz et al., 2011; Lynott et al., 2019; Siakaluk et al., 2008; Sidhu et al., 2014) or mental rotation task (Suggate et al., 2019), and overlap in sensorimotor information between a category and its member (e.g., animal – dog) facilitates performance in a category production task (Banks et al., 2020). This is in line with the semantic richness effect (Buchanan et al., 2001; Pexman et al., 2008) whereby richer representations facilitate access to meaning, which supports more efficient word processing. Words higher in perceptual and action experience are also better remembered, due to a richer representation, although evidence is scarce so far (Lau et al., 2018; Sidhu & Pexman, 2016). Thus, as postulated by Harnad (1990) sensorimotor simulation supports conceptual processing in combination with linguistic information.

Language supports conceptual processing in a number of different ways. Labels perform a social function, allowing us to operate in a social context (Borghi & Cimatti, 2009; Borghi & Binkofski, 2014; Borghi et al., 2018). For example, when an encountered object or experience is not typical or familiar, language helps us communicate how it should be categorised, which is beneficial for the broader understanding of concepts within a society. Similarly, labels facilitate creating a conceptual category which encompasses a number of different experiences, such as “running on the grass, exiting prison, and taking a decision without the influence of others” which can all be categorised as “freedom” (Borghi & Binkofski, 2014, p. 29). Labels also facilitate recognition of a concept, due to their unique, easily reproducible phonological forms, as well as the property of being a stable category referent. When a word acts as a cue to activate a conceptual representation (e.g., "cat"), participants are faster to respond to a picture of a cat, than when they hear a non-verbal cue, such as meowing (Lupyan & Thompson-Schill, 2012). Words are also involved in forming representations of abstract concepts. For example, acquiring linguistic representations of numbers bootstraps children's understanding of the number systems by creating a structure which can be filled through experiencing quantities of objects (Carey, 2004).

The idea of linguistic labels acting as a "frame" to facilitate conceptual processing is demonstrated when comparing speakers of different languages. According to the Sapir-Whorf hypothesis of linguistic relativity, our perception of the world and the categories within it are influenced by the language we speak (Whorf, 1956). It is important to note, that the theory proposed by Whorf (1956) was general and based on anecdotal evidence and personal observations of different languages (Latkowska, 2015; Pederson, 2010), and has since evolved through decades of more rigorous research to focus on specific cognitive mechanisms. A popular example is colour perception, studied widely across languages: speakers of English tend to perceive a larger distance between green and blue colours, unlike speakers of languages where a linguistic distinction between the two does not exist, such as Berinmo from Papua New Guinea (Roberson et al., 2000) or Tarahumara in Mexico (Kay & Kempton, 1984). Similarly, speakers of Russian are faster to discriminate different shades of blue, which have individual labels, than speakers of English where both colours belong to the same category (Winawer et al., 2007). Other examples include enhanced pitch perception in speakers of tonal languages (Peng et al., 2013) or enhanced odour discrimination in speakers of a richer odour-related language (Majid, 2020). Language also allows for different conceptualisations of time and space. English speakers typically represent time in a horizontal line (left to right) while representing it on a vertical axis (starting from the past being up) is typical, for example, for speakers of Mandarin Chinese (Boroditsky et al., 2011). Similarly, children as young as 2 years old vary in their conceptualisation of space, depending on whether they speak English or Korean (Choi & Bowerman, 1991).

This is not to say that people who do not speak a particular language are not capable of distinguishing certain categories or perceiving some concepts. Indeed, exposure to novel linguistic structures later in life can easily affect the way that people conceptualise, for example, time. Typically, in a task where labelling objects as "earlier" or "later" corresponded to either top-bottom (congruent in Mandarin = faster), or bottom-top (incongruent in Mandarin = slower) response button locations, respectively, English speakers do not show a difference in performance. However, the congruency effect emerged when English speakers learned novel metaphors about time which facilitated vertical representations (Hendricks & Boroditsky, 2017). Similarly, children's use of multidigit numbers can be influenced by salience of the cues in a number task, rather than being a fixed function of the way numbers are labelled and acquired in a particular language (Towse & Saxton, 1997). It is possible that linguistic labels may enhance processing of certain categories, and reflect the kinds of concepts that are important to speakers of each language (such as colour perception in hunter-gatherer societies or pitch perception in speakers of tonal languages). From the linguistic-simulation perspective, critical to this thesis, the Sapir-Whorf hypothesis suggests that in a sense, language serves as an additional form of experience, somewhat like perception and action, its structure leading to certain characteristics being more salient in our perception of a concept than others. Nonetheless, the two ideas should not be conflated – the Whorfian approach originally suggested that language shapes our thoughts, and therefore guides cognitive processes, but it did not account for sensorimotor grounding of concepts. On the other hand, according to the linguistic-simulation theories, the two sources of information are inter-related, but form a dynamic system where they can support conceptual processing separately or together, depending on task demands.

2.1.3 Language as a shortcut

In addition to activating sensorimotor simulations, language encodes knowledge directly through statistical information about word distribution. Since communication is a big part of our everyday experience, and language use is crucial to describe situations from the past or speculate about the future, language constitutes a rough reflection of those situations. By gathering information about words which occur in similar contexts or in close proximity to each other, and storing these word associations in long-term memory, our linguistic system is able to inform some of our cognitive processes without the need for a sensorimotor simulation. This is demonstrated by strategies employed to respond to a property verification task (Solomon & Barsalou, 2004). When participants were presented with linguistically associated true pairs (e.g., watermelon-seed) and linguistically unassociated and false pairs (e.g., airplane – cake), they were fast to make the TRUE/FALSE decision on whether one was a property of the other. On the other hand, when the task included associated but false pairs (e.g., monkey-banana), relying on linguistic co-occurrence information was not sufficient, and responses were slower – instead, a perceptual simulation of the conceptual representation was necessary to verify whether the pair represented a true property or not. Linguistic co-occurrence can also predict characteristics of a concept (Durda et al., 2009), predict affective ratings of words (Bestgen & Vincze, 2012; Recchia & Louwerse, 2015), or determine to which sensory modality a word belongs (Louwerse & Connell, 2011): we can understand that the word “blinking” refers to a visual modality, and the word “sweet” refers to gustatory modality, due to their association with words like “light” or “vision”, and “taste” or “chocolate”, respectively. Additionally, while sensorimotor information plays a big role in learning about the world, some things cannot be learned via sensorimotor experience, either because the situation is not available (for example, in a tropical region one cannot experience the concept of “snowman” through building one on a winter’s day), because the modality is unable to provide experience (e.g., blind individuals cannot perceive colours), or simply because knowing things only after experiencing them would not be practical. Linguistic co-occurrence also allows us to understand novel conceptual combinations, such as “elephant complaint” (Connell & Lynott, 2013) and spatial relations in the world (Zwaan & Yaxley, 2003; Goodhew et al., 2014; Louwerse, 2008; Willits et al., 2015). For example, words which co-occur with spatial location (e.g., “dream” – “up”) prime people to look towards the location (Goodhew et al., 2014). Linguistic distributional information can guide behaviour, just like sensorimotor information, but in reality, they work in conjunction with each other, one activating the other according to task or resource demands.

Linguistic associations create networks of information that we can use without having to activate perceptual simulations every time, but are not as detailed as the sensorimotor system. For example, visual- and haptic- related words tend to co-occur, as do gustatory and olfactory words (Louwerse & Connell, 2011), leading to a less nuanced but faster response to, for example, a property verification task. The cost to this is that sometimes the linguistic system can inform the response incorrectly. When asked to answer questions such as “how many animals did Moses bring on the ark?” or “where were the plane crash survivors buried?”, people often miss the inaccuracies in the question (i.e., Moses did not build the ark, survivors are not buried), and try to come up with genuine answers (Erickson & Mattson, 1981; Barton & Sanford, 1993). This is likely because of the linguistic associations between Moses and Noah, or survivors and victims, which co-occur in similar contexts, therefore leading participants to create a rough representation of, say, “the person in the Bible” which does not conflict with “the person who built the ark”, much like a rough representation of “the people who are in a plane crash” does not conflict with “buried”. Hence, people respond as normal, to the best of their knowledge, about the ark or the plane crash victims.

At the same time, the benefit of relying on the linguistic system is speed. The peak of linguistic activation occurs earlier than the peak of the simulation, seen both in EEG recordings (Louwerse & Hutchinson, 2012) and behavioural studies where responses informed by linguistic co-occurrence are faster (Barsalou et al., 2008; Connell & Lynott, 2014b; Louwerse & Jeuniaux, 2008). As suggested by Barsalou et al. (2008) in Language and Situated Systems Theory (LASS), the linguistic system also informs the decision on whether the simulation is necessary (as in the cross-modality examples in Solomon & Barsalou, 2004 or Louwerse & Connell, 2011), or not – for example, when two words almost never co-occur together, participants can make a quick judgment that they cannot plausibly form a conceptual combination before they need to engage the sensorimotor system to try to simulate it (Connell & Lynott, 2013). Thus, linguistic representations only have to be "good-enough" to provide a rapid response or signpost the need for further activation. The two types of information: linguistic and embodied, are thus interdependent and reinforcing, and used depending on task and resource demands (Louwerse, 2011; Wingfield & Connell 2020; Willits et al., 2015).

An important role of the linguistic system within the linguistic-simulation framework is that it is computationally cheaper. While the full concept might consist of a large amount of sensorimotor information associated with perceptual and action experience, a linguistic label is a stable, concise piece of information (e.g., "dog" always refers to a dog, whether it is big or small, brown or black, loud, soft, fast, etc.). The simple linguistic label demarcates a specific concept without the need to activate the full sensorimotor representation. This allows us to use language as a bootstrapping mechanism (Connell & Lynott, 2014b). A label can represent a concept when time and resource constraints make it difficult or impractical to activate a full sensorimotor representation, as well as when ongoing attentional and perceptual demands require that the modality-specific resources are preserved for another task (Connell & Lynott, 2014b). For example, labels help discrimination of and memory for odours (de Wijk & Cain, 1994; Cornell

Kärnekull et al., 2015; Rabin, 1988), which are otherwise difficult to simulate. Miller et al. (2018) found that participants were better at discriminating minimally different tactile-patterned stimuli when they were implicitly taught labels for them by having the tactile patterns and pseudowords presented simultaneously. Similarly, participants categorised novel stimuli as an approachable alien and a hostile alien more accurately when they had learned what features are typical for the approachable and hostile type of aliens (labelled with two pseudowords; Lupyan et al., 2007). A possible mechanism behind this is that because the novel stimuli involved a large number of sensorimotor characteristics, many of them overlapping, the task was difficult when participants had to rely on the visual features (e.g., "rounder base" or "smoother head"), as they would have to consider the object in front of them and the referent objects in their mind, and compare their features while holding all three representations in working memory. On the other hand, when participants were able to label the alien based on its characteristics, all they had to do was compare the label of the object in front of them and the two labels in their memory and decide whether or not it was the approachable type or not. Additionally, there was less room for confusion regarding which category contained which characteristic, as each characteristic was assigned to one of the labels, making the categories more easily distinguishable. Linguistic labels thus enable more efficient processing in a task which requires excessive sensorimotor resources.

2.2 Other theories of concepts

2.2.1 Non-embodied theories

Theories of how concepts are represented in memory range from amodal, where concepts are represented by abstract symbols, to different degrees of embodiment, where sensorimotor information may play a secondary role in semantic activation (Meteyard et al., 2012). In traditionally amodal theories, concepts are defined according to their associated characteristics, and language serves as a symbol to represent them. According to the semantic network model (Collins & Quillian, 1969; Simoni, 1979; Quillian, 1967; Quillian, 1969), names of concepts, such as “bird”, are stored in individual nodes, and connected with their properties or with other related concepts using superordinate, subordinate or property relations (e.g., “bird”-“eagle”; “bird”-“flies”), with the full network of connections representing the word’s meaning. This allows us to make inferences, (e.g., “eagle”-“flies”), rather than store every property of every word concept separately. While some information has to be linked to a specific concept (e.g., “penguin”-“cannot fly”), most of the time the inference can be made by searching along the multiple levels of connections, which means a large amount of knowledge can be stored in a limited capacity system. However, the verification of “eagle”-“can fly” takes less time than the verification of “eagle”-“has skin”, which requires moving through a larger number of nodes (“eagle”-“bird”-“animal”-“has skin”), to arrive at an answer (Collins & Quillian, 1969). Searching along the pathways of connections can also inform property verification if the property is not true (e.g., “cola-cola is blue” is false, because its stored property is “brown”). The model also accounts for resolving semantic ambiguity – when talking about a bank (e.g., “he went to the bank to deposit money”), we know that it is a financial institution, not the side of the river, because the concept of “money” is linked closely to the “bank” as a financial institution. However, the semantic network models focus on (mostly) explicitly acquired knowledge about hierarchies of connections between objects and their properties, and do not take into account many other ways that knowledge can be acquired and represented.

The mechanism and timescale of retrieving semantic information in tasks such as property verification can be explained by linguistic-sensorimotor theories of conceptual representations, which suggest that knowledge about concepts relies on sensorimotor information from the environment, as well as information about patterns in language. For example, property verification performance can be driven by linguistic co-occurrence only, as demonstrated in a study where linguistic co-occurrence information was sufficient to quickly verify two frequently co-occurring words as an item-property pair (e.g., “watermelon – seed”; Solomon & Barsalou, 2004; see Section 2.1.3). Additionally, when linguistic co-occurrence information was not sufficient, sensorimotor information was used to inform response. Because “eagle” and “flies” co-occur frequently, this information is activated quickly and allows for an efficient response. Since “eagle” and “skin” do not co-occur frequently, they require participants to perform a simulation of eagle, and then make a decision on whether or not the property is correct, which takes more time. These findings support the idea that linguistic and sensorimotor information guide conceptual processing.

More complex types of models which describe the semantic system are connectionist models (Rumelhart et al., 1986; McClelland & Rogers, 2003), where every instance of an encountered concept is represented as a pattern of activation over a system of nodes, corresponding to features. When a concept is encountered (e.g., when we encounter a cat), an input layer of nodes which correspond to the perceived features, such as four legs, a tail, or fur, is activated, and spreads activation to the next layer. The network then outputs a layer of nodes which informs a classification decision. If the decision is correct, a cat is recognised, and the pattern of activation is reinforced to represent the same concept in the future. If a decision is incorrect, for example the input is classified incorrectly as “dog”, the weights of each node are adjusted based on feedback from the environment, such as being told that dogs bark or that cats have whiskers. In contrast to the semantic network model, in connectionist models concepts are represented over a number of distributed nodes. However, many features still appear to be amodal representations with no clear sensorimotor correspondence, for example “is domesticated” or “drinks milk” for a cat. Alternatively, very specific features are abstracted from sensorimotor experience: “has four legs” “has a tail” refers to visual experience, but many other sensorimotor features such as the softness of the fur, the sound of a meow and so on, are usually not accounted for in the model. In fact, the sensorimotor characteristics might actually be more useful, because they correspond to a full spectrum of perceptual and motor experience, and they help ground the symbolic representations in corresponding real-life experience (Andrews et al., 2014). Banks et al. (2020) proposed a distributional model which used these characteristics, and found that it performed better when both linguistic *and* sensorimotor representations were included, performing as well as a typical human on category production. Similarly, a linguistic distributional- and sensorimotor-driven model produced concepts that were closer to human conceptual representations than a purely distributional model (Steyvers, 2010; Andrews et al., 2009). Although this approach has not been taken by connectionist models, they could be a good characterisation of how concepts are represented if they were grounded by combining linguistic distributional and sensorimotor information, for example if each node represented a perceptual property.

2.2.2 Weak and secondary embodiment

Many theories acknowledge the contribution of perceptual and action information to conceptual representation in a secondary or partial role. For example, it has been suggested that activation of sensorimotor brain areas is epiphenomenal, and is an aftereffect, rather than a prerequisite, of semantic processing (Mahon & Caramazza, 2008; 2009; 2011). More specifically, an abstract representation of, for example, a hammer, can activate motor neurons if an action is required, such as when we need to grab a hammer in order to use it. Alternatively, sensorimotor simulation could be necessary only to solve specific tasks, which require, for example, visuo-spatial reasoning, such as finding our way around a location or getting from one place to another by following instructions about landmarks, and thus simulation is only activated as an offloading mechanism to support the amodal system (Machery, 2016). However, a semantic system which encodes information in sensory modalities, then abstracts an amodal representation, but then retranslates it to modal concepts for processing does not seem like the most parsimonious. Further, as discussed in Section 2.1, sensorimotor brain areas are involved in semantic processing when no action is required (Kiefer et al., 2008). If the activation was present only when an action needs to be performed, then there would be no reason to see activation in modality-specific cortices while reading action-related words (Hauk et al., 2004). The timescale of sensorimotor activation also implicates sensorimotor information in word comprehension: activation occurs between 116ms (Hoenig et al., 2008) and 150ms after word presentation (Kiefer et al., 2008) for the visual and auditory cortex, while in the motor cortex activation is much stronger for action verbs compared to other types of words, occurring between 165-189ms (Garcia et al., 2019), which is before full semantic processing is complete. Additionally, encoding action verbs interferes with performing a reaching movement within the first 200ms of the word onset (Boulenger et al., 2006), which suggests that motor pathways involved in performing the action are also needed to process the action word. However, processing the action word before the action is required actually facilitates its execution (Boulenger et al., 2006; Nazir et al., 2008). That is, once an action verb is presented, it activates the neuronal pathways responsible for action information, which can then be more easily used to perform the reaching movement 500ms later. On the other hand, when a word does not involve action information, the reaching movement is performed slower, because the action-related activation needs to start from scratch. Hence, evidence from the time course of action word processing and action movements suggests that language and motor action share overlapping neuronal representations. Nonetheless, this does not unequivocally prove that action-related activation is necessary for language comprehension. That is, participants did not demonstrate their semantic knowledge about the word while performing the action. It is, however, apparent that activation of sensory and motor information is not merely an aftereffect of language processing, or an artifact of having to perform the reaching movement.

Similar evidence against the epiphenomenal view can be taken from behavioural findings. When a modality is occupied with another task or unavailable to perform the simulation, then conceptual processing may be impaired (Bidet-Ildei et al., 2017; Song et al., 2019). Alternatively, an activation of a specific modality will facilitate its use in a later task, if it relies on similar sensorimotor resources and does not conflict with attentional demands (cf. Connell & Lynott, 2012b). For example, Connell et al. (2012) asked participants to decide which out of two object names presented on the screen referred to a larger object. During the task, participants were exposed to tactile vibration on their hands or feet, or were asked to hold a beach ball between their hands or knees, for proprioceptive stimulation. Size judgment of manipulable objects was faster when participants received tactile or proprioceptive stimulation of their hands, but not legs. In other words, when neurons associated with holding or touching small objects were activated, subsequent simulation of the presented object in a size judgment was more efficient. If the activation of sensorimotor information was epiphenomenal, the size judgment would have been made at the same speed, regardless of tactile stimulation. Instead, the simulation of tactile experience was a functional part of making size judgment, and thus could be considered an integral part of representing the concept characteristics.

An alternative view of semantic memory which only considers embodiment to be secondary to conceptual representations is the hub approach. For example, the convergence zone theory proposes that during perception "sensory channels engage geographically separate sensory regions of the brain" (Damasio; 1989a, p. 123), which are then bound together to create semantic representations. The combinatorial code of the sensory information is stored in convergence zones that bind them into item representations, and those are further combined into higher-order convergence zones which bind them into more complex information such as event representations. The convergence zones do not constitute a representation, but merely control other parts of the brain (Lakoff, 1993), and are "uninformed as to the content of the representation" (Damasio, 1989b, p. 46), containing the combinatorial code of the lower level perceptual information.

Convergence zones are thought to develop based on which combination of sensory and motor neurons is activated at the same time, for example when perceiving an object, in line with the idea of Hebbian learning and the Perceptual Symbols System (Barsalou, 1999; Matheson & Barsalou, 2018; Simmons & Barsalou, 2003). This means that the lower-level information is modality-specific. Damasio et al. (2004) found that impairment in naming musical instruments was associated with lesions in the auditory cortex, and that naming tools activated brain areas associated with manual actions. Indeed, current theories of embodiment support the idea that conceptual representations rely on higher order convergence zones which combine the input from multiple modalities (Barsalou, 1999; Meteyard et al., 2012). Perception and conceptual processing do not rely on exactly the same neural activity, as knowledge about concepts constitutes knowledge about how to combine, separate, classify, ignore or attend to the rich perceptual input the world provides, and this must be informed through a sort of top-down mechanism. However, the idea of the convergence zones being an amodal "control centre", which has been removed from modality-specific information, is at odds with findings from studies on conceptual processing. For example, directing attention to a particular modality disrupts processing words associated with that modality, even when they express very different concepts (Connell & Lynott, 2014a), or primes responses to words within that modality (Pecher et al, 2004). If the activated word representation triggered the higher-order amodal combinatorial code, this should not interfere with other combinatorial codes of other concepts, regardless of modality. Thus, the current simulation theories appear to build on the convergence zone theory, as both state that information from sensory and motor association cortices are combined to represent a multimodal entity, but the exact nature of the convergence zones may be disputed.

Similar evidence pointing to the existence of a semantic hub which integrates multimodal information comes from three main areas of research. First, semantic dementia patients who suffer progressive damage to the anterior temporal lobes (ATLs), and lose semantic knowledge across all modalities (Bozeat et al., 2000; Hodges et al., 1995; Rogers et al., 2004), such that they are unable to name, recognise, draw or use common objects. Further evidence comes from activation of anterior temporal lobes found during conceptual processing in fMRI studies (Chiou et al., 2018; Visser & Lambon Ralph, 2011). Similarly, applying Transcranial Magnetic Stimulation (rTMS) to the temporal lobes disrupted performance in a semantic association task, where participants had to choose which word pairs were more closely associated (e.g., "rabbit – shotgun" vs "rabbit – pistol"; Pobric et al.; 2010). Findings from the research described above led to the development of the hub-and-spokes theory, whereby sensory and motor information is processed in modality-specific brain areas, but converges in ATLs where the representation is stored. However, the pattern of deficit in semantic dementia patients suggests that the nature of representations is more nuanced. In a picture naming task, Reilly et al. (2012) found that patients showed a pattern of functional errors, that is, they could not name some objects, but could explain what they were used for (e.g. knife – you cut with it). Further, Merck et al. (2012) found that patients with semantic dementia showed no priming effect in a lexical decision task when the prime and target were associated contextually (e.g. bed and pillow), contrary to healthy participants. While these findings implicate the ATL in semantic knowledge, they do not necessarily mean that it serves as a central hub for multimodal representations. Instead, Damasio et al. (2004) propose that the temporal poles may be associated with lexical retrieval, and thus their damage causes difficulty with manipulation of the linguistic part of the representation – that is, the label. In light of linguistic-simulation theories, this could mean that the linguistic distributional information about word co-occurrence are stored and manipulated in the temporal lobes and that is where they connect to sensory and motor information. The ATL impairment then makes it difficult to both match the sensorimotor representation to the right label, and to use linguistic association information to inform task response, instead of disrupting the entire conceptual representation.

The distinction between the sensorimotor and linguistic systems is captured in dual-coding theory (Paivio, 1971), where semantic representations are not stored in a centralised hub, but rely on two different types of information which are interconnected. According to this approach, conceptual representations comprise of a linguistic code, such as a label, and an imagistic code which can be imagined and represented by our senses. Concepts which are more directly connected with the sensory code are more concrete, while concepts which are more abstract are represented using verbal associations (e.g. the concept of "justice" may be associated with "fairness"), and are therefore indirectly linked to sensory experience (e.g. an experience of "fairness"). As a result, Paivio suggested that processing of concrete, highly imageable concepts is faster and more efficient than processing abstract concepts.

Much evidence for dual coding theory rests on experimental work using imageability, the psychological construct that measures how easy it is to generate a mental image for a given word. Imageability is typically rated on a 7-point Likert scale (Paivio et al., 1968), where a low-imageability word like "aptitude" is difficult to image whereas a high-imageability word like "lake" is easy to image. High-imageability words are concrete concepts with direct connections to the imagistic code, whereas low-imageability words are abstract concepts that have little if any direct connection to the imagistic code. Since concrete concepts are encoded with both the image and its label, this additive feature makes them easier to process. Imageability should facilitate conceptual processing, since high imageability words are associated with more information (dual modality), the association is stronger (because it is directly mapped onto the multisensory characteristics of a real-world object), and the image is more vivid than the label because it is embedded in sensory, rather than phonological or orthographic processing. Indeed, words rated higher in imageability are recognised faster in lexical decision (Balota et al., 2004; Cortese & Schock, 2013; Reilly & Desai, 2017) and recalled more accurately in a memory task (Rubin & Friendly, 1986; Lau et al., 2018). In a way, this principle makes the theory resemble the linguistic-sensorimotor approach, as words higher in sensorimotor strength were faster to recognise in a lexical decision task (Connell & Lynott, 2012a). In fact, sensorimotor strength has been found to actually outperform imageability as a predictor of lexical decision (Connell and Lynott; 2012a). Additionally, evidence suggests that perceptual characteristics of objects and conscious imagery do not represent the same information. Perceptual imagery abilities do not predict modality-switch performance, suggesting that generating conscious imagery does not involve the same processes as simulating objects and situations, which happen unconsciously and effortlessly (Pecher et al., 2009). When rating imageability, participants tend to rely heavily on vision but ignore touch, and misinterpret sound, taste, and sometimes smell information (Connell & Lynott, 2012a). More generally, participants find it difficult to consciously consider the full range of perceptual modalities when asked to provide a single rating (Connell & Lynott, 2016). As a result, ratings of imageability neglect and distort sensorimotor information about a concept, such as auditory or olfactory imagery, which is not the focus of the task. Imageability ratings skew the perceptual information underlying a concept because they require the simultaneous consideration of multiple perceptual dimensions in a way that participants find difficult to manage. Thus, imageability does not reflect the full sensorimotor experience which contributes to a representation of a concept.

Dual-coding theory is somewhat similar to the linguistic-simulation approach in that it postulates that concepts are represented using both sensory (and motor) experience and linguistic associations. However, it distinguishes between concrete and abstract concepts based on their characteristics and meaning, while the linguistic-simulation approach shows that concepts can be represented using different types of information (i.e., sensorimotor or linguistic), depending on task demands. Concepts which are traditionally considered “concrete”, or which are high in imagery, may also rely on linguistic associations when information about their linguistic co-occurrence suffices to guide a response (Solomon & Barsalou, 2004). On the other hand, concepts considered “abstract”, which are rated low in imagery, can be grounded in perceptual, action and emotional information (and more) through our social or bodily experiences (Connell et al., 2018). Therefore, dual-coding theory does not accurately describe the nature and characteristics of conceptual representations.

2.3 Summary

While each theory of semantic memory has put forth an explanation of some aspect of how knowledge is stored and used, the linguistic-simulation theories bring together the idea that information from language and from our senses both contribute to representation of a concept, which consists of our situational, social, emotional, bodily and linguistic experience, rather than amodal symbols stored in a centralised system which can be consciously accessed. The linguistic-simulation approach can also explain a number of previously reported processes. Modality-specific information is recruited to perform conceptual tasks, but can also come together to form multimodal representations, or it can be bypassed altogether when the linguistic system is engaged and can perform a task on its own. It is clear that conceptual representations are complex and different aspects of the representation can be activated depending on incoming information, our attention, or task demands. The linguistic-simulation perspective does not claim that nothing but sensorimotor simulation and sensory brain regions are relevant in conceptual processing. While grounding in sensorimotor experience is clearly useful, there is no reason to deny that complex processing also requires more complex structures, where information from different modalities converges (Damasio, 1989; Gainotti, 2011; Meteyard et al., 2012). Additionally, language, which in itself is symbolic, can contribute to complex representation (Connell, 2018), or can serve as a placeholder for faster and less demanding processing (Connell & Lynott, 2014b). However, questions about the part of the simulation that gets activated remain unanswered. For example, dual-coding theory suggested that consciously available aspects of the representation’s content are used to facilitate performance in a conceptual task. However, the perceptual information about a concept which is automatically recruited to perform a semantic task does not necessarily correspond to consciously generated imagery (Pecher et al., 2009) used to test dual-coding predictions, and therefore ease of generating imagery does not represent what information gets activated during conceptual processing. Additionally, when multiple representations are retrieved and manipulated to complete a task, it is not clear what the limit is on the number of representations that may be activated at one time. The next chapter will address this question.

3 Working memory for concepts

The system responsible for temporarily holding information in mind is referred to as short term or working memory (Atkinson & Shiffrin, 1971; Baddeley, 1992; Baddeley & Hitch, 1994). Research in working memory has not yet addressed the linguistic-sensorimotor nature of conceptual processes. In this chapter, I will outline some of the theories of working memory, and how much information it stores at one time, as well as address the challenges that the linguistic-simulation account of concepts poses for the current state of research on working memory.

3.1 Working memory

One of our most important cognitive functions is the ability to keep and manipulate information in mind in an on-line manner, which is referred to as short term or working memory. A number of models attempting to define the process have been developed (see e.g., Adams et al., 2018 for discussion). Some present a unitary system which encompasses whatever information is activated from long-term memory or via perception (Cowan, 1988) and where the dynamic processing of information takes place (Newell, 1990). Others propose separate modules for different types of information, such as the Interacting Cognitive Subsystems model of cognitive processing (Barnard, 1999) where modality specific information (e.g., auditory, visual, bodily state, limb movement) is encoded via specific subsystems which come together to represent conceptual knowledge. This model, and the modular approach in general, resembles the sensorimotor approach to cognition discussed in Chapter 2.1. However, in order to estimate how many complex representations can be held in mind at one time, we turn to a model which has been the most extensively used to address the question of capacity: the Working Memory model (WM, see Figure 1) proposed by Baddeley and Hitch (1974; 1994; Baddeley, 2000). WM also consists of subcomponents responsible for processing information in different modalities, where information is activated temporarily, and has to compete for limited “space”: the phonological loop receives and maintains auditory and verbal information, such as speech or written words and nameable pictures, in short-term memory, particularly through subvocalization. The visuospatial sketchpad is responsible for maintaining visual information, for example images, or the combination of visual stimuli and their spatial location, such as a spot of light moving on a screen (Baddeley, 2007). Input from the two modalities is combined in a central executive system. Its exact role has been debated (May, 2001; Towse & Houston-Price, 2001a) since it is easy to use its functions as a way around anything that cannot be attributed to other systems. It is generally considered to coordinate the input and processing in the slave systems (Beddeley & Hitch, 1974; Baddeley & Logie, 1999) and performance of concurrent tasks (Baddeley et al., 1986; D'Esposito et al., 1995). The ability to pay attention to stimuli plays a role in the capacity to hold information in mind (Kane et al., 2001; Unsworth & Robison, 2015), but the central executive is thought not to have memory capacity in itself (Baddeley, 1996; Morris & Jones, 1990). While the need for some kind of attentional mechanism was acknowledged in short-term memory research even before the working memory model (Atkinson & Shiffrin, 1968), it is likely that the complexity of the central executive system means that its multiplt functions can actually be assigned to multiple subsystems within (Baddeley, 2002). These could be responsible for voluntary control of processing versus involuntary orienting of attention (Cowan, 1988) or for different types of executive functions (Parkin, 1998). For the purpose of this thesis, I focus on the role of executive control (which may include subsystems of a larger mechanism) in access to conceptual representations independently of modality-specific processing. The visuo-spatial sketchpad and the phonological loop work independently, that is, any modality-specific interference affects processing in only that domain: interference of the phonological loop should therefore prevent verbal information from being encoded, or lead to earlier decay, but does not impair visual processing, while visual interference does not affect phonological processing (Allen et al., 1978; Baddeley, 2012). This has frequently been tested in a dual-task paradigm. When participants were asked to attend to and remember two types of stimuli from the same modality, for example, two colours of a square, their time and accuracy of response in a later recognition memory task was impaired, because the two colours were encoded by the same component (Delvenne & Bruyer, 2004).

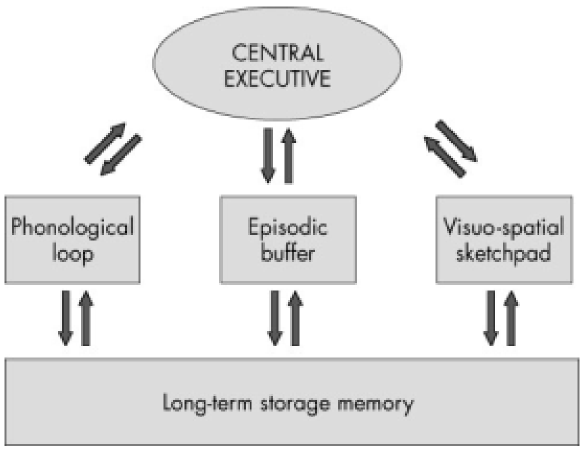


Figure 1: Working memory model illustration (From “The Episodic Buffer: A New Component of Working Memory?” by A. Baddeley, 2000, Trends in Cognitive Sciences, *4*, p. 421)

However, in everyday life working memory rarely needs to attend to separate, single-modality information. When we encounter a new object, encoding its colour, shape, texture, or whether we had seen it before, can all be useful for encoding full information about a particular object. To account for the complex nature of memory, Baddeley (2000) later added another module, the episodic buffer, which combines information from different sources (such as auditory and visual perception from the other modules in the working memory) to form multidimensional representations. In everyday life, we also rarely have to remember entirely new concepts, and a lot of the time we have to remember concepts that are already stored in long-term memory. The episodic buffer is also linked to long-term memory, which can influence representations held in the temporary storage, to form associations between new and existing information, such as using a recipe for a novel dish based on previous knowledge about cooking, or using existing knowledge in solving a new problem, for example, deciding how to approach eating a novel type of food. Evidence for a more centralised, cross-modal working memory component, such as the episodic buffer, comes from the findings that disrupting one modality (vision, speech, etc.) does not interfere with processing multi-modal stimuli. Performance on a visuo-spatial memory task was not affected by a spatial tapping task (Allen et al., 2009; Allen et al., 2015). Similarly, visual or auditory interference did not affect a verbal-spatial task performance, where letters displayed within squares randomly placed around the screen had to be remembered (Langerock et al., 2014). Articulatory suppression, which consisted of merely repeating a meaningless syllable, such as "lalala" (Delvenne & Bruyer, 2004) did not disrupt binding multiple visual features (squares and colours) into objects. However, activating or holding information in working memory can be disrupted by an interference task which consists of processing a multimodal object, for example mental rotation of a visually presented letter (He et al., 2020), or backward counting out loud (Allen et al., 2009). Cowan (2001; 2010) proposed a core working memory storage that does not rely on perceptual information, much like the episodic buffer. According to this theory, when the perceptual trace is erased by auditory interference or visual masking, participants still remember around 3-5 visually presented shapes or auditorily presented digits (Chen & Cowan, 2009; Saults & Cowan, 2007). Although the core working memory theory suggests that perceptual information is erased rather than combined (as the episodic buffer is proposed to do), both theories support a similar storage capacity limit for complex information.

Working memory was initially proposed to have a limited storage capacity of up to 7 items or chunks of items (Miller, 1956), but the hypothesis had little empirical support. The actual capacity seems to vary widely between tasks and participants: Memory for word lists ranges from 4-6 items (Tulving & Patkau, 1962) and 2-7 items (Daneman & Carpenter, 1980), while memory for shapes and colours oscillates around 4 items (Alvarez & Cavanagh, 2004; Brady et al., 2016; Luck & Vogel, 1997). Of interest to the present work is the capacity of the episodic buffer, which is the component of working memory where multimodal conceptual representations of stimuli are likely to be stored and supported by activating conceptual information from long-term memory. In experimental studies which look at combining multimodal information, when both colour and orientation are bound and remembered together, the number of encoded cross-modal items is still 4 (Luck & Vogel, 1997; Vogel et al., 2001). A similar estimate emerges when digits and spatial locations are remembered together, compared to when they are remembered individually (Towse & Houston-Price, 2001b). There seems to be a similar capacity limit when integrating information with knowledge in long-term memory. When looking at working memory performance of visuo-spatial information, Darling and Havelka (2010) presented participants with a sequence of digits displayed either in a single location, in a horizontal line, or in a phone keyboard layout, where each digit had an assigned spatial location known to participants from their everyday experience with phone keyboards. In a later verbal serial recall task, participants were able to recall a larger proportion of items in the keyboard layout condition (by approximately 1.5 items), regardless of the length of the sequence. This was not the case when the numbers on a keyboard were presented in a different order than a standard phone keyboard (Darling et al., 2012), which indicates that relying on existing knowledge from long-term memory allowed participants to make better connections between digits. Thus, information from different modalities and long-term memory can come together to form complex representations, and those representations are able to carry more information in one unit. Similarly, Brady et al. (2016) investigated the role of existing knowledge in working memory by testing recognition memory for colours and real-life objects. In a condition where participants were presented with colours, they were able to recognise up to about 3.7 items. On the other hand, when participants were presented with real-life objects, such as a jug or a cookie, they were able to recognise up to around 4.7 objects. Critically, in both conditions participants performed an articulatory suppression task during both encoding and retrieval, to prevent them from relying on, and possibly rehearsing, verbal information. The capacity estimate in the study is therefore based only on sensory information about the stimuli.

The most accurate estimate of the episodic buffer capacity appears to be obtained by Langerock et al. (2014), who used an increasing length of the stimuli sequence in order to identify the exact number of multi-modal objects which can be remembered before the capacity limit is reached, while also avoiding a ceiling effect. In their experiments, participants were presented with a series of 2 to 7 letters displayed within squares randomly placed around the screen – that is, visual information and spatial information was combined, forming a multimodal item. They were then asked to recall what letter appeared in which location. The results indicated that the capacity of the episodic buffer was around 2.3-2.9 objects. This was lower than the capacity for items encoded via individual modalities, which could be attributed to the fact that participants encoded both letters and spatial locations, which add up to the capacity of 4 when considered as separate units of information. However, this is unlikely to be the case. Indeed, other studies found that when participants encoded integrated letters and locations, they were better at recognising the letter-in-location stimulus as previously seen, than the individual letter or location separately (Prabhakaran et al., 2000), and faster to attend to those types of objects (Bao et al., 2007), which supports the idea that multimodal objects are encoded as one unit of information. Instead, the low capacity of the episodic buffer estimated by Langerock et al. might reflect difficulty of the task, or the lack of broader context which would link stimuli together. Langerock and colleagues also showed that processing of the cross-domain stimuli was not affected by modality-specific interference, (e.g., a verbal interference task where participants were presented with a noun, and had to decide whether its referent was an animal or not), further supporting the multisensory nature of information held in the episodic buffer. These studies provide some insight into the capacity to remember multimodal objects, although the estimates vary widely between 2.3 and 6 items, and are not conclusive because they are based on different types of stimuli, and only focus on a few specific sets of items

Most working memory studies have used artificial and simplistic stimuli such as digits, colours or shapes (usually simple geometrical shapes in different colours and different spatial locations on the screen (Allen et al., 2015; Daring & Havelka, 2010; Langerock et al. 2014). Even in Darling et al. (2012), where participants were supposed to rely on knowledge from long-term memory, this was limited to knowledge about something simple and very specific - the layout of the phone keyboard. To mitigate this, Brady et al. (2016) used real-life objects as stimuli, which allowed participants to use their knowledge about the world to support individual objects (e.g., we know that a round object with chocolate chips, made of flour, is a cookie). However, these objects were not contextually related, and therefore participants could not take advantage of their knowledge about how different objects may be inter-related in their use (e.g., a jug of milk can be put next to a plate of cookies as a snack). This use of information from long-term memory (which can allow items in the episodic buffer to mutually support each other) is what increases working memory capacity for contextually related objects compared to unrelated objects. In fact, this attribute of memory has been taken advantage of since ancient times in the Method of Loci (Bower, 1970), also known as a Memory Palace, where mentally placing discrete items in a familiar location, such as a house, makes them easier to remember since it provides a context that links them together. This technique has been found to enhance memory performance (Legge et al., 2012), lending support to the idea that working memory capacity is higher for contextually related information supported by long-term memory.

Moreover, studies of cross-modal representation are limited to visual, spatial, and sometimes auditory information. While these types of stimuli give a glimpse of how information retrieved from long-term memory can support working memory, they do not fully reflect the richness of the world around us and the complexity of the concepts we use on a daily basis, for example when we are trying to remember a recipe for a cake or to mentally plan how to decorate our house. These concepts, which are what WM has to manage in everyday life, and which are normally stored in long-term memory, seem to fit the description of the multi-modal, long-term memory-dependent objects stored in the episodic buffer (Baddeley, 2000). However, little attention has been given to studying real-life, contextually related stimuli, which reflect all aspects of conceptual representations in working memory, such as use of objects (e.g., what is used for cutting or holding things), action and other motor information (Banks et al., in prep; Jaroslawska et al., 2018; Pezzulo et al., 2010) or texture, smell, taste and other sensory information. For example, working memory for smells is not negatively affected by modality-specific interruption (Zucco, 2003), which raises further questions on how sensorimotor information from long-term memory is represented in working memory tasks.

3.2 Supporting working memory with long-term memory

Some tasks, especially those more closely reflecting real-life tasks, such as reading a story, clearly exceed the capacity for temporarily holding information in mind suggested by the working memory model. For example, in a word recall study, participants were able to temporarily maintain representations of over 15 words when they were presented within meaningful sentences (Baddeley et al., 2009). According to the Long-Term Working Memory Model (LTWM, Ericsson & Kintsch, 1995), in real-life tasks, short-term memory may merely serve as a cue to activate a larger amount of information in long-term memory. The LTWM model also predicts that working memory is resilient to interruptions: after prose reading is interrupted, reading a sentence serves as a cue to bring back information about previously read segments: the characters, the main premise, the point which was previously made, and the point which is currently being made, etc. (Ericsson & Kintsch, 1995; Delaney & Ericsson, 2016; Kintsch et al., 1999; Oulasvirtaa & Saariluoma, 2006). Interruption, which in working memory would have wiped the trace of objects held in mind, does not impair performance in other tasks requiring high domain knowledge. Recognition memory for objects embedded within visual scenes was the same regardless of whether encoding and retrieval were separated by up to 10 other objects (Hollingworth, 2004). Similarly, when participants were asked to produce a written narrative about baseball, and had to perform an interference task in the middle, the quality of the text (evaluated through grammar, coherency of ideas and effectiveness of communicating them) was affected less in participants who had higher knowledge of baseball (Kellogg, 2001). Clearly, when relying on expert knowledge, the memory system can process both the task at hand, and the interruption, so its capacity is larger than the capacity proposed by the working memory model.

The actual capacity of LTWM is not fixed, but can be predicted by domain knowledge: experts in soccer have greater reading span of soccer information than novices (Postal, 2004), and users of sign language remembered novel signs more easily if the signs shared their structure with existing, familiar signs in British or Swedish Sign Language (Rudner et al., 2016). This supports the notion of LTWM by showing that when a task engages information stored in long-term memory, a potentially unlimited system of LTWM is used to support task performance. Indeed, the idea of LTWM has mostly been applied to domain-specific expert performance, such as playing chess or performing mathematical calculations, which clearly exceed the “traditional” capacity of working memory of up to 6 or 7 chunks of information. For example, the digit span memory for individuals exceptionally adept at maths is estimated at between 12 and 16 digits, and chess masters can recall up to nine different chess board configurations (Ericsson & Kintsch, 1995). This happens because seeing familiar cues leads to activation of previously existing knowledge, in order to perform complex cognitive tasks (Ericsson & Delaney, 1999). Contrary to the assumptions of the episodic buffer, which state that information from long-term memory is retrieved and linked with short term memory to support maintenance of information, LTWM is active and processes information directly in long-term memory. In the LTWM framework, a large number of complex objects and ideas can be manipulated in working memory, but its capacity is not fixed, and could potentially be unlimited, activating more and more information stored in long-term memory. It appears that with the right type of cues, for example passages from a book, one could continue to perform well for a nearly unlimited amount of time and number of stimuli by activating more and more information from long-term memory.

However, the question of working memory storage capacity, regardless of how it is linked with previously acquired information, refers to the information that we can simultaneously hold in mind. There has been little attempt to quantify LTWM units of processing, or units of measurements of capacity for reading text, for example, or to quantify how much information can be retrieved from a given set of cues. Neither of the existing theories can provide an estimate on how many units of complex, real-life information the episodic buffer or the LTWM can hold. In other words, while our ability to hold information simultaneously is limited, it is difficult to precisely pin down its constraints beyond a few specific sets of information such as words, colours, or chess moves. Although theoretical accounts of both the episodic buffer and the LTWM address the links between objects encountered in real-time and long-term memory storage, no research focuses explicitly on the amount of information that can be retrieved from long-term memory simultaneously.

3.3 Memory in the linguistic-simulation perspective

The LTWM model is somewhat closer than Baddeley's model to the simulation perspective of conceptual representations (discussed in section 2.1), which states that knowledge can be represented using similar neural pathways to those used at the point of encoding information, which are stored in long-term memory. However, working memory has not received much attention from the linguistic-simulation perspective, even though representations in working memory are clearly modality-specific (Buchsbaum & D'Esposito, 2019; Wilson, 2001) – that is, with separate processing of vision, sound and language. The involvement of sensory modalities in working memory is evidenced both by traditional working memory research (where the visuospatial sketchpad encodes visual properties of the stimulus and the phonological loop encodes phonological and articulatory properties, with modality-specific interference disrupting this process; Baddeley & Hitch, 1974), as well as neuroimaging studies showing activation in perception- and action-related brain regions during verbal short term memory tasks (Buchsbaum & D'Esposito, 2008; Koenigs et al., 2011). For example, the maintenance of phonological and auditory information is associated with brain activity in the auditory dorsal stream (Hickok & Poeppel, 2007; Kummerer et al., 2013; Markiewicz & Bohland, 2016). If concepts consist of sensorimotor simulations (as proposed in section 2.1 of this thesis), then it would make sense to rely on them to temporarily represent and manipulate conceptual information even in the absence of input (Wilson, 2001). Additionally, the temporal nature of working memory suggests that we should move away from a rigid object-based framework to transient activation and deactivation of multi-modal networks of neurons associated with incoming stimuli (Langerock et al., 2018; Wilson & Emmorey, 1998). Indeed, it was already suggested by Hebb (1958) that complex cognitive processes like memory are supported by the same mechanism as simple stimulus perception and reaction. That is, as perceiving and interacting with an object activates neurons responsible for the sensorimotor information associated with its representation, the neurons which remain temporarily activated are what constitutes working memory content; they can then be used to inform a task until the activations fade or are pushed out by more incoming stimuli.

Studies on the involvement of sensorimotor simulation in memory suggest a similar system to the working memory model, where multiple modalities come together to represent and hold an object. Much as in language processing, evidence suggests that modality-specific resources are occupied with word retention, and interference impairs modality-specific memory. For example, performing an action with the hands or feet resulted in poorer memory for hand- and feet-related words (e.g., “clap”, “kick”), respectively (Shebani & Pulvermuller, 2013). When a visual or auditory distractor was presented (and attended to) between word presentation and recall, recall of words related to visual experience (e.g., “light”) was better when the distractor object was presented aurally, and vice versa – recall for words related to sound experience (e.g., “echo”) was better when the distractor object was presented visually (Vermeulen et al., 2013). Similarly, action verb recall was impaired when participants sat with their hands behind their back during encoding and recall, which was intended to inhibit simulation of hand action (Dutriaux et al., 2018). The impairment of action verb recall was also the case for individuals who were born without hands, and thus had limited ability to perform simulation of hand movements, which affected their memory for action verbs compared to healthy individuals (Vannuscorps & Caramazza, 2016). Vermeulen et al. (2008) also touched on the capacity of working memory from the simulation perspective, as well as its role in problem-solving rather than just remembering and repeating information. Participants had to perform a property-verification task (e.g., is “lemons are yellow” – true/false), while holding 1 or 3 visual or auditory items in working memory. Responses to auditory properties were slower when WM load consisted of auditory stimuli, and responses to visual properties were slower when WM load consisted of visual stimuli. In other words, holding stimuli in a specific modality slowed down processing of sentences in the same modality. Notably, the effect was stronger when WM load was higher, indicating that when participants had to hold 3 items in mind and process sensorimotor information in the same modality, their working memory was under strain, which led to slower, more effortful processing, which was not the case in the low load condition. Dijkstra et al. (2007) also found that body posture facilitated retrieval of autobiographical memories – when retrieval of a memory for autobiographical information (e.g., attending a concert) was accompanied by associated posture or movement, participants recalled the same event better later on. These results suggest that memory for modality- and effector-specific stimuli is processed in the same sensorimotor system as grounded concepts and language.

The role of language in short-term memory has been touched on in a few theories. Both the redintegration account (Hulme et al., 1997), and the semantic richness theory (Buchanan et al., 2001; Yap et al., 2012) predict that words with higher frequency of occurrence (or potentially words rated higher on other semantic dimensions such as number of features, emotional valence, or semantic neighbourhood) are easier to remember in a serial recall task, because of stronger representations in long-term memory, which make them easier to reconstruct from decaying traces in short-term memory. Of interest, when low frequency words are repeatedly presented in the same pairs, they are remembered much better, while memory for high-frequency words is not affected by word pairing (Stuart & Hulme, 2000), suggesting that linguistic co-occurrence might be what drives the frequency effect in memory. After all, linguistic labels are fast at activating their frequently co-occurring neighbours (Solomon & Barsalou, 2004; Connell & Lynott, 2013), and there is evidence that statistical co-occurrence allows for chunking words and thus increasing WM capacity (Brady et al., 2009). Additionally, instructing participants to explicitly label stimuli increases the probability of correctly recognising them later (Zormpa et al., 2018), and increases the number of remembered objects by half an item in a set of 4 (Souza & Skóra, 2017). Hutter et al. (2016) found that novel social combinations, which require activating knowledge about two unrelated concepts, such as “female blacksmith”, are created in the working memory. However, the utility of linguistic labels in working memory, compared to sensorimotor information, has not been directly investigated.

3.3.1 Linguistic bootstrapping in short-term memory

As discussed in Chapter 2, conceptual representations comprise both linguistic and simulated representations, and both types of information can be used to hold concepts in WM. For example, when we want to remember something in real life, such as a cake recipe, information from long-term memory about what cake usually consists of, and the way and order in which ingredients combine, encoded through both perceptual and action experience, as well as linguistic experience, can mutually reinforce one another and help us to remember the right ingredients. But since linguistic information occupies less space within working memory, the linguistic bootstrapping mechanism makes it more efficient to remember linguistic labels and then rely on their ability to activate larger and more complex associated sensorimotor information.

Section 2.1.3 outlined how linguistic representations are computationally cheaper, and therefore can potentially be used as a bootstrapping mechanism, either by representing information on their own, or by activating sensorimotor representations when possible and necessary. In everyday life, an object is encountered either through direct sensorimotor input, or through a linguistic label which allows for offline processing (Wilson, 2002). Incoming sensorimotor information, such as seeing a dog, an apple or a glass of milk, can be matched with a representation in long-term semantic memory, so that the concept can be recognised, and part of that information will be held in working memory to address the task at hand (such as patting a dog or making an apple pie). Similarly, an incoming linguistic label can do the same – activate the associated sensorimotor information held in long-term memory, so that it can be held and manipulated in working memory. Alternatively, as outlined in the linguistic bootstrapping hypothesis (Connell & Lynott, 2014b), the label itself can be held in working memory, which reduces the load when there are time or space constraints, or simply because it may be sufficient to store a linguistic label when, for example, trying to memorise a shopping list. The size and features of linguistic representations are therefore an advantage in manipulating a number of concepts at the same time.

3.4 The present thesis

The linguistic bootstrapping hypothesis has not been explicitly applied to working memory research before. In the present thesis, I will address the question of whether linguistic bootstrapping allows an increase in working memory capacity. Assuming that working memory has a stable, limited capacity, the number of linguistic representations simultaneously activated in working memory is expected to be higher than the number of sensorimotor representations held in working memory at the same time. The reliance of working memory on language for more efficient processing and higher capacity can be investigated using a paradigm where language is not available (when articulatory suppression is performed). In such case, one has to refer to perceptual simulations which take longer to retrieve (Louwerse & Connell, 2011) and are more complex because they consist of information from different modalities. Working memory capacity for complex, contextually related everyday-life objects when language is and is not available will be estimated in Chapter 4.

The present thesis will also address the nature of conscious imagery in conceptual representations (Chapter 5). I will examine how consciously available information relates to sensorimotor experience, and whether different measures of conscious imagery capture the same types of information. If conscious imagery reliably captures the same aspect of meaning (i.e., whether the concept has an imagistic code) that is essential to conceptual processing, as predicted by dual-coding theory, then different sources of imageability ratings should consistently predict performance in word recognition tasks. Additionally, Chapter 6 will look at the role of conscious imagery in word memory. Generating mental images may be particularly useful in a setting where participants are explicitly required to remember something. However, it may not play the same role in everyday life situations when people do not always use strategies to remember things around them. This will be tested using a surprise memory task, and by analysing the contribution of imageability to word memory performance when information about word meaning is accounted for by sensorimotor information.

Finally, Chapters 5 and 6 will also consider the role of sensorimotor information in memory of concepts. First, Chapter 5 will use sensorimotor strength as a predictor of word recognition to consider whether it supports retrieval of concept knowledge from long-term memory in a word recognition task. In Chapter 6 I will test the contribution of sensorimotor information to word memory. This will investigate the role of different types of sensorimotor information and will test the predictions of the semantic richness theory that an increase in variable strength will always facilitate performance in a conceptual task. If the additive effect of semantic richness appears, words rated higher in sensorimotor experience should always be remembered better. This will also be compared between a surprise and an expected memory task, but the subconscious nature of sensorimotor experience should in principle support word memory to the same extent.

Chapter 7 will summarise the main findings of the thesis, discuss their contributions to the study of cognitive processes, as well as their limitations, and the potential for future research. The findings will contribute to a more comprehensive understanding of conceptual representations, and the interplay of linguistic and sensorimotor information in memory processes.

4 Language Increases Working Memory Capacity for Object Concepts

This chapter sets out to examine the role of linguistic labels in working memory. In particular, I will investigate whether the linguistic bootstrapping hypothesis applies to working memory, where participants have limited time and capacity resources to remember a number of objects. This hypothesis will be tested in a paradigm where some participants perform an articulatory suppression task at encoding or retrieval to block their access to language and encode sensorimotor information only, while others are free to rely on linguistic placeholders. If linguistic bootstrapping is employed as a strategy to save limited resources, participants with access to language will remember a higher number of objects when their working memory is under strain than participants who cannot rely on labels due to the articulatory suppression task.

This paper was submitted for publication in March 2021 and was produced in collaboration with Professor Louise Connell and Dr Briony Banks as co-authors. It is available as a preprint at:
Dymarska, A., Connell, L., & Banks, B. (2021). Language Increases Working Memory Capacity for Object Concepts [preprint]. https://doi.org/10.31234/osf.io/5v6de

Language Increases Working Memory Capacity for Object Concepts

Agata Dymarska, Louise Connell, and Briony Banks

Department of Psychology, Lancaster University.

Author Note

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 682848) to LC.

Correspondence concerning this article should be addressed to Agata Dymarska, Department of Psychology, Fylde College, Lancaster University, Lancaster LA1 4YF, UK. Email: a.dymarska@lancaster.ac.uk

 Stimuli, data and analysis code can be found at: https://osf.io/8sm3p/?view_only=853111f26db840b4a0558275496d77e9.

Abstract

The linguistic-simulation approach to cognition predicts that language can enable more efficient conceptual processing than purely sensorimotor-affective simulations of concepts. We tested the implications of this approach in working memory, where use of linguistic labels (i.e., words and phrases) could enable more efficient representation of concepts in a limited-capacity store than representation via full sensorimotor simulation; a proposal called *linguistic bootstrapping*. In four pre-registered experiments using a nonverbal recognition memory paradigm, we asked participants to remember sequences of real-world objects, and used articulatory suppression to selectively block implicit activation of linguistic labels, which we predicted would impair object memory performance. We found that blocking access to language at encoding impaired memory accuracy, though not latency, and that this impairment was not simply dual-task load. Results show that a sequence of up to 10 contextually-situated object concepts can be held in working memory when language is blocked, but this capacity increases to 12 objects when language is available. The findings support the linguistic bootstrapping hypothesis that working memory for familiar object concepts normally relies on language, and that implicitly-retrieved object labels, used as linguistic placeholders, enhance the achievable capacity of working memory beyond what sensorimotor information alone can accomplish.

Keywords: concepts; sensorimotor simulation; working memory; linguistic labels; linguistic bootstrapping

Language Increases Working Memory Capacity for Object Concepts

The conceptual system consists of simulation- and linguistic-based components (Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2008; Vigliocco et al., 2009). Simulated representations engage the neural subsystems involved in sensorimotor, affective, introspective, and other situated experience of a concept (e.g., Barsalou, 1999; Martin, 2007). For example, the experience of a dog may include its visual shape and colour, the action and feel of patting its fur, the sound of its bark, the broader situation of walking it on a leash, and the love and positive feelings towards a pet. The neural activation patterns involved in processing these experiences can be partially re-activated (i.e., simulated) at a later time when representing a concept. Linguistic representations of concepts, on the other hand, comprise word (and phrase) labels associated with these sensorimotor-affective simulations; for instance, seeing a terrier or hearing a bark will activate the label “dog”, as well as other associated words that represent experiences in related contexts, such as “tail”, “walkies”, or “leash” (e.g., Louwerse, 2011; Wingfield & Connell, 2019). These simulated and linguistic components are interrelated and mutually supportive, and recent theories argue that both are intrinsic to conceptual representation (e.g., Connell & Lynott, 2014; Louwerse, 2011). That is, linguistic labels are part of concepts, and conceptual processing utilises simulation and linguistic information to varying extents depending on task demands, available resources, and other factors (Connell, 2018; Connell & Lynott, 2014).

The role of both simulation and linguistic components in conceptual processing is supported by a range of empirical evidence. Support for sensorimotor simulation comes from neuroimaging of sensory and motor cortices during word processing (e.g., Goldberg et al., 2006; Hauk et al., 2004), neuropsychology of motor impairment (e.g., Boulenger et al., 2008; Fernandino et al., 2013), and a variety of behavioural paradigms involving perceptual or action manipulations (e.g., Bidet-Ildei et al., 2017; Connell et al., 2012; Davis et al., 2020).

For instance, Shebani and Pulvermuller (2013) found that performing an arm or leg movement interfered with working memory for action words associated with those effectors (e.g., arm movements impaired memory for words like *grasp* and *clap*). Support for the linguistic component comes from computational modelling of conceptual information captured in language (e.g., Banks et al., 2020; Riordan & Jones, 2011; Wingfield & Connell, 2019) and from behavioural paradigms showing that information from language alone can inform responses in diverse conceptual tasks (e.g., Connell & Lynott, 2013; Goodhew et al., 2014; Louwerse & Jeuniaux, 2010; see Connell, 2018, for review). For example, because the linguistic component has a relative speed advantage over the simulation component (Barsalou et al., 2008; Connell, 2018; Louwerse, 2011), responses that rely on language tend to be faster and less effortful than those that rely on sensorimotor simulation (e.g., Louwerse & Connell, 2011; Santos et al., 2011). However, much evidence for the linguistic component centres on the usefulness of linguistic distributional knowledge (i.e., the statistical patterns of how words/phrases co-occur across language), which does not encompass the full role of language in conceptual processing. The very existence of linguistic labels – that is, being able to concisely name a complex multimodal experience with a word or phrase – provides another means for the linguistic component to enhance the efficiency of conceptual processing.

The idea that language is beneficial for our cognitive processing has existed for some time (e.g., Paivio, 1971; Vygotsky, 1934/1986). Recent theories have, however, developed the role of linguistic labels in a number of new directions (e.g., Borghi et al, 2018; Connell, 2018; Lupyan, 2012). Most relevant to the present article, Connell and Lynott (2014, p. 7) propose that having labels for concepts enables a process of *linguistic bootstrapping*, whereby words and phrases act as linguistic placeholders in an ongoing representation when there are insufficient resources to maintain a sensorimotor simulation in full, thus enhancing the achievable size and complexity of what can be held in mind. That is, when the sheer scale or complexity of what one is trying to simulate at a given moment outstrips the limited resources of the human cognitive system, replacing a portion of the simulation with a linguistic placeholder can preserve structure in the representation while freeing up resources to maintain or extend the simulation as needed. These linguistic placeholders can later be fleshed out into a simulation again at any time if resources become available. While not directly framed as a working memory hypothesis, the implication of linguistic bootstrapping for memory capacity is clear: having language available to label concepts should enable more efficient use of working memory and ultimately increase the number of concepts that can be remembered.

To date, the linguistic bootstrapping hypothesis has remained theoretical and has not been tested directly. There is some evidence that memory relies on linguistic representations and the sensorimotor representation is dropped where possible. This may be reflected in the overshadowing effect, where activating a verbal representation leads to impaired processing of visual stimuli (e.g., perception of a face, Schooler & Engstler-Schooler, 1990). In a study by Brandimonte et al. (1992), when participants were presented with easily nameable objects in a study phase, they encoded them using linguistic labels by default. This led to impaired performance in a later image manipulation task, which required memory of specific visual characteristics (i.e., the shape and orientation of the presented stimulus), rather than just memory of a concept (e.g., a fish, a car). This was not the case, however, when participants performed articulatory suppression (while encoding items), and therefore were forced to rely on sensorimotor information to remember the specific characteristics of the presented stimulus. Similarly, Hitch et al. (1995) and Walker et al. (1997) asked participants to perform articulatory suppression when learning to combine two images into one object. Memory for the objects was enhanced under the articulatory suppression condition, when participants could perform the task by retrieving part of the image from memory, instead of using more abstract representations encoded verbally. These findings may appear to contradict the linguistic bootstrapping hypothesis predictions that availability of labels enhances memory performance. However, it is important to note that in these studies verbal encoding facilitated access to a more general representation, which in turn impaired the memory for a specific image with its unique visual characteristics. Since participants were asked to manipulate a specific representation in a novel way, and not tested a general ability to remember a previously presented concept, it is expected that using language to label the concept could have affected the visual representation negatively. On the other hand, linguistic bootstrapping predicts that a label allows for manipulating a conceptual representation without the need for activating the complex sensorimotor information (Connell & Lynott, 2014b), when performance need not rely on perceptual features such as the shape or colour of a specific example of the concept.

However, little research has focused on memory from the linguistic-simulation perspective, and what research exists has concentrated on the role of sensorimotor simulation in memory (e.g., Dutriaux et al., 2018; Vermeulen et al., 2013) rather than examining the interplay of simulated and linguistic information in working memory capacity limits. Nonetheless, there is indirect support for the idea in the wider literature, particularly in working memory research. According to the most recent versions of the multi-component working memory model (Baddeley, 2000; 2012), when processing complex stimuli, information from multiple modalities is integrated with conceptual representations from long-term memory and stored in the episodic buffer. This episodic buffer storage is necessarily limited in capacity: that is, there are only so many concepts that can be maintained and manipulated at once. Empirical studies estimate the capacity of the episodic buffer (for combination of letters/digits and their spatial location) to be from 3 items (Langerock et al.,

2014) to 5 or 6 items (Allen et al., 2015), which is similar to other estimates of a central capacity limit comprising 3-5 items (Cowan, 2001). Critically, other studies suggest that linguistic information is more economical in representation (i.e., may occupy less "space" in working memory) than sensory information, and may thus allow this capacity to be increased. For example, explicitly labelling simple visual stimuli (e.g., dots of different colours), appears to increase memory capacity compared to unlabelled stimuli (Souza & Skóra, 2017; Zormpa et al., 2018). It is therefore possible that when working memory capacity is strained to its limit, such as when trying to maintain a representation of numerous concepts, a linguistic label could deputise for its referent sensorimotor information (e.g., the word "dog" could replace the simulation of a *dog*) in order to free up space. Examining working memory capacity for concepts may thus provide a means for directly testing the linguistic bootstrapping hypothesis, as well as for estimating the potential benefit to working memory capacity afforded by linguistic labels.

When conceptual information from long-term memory can actively support working memory, it follows that the extent of such support will vary according to the nature of the stimuli to be remembered. Critically, it may also be the case that the *capacity* of working memory will vary according to such support. Working memory research has often concentrated on using relatively simple, artificial stimuli (e.g., visual feature conjunctions such as a *red triangle*, random word pairs such as *desk-ball*) which, while useful for examining subcomponents of memory, are not ecologically valid instances of what humans typically hold in memory during daily life. Such unrepresentative, contextually-unrelated stimuli – and their derived limits of working memory capacity – do not easily generalise to more naturalistic, real-world concepts that comprise rich sensorimotor and linguistic information from long-term memory, and that are typically represented in broader situated simulations that allow concepts to reinforce and cue one another (e.g., a *dog* that is *running* with a *ball*).

Support from long-term memory makes familiar, real-world, situated concepts easier to remember than more artificial, novel experimental stimuli, because even when items are partially erased from working memory, they may be retrieved based on activation of associated information, serving as cues. For example, children and adult participants were able to recall a previously heard item when given a semantic category it belonged to (Roome et al., 2019). This is also the case when participants were presented with meaningful sentences, where words cue one another and allow for retrieving a larger number of items compared to lists of unrelated words (Baddeley et al., 2009). Indeed, retrieval of associated items occured even when the items themselves were not presented, and participants reported them falsely (see Roediger & McDermott, 1995). This supports the idea of two kinds of systems or processes in working memory, where conscious awareness allows for remembering a core set of cues (such as linguistic labels) in the primary working memory (Unsworth & Engle, 2007), and these may be expanded on by retrieving information from secondary memory. As a result, working memory capacity in a task which allows for activation of supporting information from long-term memory, based on semantic cues, may be higher than previously estimated based on simple, unrelated stimuli. For example, presenting participants with a recipe for a cake which includes the name of the recipe and the order of mixing ingredients, and situates the participant in the context of making a cake for a birthday party, allows for the kind of support from long-term memory that occurs when performing these tasks in real life.

This idea is addressed to some extent by the Long-Term Working Memory model (LTWM; Ericsson & Kintsch, 1995), which suggests that information received by short-term memory serves as a cue to activate knowledge from long-term memory. According to the model, we can remember a large number of complex memories (e.g., objects and ideas during domain-specific expert performance, such as playing chess or reading a book) if they are already present in our long-term memory storage. However, the model does not provide an estimate of LTWM capacity, instead suggesting that performance on complex cognitive tasks and the capacity of LTWM is predicted by domain knowledge. Nonetheless, there is some evidence that people might have a higher working memory capacity for familiar, more complex concepts than for simpler, artificial stimuli. For example, Brady et al. (2016) found that participants could hold more real-world objects than colours in working memory, despite their greater complexity as stimuli (4.7 vs. 3.7 colours). Similarly, users of sign language discriminated studied and non-studied novel signs in an *n*-back task more easily if the signs shared their structure with existing, familiar signs in British or Swedish Sign Language (Rudner et al., 2016). While such findings are suggestive of greater capacity for real-world concepts, their focus on contextually unrelated stimuli meant that a potentially important aspect of conceptual support from long-term-memory – mutually reinforcing contextual situations – could not be utilised. Hence, the limit of working memory capacity for such real-world concepts when contextually situated in a naturalistic sequence (e.g., remembering a list of ingredients for a recipe) remains unknown.

The Current Study

The present study had two main aims: to examine the role of linguistic bootstrapping in working memory for real-world object concepts, and to establish the capacity of working memory for recognition of real-world object concepts both when access to linguistic labels is available and when it is not. In four pre-registered experiments using a nonverbal paradigm, we presented ecologically valid sequences of object pictures from naturalistic situations (e.g., ingredients for a novel recipe) and then tested recognition memory by asking participants to select the previously presented objects from arrays of distractors. Critically, in Experiments

1-2, participants performed articulatory suppression (i.e., repeated aloud "the") during item encoding and/or retrieval, in order to block access to linguistic information (i.e., object labels) while leaving access to sensorimotor simulation unaffected. Articulatory suppression has been widely used in working memory research (e.g., Baddeley, 1992), where it has been shown to interfere with verbal encoding but to have little effect on the central executive and general cognitive processing (e.g., De Rammelaere et al., 2001; Jaroslawska et al., 2018; Larsen & Baddeley, 2003). We hypothesised that, even in ostensibly nonverbal paradigms, storage of object concepts in working memory would normally rely on language (i.e., implicitly-retrieved object labels, used as linguistic placeholders), and that blocking access to language would impair speed and accuracy, and reduce the capacity, of working memory for object concepts. In Experiment 3, we included an additional control condition of foot tapping in order to compare performance in the articulatory suppression condition with a secondary task that had some attentional demands but that would not affect access to linguistic information. Finally, in Experiment 4, we addressed the possibility of ceiling effects in earlier studies by using sequences of increasing length to determine the upper limit of working memory capacity for contextually situated, real-world concepts when linguistic labels were fully available (i.e., without articulatory suppression).

Experiment 1: Articulatory Suppression

In this study (pre-registration, materials, data, analysis code, and full results are available as supplemental materials on OSF) we presented participants with images of natural and artifact objects, arranged in sequences of six items that would plausibly be experienced in a real-world setting, and asked them to remember each sequence. After each sequence, we tested memory for the objects by asking participants to choose each remembered object from an array of related distractors, and measured speed and accuracy of performance. Participants performed articulatory suppression during encoding and/or during retrieval. Following the linguistic bootstrapping hypothesis, we predicted that performance would be impaired when access to language was blocked, such that articulatory suppression at either stage would lead to slower responses and more errors in identifying remembered objects. We expected memory performance to be best with no articulatory suppression at either encoding or retrieval, where participants would be free to utilise both linguistic and sensorimotor information to remember the objects. In addition, we expected performance to be worst with articulatory suppression at retrieval only, due to participants employing linguistic placeholders to replace sensorimotor information when encoding some objects, and then losing access to those placeholders (and thus the object representations) at retrieval when access to linguistic information was blocked. Finally, by calculating the average number of objects correctly retrieved with articulatory suppression at both encoding and retrieval (i.e., when linguistic information was fully unavailable throughout the task), we planned to estimate the capacity of working memory for sensorimotor representations of real-world concepts.

Method

Participants

Thirty-two native speakers of English (27 female; mean age = 21.2 years, *SD* = 3.2 years) were recruited from Lancaster University, and received course credit or a payment of £3.50 for participation. One participant was replaced due to a procedural error during testing. The sample size was determined using sequential hypothesis testing with Bayes Factors (Schönbrodt et al., 2017). As pre-registered, we stopped at the minimum sample size N_{min} = 32 when our Step 3 models for both RT and accuracy cleared the specified threshold of evidence $BF_{10} < 0.20$ (see Design and Analysis section for model details; full statistics are reported in the Results section).

Materials

Test items comprised a total of 72 target objects, divided into 12 sequences which were each designed to be an ecologically valid order of objects that would be plausibly used in a real-world setting, such as ingredients used in the process of making a cake, a set of tools used in order to hang a picture, or an outfit one might dress in for a particular sports activity. Each sequence therefore consisted of 6 target objects for study during encoding, and each target object was assigned five distractor items for display in an object array during the retrieval (testing) stage. The sequence length of 6 target objects was chosen to be greater than the estimate of 2-3 items for episodic buffer capacity proposed by Langerock et al. (2014), and comparable to the upper estimate of 5-6 items found by Allen et al. (2015). Five distractor objects were selected from the same semantic category as the target (e.g., food items, clothing) of which three were chosen to share the same colour, shape or function as the target object. The target and distractor items in each sequence (e.g., a recipe) could all plausibly be used for similar activities, so that the task maintained ecological validity, and it would not be obvious from the nature of the sequence which item in the array was the correct one (see sample stimuli in Figure 1). Each sequence was named according to the real-world scenario it represented (e.g., “recipe for a cake”) to provide participants with the situated context prior to encoding.

We sourced photographic images for all objects from license-free online resources and edited them to appear on a uniform transparent background. Critically, in order to ensure that participants were tested on memory for object concepts, and not perceptual matching of a specific image, we prepared two different images for each target object: one for study during encoding and one for display in the distractor array during retrieval. Both images represented good examples of the target object and differed only in minor aspects (e.g., showing a vegetable from a different perspective, or a piece of clothing in a different colour). Images were scaled to be 840 pixels along the longest dimension for target objects presented during the encoding stage, and 470 pixels along the longest dimension for objects (targets and distractors) presented in the object array during retrieval. This process resulted in a total of 504 object images: 72 target objects presented at encoding, 72 target objects presented at retrieval, and 360 distractor objects presented at retrieval. Figure 1 shows sample stimuli in a trial sequence diagram.

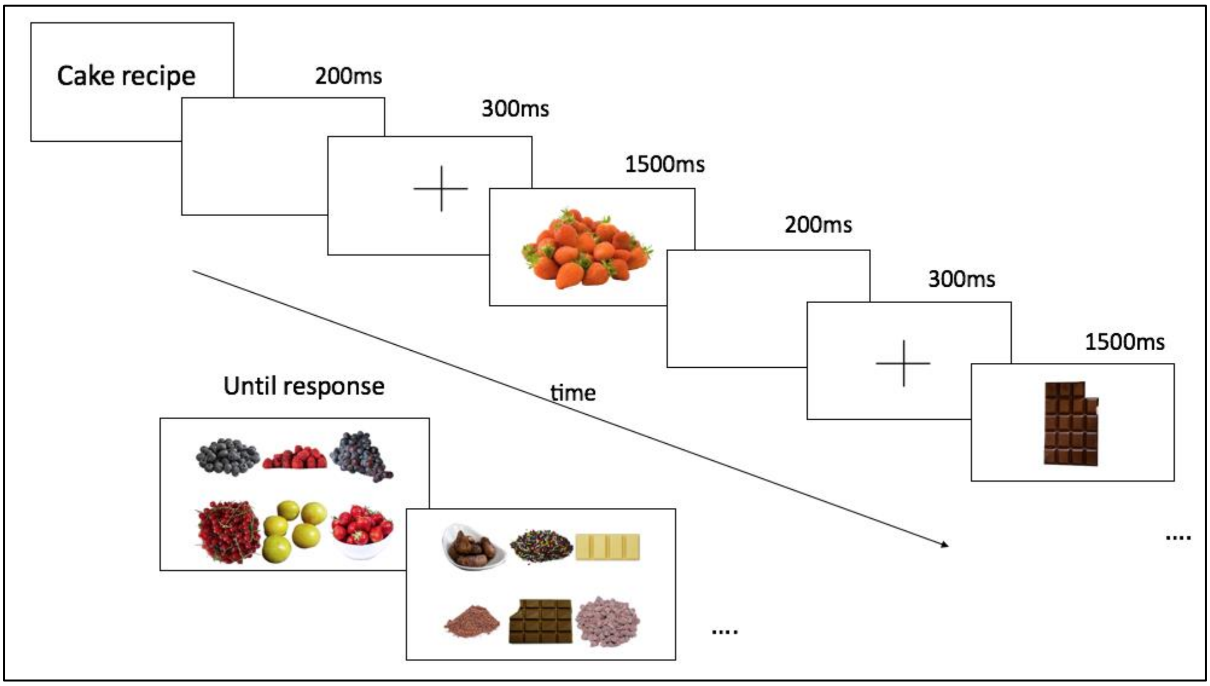


Figure 1: Diagram showing trial sequence and example stimuli at encoding (above) and retrieval (below) stages in Experiment 1.

To ensure the order of target objects within each sequence was ecologically valid, we asked 9 naïve participants (who did not take part in the main studies) to rank-order the items according to how they would be used in the given situated context. For example, in the context “Tools for hanging a picture on the wall”, participants had to decide the order in which they would use the following objects: “spirit level”, “drill”, “screw plug”, “screw”, “screwdriver”, “picture frame”. We then finalised each sequence according to the mean rank per object. Target objects were always presented in the same ecologically-valid order at both encoding and retrieval.

Procedure

Participants were tested individually. After signing the consent form, which included consent to publicly share their anonymised data, they sat in front of a computer and were informed that they would perform a memory task, and that they would be asked to repeat the word “the” at some point during the task. We chose to use the word “the” for the articulatory suppression task (as opposed to pseudo-nonsense syllables such as “la” or numbers such as “one”) because it was a real word that participants were practiced at articulating, but, as a function word, was semantically empty in isolation and so unlikely to activate any linguistic or sensorimotor information that could interfere with the task. The experimenter then explained and demonstrated articulatory suppression, and asked the participant to practice it. Once the participant confirmed that they understood and could perform articulatory suppression correctly, they provided demographic information and read the instructions onscreen.

Participants were instructed that they would see a sequence of everyday objects appear one-by-one onscreen, and their task was to remember the objects; later, they would see groups of objects onscreen and they should click on the object that belonged to the sequence they had been asked to remember. Each sequence was preceded by its name (e.g., “pasta” or “cake”). Participants then commenced a practice sequence of six items (not used in the main experiment), without any articulatory suppression at encoding or retrieval. After the practice session, when the participant confirmed that they understood the task and were happy to continue, they were given verbal instructions regarding when to start and stop articulatory suppression at both encoding and retrieval, and commenced the experimental trials. Verbal reminders were given between sequences if the participant stopped performing the suppression task during the trial. Articulatory suppression was manipulated between participants at encoding and within participants at retrieval, producing four crossed experimental conditions: no-suppression/no-suppression, no-suppression/suppression, suppression/no-suppression, and suppression/suppression. The order of retrieval conditions was counterbalanced, and six sequences were presented in a randomised order within each condition. Experiment presentation was controlled by PsychoPy software (version 1.84.1; Peirce, 2009).

In the encoding stage, participants in the articulatory suppression condition commenced repeating "the" aloud before each sequence began. The name of the sequence was first presented onscreen for 300ms. Target objects were then presented individually in their fixed sequence, starting with a blank screen for 200 ms, followed by a central fixation cross for 300 ms, and then the target object for 1500 ms (see Figure 1). Once a full sequence of six target objects had been presented, participants saw a "wait" screen of 3 asterisks ("***") for 10 seconds before the retrieval stage began. If participants were performing articulatory suppression at encoding, they continued repeating "the" aloud until the wait screen timed out. In the retrieval stage, participants in the articulatory suppression condition commenced (or continued) repeating "the" aloud before the first array appeared. On each trial, participants saw a 2x3 array of six objects, comprising one target object and five distractors in random locations within the array. Response times were measured from the onset of the array display until the onset of the mouse click. There was no time limit for the response. After six arrays had been displayed (for retrieval of six target objects), a message appeared on the screen asking participants to press space when they were ready to proceed to the next sequence of objects.

After participants had completed six sequences, they were instructed to take a self-paced break. The experimenter then instructed them to perform/not perform articulatory suppression at retrieval depending on their condition (the encoding condition remained constant), and participants then completed the task for six further object sequences. The entire experimental procedure took approximately 15-20 minutes.

Ethics and Consent

The study received ethical approval from the Lancaster University Faculty of Science and Technology Research Ethics Committee. All participants read information detailing the purpose and expectations of the study before giving informed consent to take part. Consent included agreement to share publicly all alphanumeric data in anonymised form.

Design and Analysis

We analysed accuracy (with incorrect responses coded as 0, and correct responses coded as 1) with a mixed-effects hierarchical logistic regression (binomial, logit link). Participants and items (nested within sequences) were included as crossed random effects. We included fixed effects of articulatory suppression at encoding and at retrieval (dummy coded: no-suppression coded as 0, articulatory suppression coded as 1), and their interaction. Response times (RT; ms) for correct responses were analysed in a mixed-effects hierarchical linear regression with the same random and fixed effects as above.

In all regression analyses, Step 1 entered random effects, Step 2 added encoding and retrieval as fixed effects, and Step 3 added the interaction of encoding and retrieval. We ran Bayesian model comparisons between steps, with Bayes Factors (BF) calculated via Bayesian Information Criteria (Wagenmakers, 2007), in order to quantify the evidence for or against the added step (threshold for inference was $BF_{10} = 5$ or its reciprocal 1/5). We also report null hypothesis significance testing (NHST) statistics for parameter coefficients in the Step 3 model, and used these coefficients to estimate the marginal average accuracy for each condition of articulatory suppression. The marginal mean was used to calculate a range of working memory capacity limits. All analyses were run in R software (lme4 package, Bates et al., 2015; lmerTest package (Kuznetsova et al., 2017); R version 3.4.1, 2017).

Results

Based on our pre-registered criteria, no trials were excluded for the accuracy analysis. For analysis of correct RTs, one trial was excluded as a motor error (faster than 300ms), and 27 trials were removed as outliers more than 3 standard deviations from the individual participant's mean (total 0.015% data excluded).

Confirmatory Analysis

Accuracy. Bayesian model comparison showed equivocal evidence for Step 2 over Step 1, $BF_{10} = 1.58$; that is, there was only weak evidence in favour of effects of articulatory suppression at encoding and retrieval over a model containing only random effects. There was strong evidence at Step 3 *against* the presence of an interaction between articulatory suppression at encoding and retrieval, $BF_{10} = 0.02$, meaning the data were 47 times more likely under the Step 2 model without the interaction than the Step 3 model with the interaction.

Marginal means from the coefficients in the Step 3 model (Table 1, Figure 2) indicated that both encoding and retrieval parameters had negative coefficients but only retrieval had a significant effect in NHST terms. That is, as hypothesized, articulatory suppression during retrieval impaired performance accuracy. As predicted, accuracy was better when there was no suppression at encoding, as well as retrieval, with participants correctly recognizing 5.6 objects (SE = 0.1) out of 6 objects per sequence, on average, in the no-suppression/no-suppression condition. However, against our predictions, accuracy was worst when there was suppression at encoding, as well as at retrieval. Object memory was least accurate when language access was blocked at both encoding and retrieval, and participants recognized 5.0 ± 0.2 ($M \pm SE$) objects per sequence.

Table 1: Experiment 1 unstandardized regression coefficients, standard errors, and associated statistics from Step 3 models of Accuracy (logistic mixed-effect regression) and RT (linear mixed-effect regression), for articulatory suppression effects at encoding, retrieval, and their interaction.

DV	Parameter	Coefficient	SE	*df*	*z*	*p*
	Intercept	2.564	0.269	-	9.567	<.001
Accuracy	Encoding	-0.547	0.306	-	-1.787	.074
	Retrieval	-0.416	0.184	-	-2.265	.024
	Encoding*Retrieval	-0.036	0.245	-	-0.145	.885
					t	
	Intercept	2498.90	133.86	42.13	18.668	<.001
RT	Encoding	417.35	177.49	36.01	2.351	.024
	Retrieval	-210.92	60.67	1769.51	-3.476	<.001
	Encoding*Retrieval	-18.12	87.81	1770.80	-0.206	.836

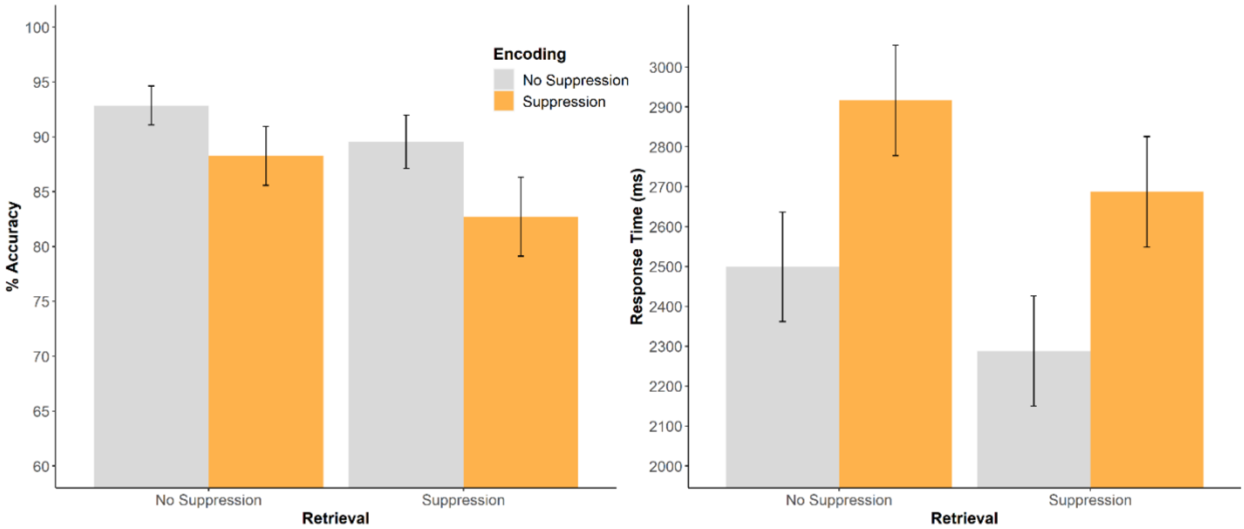


Figure 2: Mean % accuracy and RT per articulatory suppression condition in Experiment 1, calculated as marginal means from the Step 3 models. Error bars represent ± 1 Standard Error.

Response Times. Model comparisons showed very strong evidence at Step 2 for the effects of articulatory suppression at encoding and retrieval, $BF_{10} = 1808.04$. However, there was strong evidence at Step 3 *against* the presence of an encoding*retrieval interaction on RT, $BF_{10} = 0.03$: that is, the data were 33 times more likely under the Step 2 model without the interaction than the Step 3 model with the interaction.

Next, we used the coefficients in the Step 3 model (Table 1) to estimate the marginal mean RT for each articulatory suppression condition (see Figure 2). While the encoding parameter had a positive coefficient, indicating as hypothesized that articulatory suppression resulted in slower RTs, the retrieval coefficient was unexpectedly negative (i.e., faster under articulatory suppression). Against our predictions, recognition of target objects was faster when access to language was blocked at the point of retrieval. On the other hand, when language was available at encoding participants were faster at later object recognition. Performance was best (fastest) in the no-suppression/suppression condition (i.e., when language was available at encoding but not at retrieval), and worst (slowest) in the suppression/no-suppression condition (i.e., when language was available at retrieval but not at encoding). That is, participants were slowest to recognise remembered objects when language was blocked at the point of encoding but was available at retrieval.

Exploratory Analysis

Because our confirmatory analysis produced some unexpected results – in particular, there was no evidence for the predicted interaction, and articulatory suppression effects did not consistently impair performance – we ran exploratory analyses to determine the best-fitting model for accuracy and RT. We first explored alternative random effects structures for the null model using Restricted Maximum Likelihood (i.e., random encoding and/or retrieval slopes for participants and/or items) and selected the best model using Bayes Factors calculated via BIC. We then examined (using Maximum Likelihood) whether the data favoured an encoding-only or retrieval-only model in comparison to the Step 1 null model that included both encoding and retrieval. In addition, for reporting NHST statistics, we Bonferroni-corrected the p-values on parameter coefficients by multiplying by 3, to correct for examining 3 different exploratory models per DV. Model comparisons are presented in Table 2.

Table 2: Exploratory analysis of Experiment 1, showing Bayes Factor comparison of each candidate model against the null model (random effects only).

DV	Candidate model	BF_{10}
Accuracy	Encoding only	0.15
	Retrieval only	10.49
	Encoding+Retrieval	1.57
RT	Encoding only	0.32
	Retrieval only	4.95
	Encoding+Retrieval	1.49

Best-Fitting Model of Accuracy. All attempts to model random slopes led to non-convergence (using glmmTMB package in R: see supplementary materials for full results). Hence, we were unable to make valid model comparisons, and subsequently used models without random slopes (i.e., random intercepts only for participants and items, as per confirmatory analysis) as the null model in explorations of fixed effects on accuracy.

Bayesian model comparisons showed that the accuracy data were most likely under a model containing a single fixed effect of retrieval (strong evidence), followed by a model containing both encoding and retrieval (equivocal evidence), and lastly a model containing only encoding (evidence favoured the null). The retrieval-only model was BF_{10} = 6.68 times better than the next-best alternative model (encoding and retrieval), and hence represented the best-fitting model of accuracy. In this model, articulatory suppression at retrieval led to lower accuracy (b = -0.436, *SE* = 0.122, z = -3.592, *p* = .001): when access to language was blocked at the point of retrieving objects, people were 55% more likely to make an error than when language was available. This effect is slightly larger than the retrieval effect in the confirmatory analysis Step 3 interaction model, where blocking access to language at retrieval (but not encoding) meant that participants were 51% more likely to make an error.

In summary, exploratory analysis of the best-fitting model of accuracy showed that blocking access to language during retrieval impaired participants' ability to correctly remember objects (as expected) but blocking access during encoding had no reliable effect.

Best-Fitting Model of Response Time. Exploration of random slope structures showed that the best fit emerged from random retrieval slopes for participants (BF_{10} = 160.77, using Restricted Maximum Likelihood model, compared to random intercepts only for participants and items). Random encoding slopes for participants, and encoding and/or retrieval slopes for items, all offered no improvement over a model with random intercepts only (see supplementary materials for full results). We therefore used random intercepts and retrieval slopes for participants, and random intercepts for items, as the null model in explorations of fixed effects on RT.

In explorations of fixed effects, Bayesian model comparisons showed that, as with accuracy, the data were most likely under a retrieval-only model (positive evidence), followed by a model containing both encoding and retrieval (equivocal evidence), and lastly a model containing only encoding (evidence favoured the null). Notably, the addition of random retrieval slopes for participants meant that, unlike in confirmatory analysis using random intercepts only, there was no longer strong evidence for the encoding+retrieval model over the null. The retrieval-only model was BF_{10} = 3.22 times better than the next-best alternative (model with encoding and retrieval), and hence represented the best-fitting model of RT. In this model, articulatory suppression at retrieval had a negative effect on RT (b = -

224.70, SE = 63.20, t(30.29) = -3.555, *p* = .004), indicating that when language access was blocked at retrieval, participants were 225ms *faster* to respond than when language was available.

To summarise, exploratory analysis of the best-fitting model of RT showed that blocking access to language during retrieval unexpectedly speeded up participants' responses when selecting target objects, but blocking access during encoding had no reliable effect. However, closer examination of the combined effects of articulatory suppression on RT and accuracy suggested that the unexpected RT pattern was due to a speed-accuracy tradeoff rather than facilitation of performance per se: when participants were asked to perform articulatory suppression at retrieval, response times were faster than without articulatory suppression, but this was accompanied by lower accuracy. We discuss possible reasons for this tradeoff below.

Discussion

Experiment 1 examined the role of language in working memory for object concepts, specifically whether blocking access to language (using an articulatory suppression task) impairs performance in a working memory recognition task. In line with our hypothesis, we found that blocking access to language during memory retrieval impairs accuracy. On the other hand, blocking language access at encoding did not have an effect on memory; while NHST of coefficients in Step 3 indicated a significant effect of articulatory suppression at encoding, this effect was not supported by the exploratory analysis of fixed effects. There was no interaction between articulatory suppression at encoding and retrieval, contrary to our hypotheses.

Furthermore, blocking access to language while retrieving objects from working memory had unexpected effects, in that it *reduced* the time participants took to respond. This unexpected effect on RT persisted in the exploratory analysis of fixed effects, and was most likely because of a speed-accuracy tradeoff: that is, participants responded quickly but their accuracy was impaired as a result. One possible reason may have been that participants always knew, before they studied the object sequence at the encoding stage, whether or not they would perform articulatory suppression during retrieval. This advance knowledge that language would be unavailable during retrieval could have led participants to strategically rely on sensorimotor information even when they had access to language at encoding. For example, while seeing a tomato as an ingredient in pasta, they would primarily encode the colour and/or shape (e.g., *a round red thing*), and then rely on that information in the retrieval phase. While the distractors were created to ensure that individual features were not enough to identify the target (e.g., distractors had the same colour, shape or use as the target), the encoded sensory information, together with the contextual knowledge that tomatoes are often used in pasta, could allow participants to quickly guess which object was the target, without fully activating the concept of tomato. In other words, taking a “quick and dirty” heuristic approach to selecting a plausible target object could lead to rapid but error-prone responses in retrieval.

Another possibility is that performance was subject to ceiling effects. Even when language was never available (i.e., suppression/suppression condition), participants correctly recognised approximately 5.0 items per sequence on average, which suggests that they were able to represent five object concepts in working memory from sensorimotor simulation alone (i.e., more than the 2-3 items suggested by Langerock et al., 2014, as the capacity of the episodic buffer, and closer to the 5-6 items suggested by Allen et al, 2015). In fact, participants correctly recognised 5 or 6 objects in a sequence 73% of the time. It is possible that working memory capacity may never have been under particular strain, meaning that people did not have to rely on linguistic placeholders to replace sensorimotor information for the objects in order to remember the full sequence. Additionally, the exploratory analysis of individual fixed effects revealed that articulatory suppression at retrieval had a much stronger effect on accuracy than suppression at encoding, suggesting that the main effects were driven by the speed-accuracy tradeoff. In Experiment 2 we address both these potential issues.

Experiment 2: Articulatory Suppression with Longer Sequences

In our second experiment (pre-registration, data, analysis code, and full results are available as supplemental materials on OSF), we made some methodological changes to the procedure and design used in Experiment 1. First, we presented 12 objects per sequence rather than 6, in order to place greater strain on working memory capacity and therefore increase the likelihood that participants would rely on linguistic bootstrapping. Second, we randomized the articulatory suppression condition at retrieval so that participants were no longer aware whilst encoding a given sequence whether or not they would have language available during retrieval. Further minor changes are detailed in the Methods section; our hypotheses remained the same.

Method

Participants

Forty-four native speakers of English (33 female; mean age = 20.3 years, *SD* = 5.4), who did not take part in the earlier experiment, were recruited as per Experiment 1. Three initially-recruited participants were replaced: one due to not following instructions correctly, one who had previously participated in Experiment 1, and one due to not being a native speaker of English. As before, we used Bayesian sequential hypothesis testing to determine sample size. Bayes Factors for Step 3 cleared the evidence threshold for the null hypothesis at Nmin = 32 for both RT (BF_{10} = 0.02) and accuracy (BF_{10} = 0.03). However, sequential analysis plots for the Step 2 model (i.e., the best-fitting model in Experiment 1) suggested that the level of evidence was still unstable for RT (i.e., BFs fluctuated with successive participants between evidence for the null, equivocal evidence, and evidence for the alternative), so we opted to deviate from the pre-registered stopping rule and tested additional participants until Bayesian evidence stabilised at $N = 44$ (see Results section for statistics). We therefore report results for 44 participants, but full analyses at the original Nmin = 32 are available in supplementary materials, and we note that BF inferences and parameter estimates remained consistent between $N = 32$ and $N = 44$.

Materials

We used the same materials as Experiment 1, but with the following changes: To reduce the risk of ceiling effects, and tap into maximum working memory capacity, we paired all sequences from Experiment 1 to create longer sequences of objects to remember, with each pair forming a naturalistic situated context (e.g., ingredients for a meal and a cocktail formed a single shopping list). This resulted in six sequences of 12 items each. Instead of a specific sequence label for each list, participants were given more general information about the context (e.g., "You are making dinner and need to remember your shopping list for a meal and a cocktail. Press space to proceed to the list of ingredients."), to provide a plausible real-life situation that was ecologically valid. We also altered some of the distractors presented during retrieval ($N = 8$; 0.015% of all items) to ensure that they were not easier to eliminate than others in the same distractor set (which could have been the case in Experiment 1), and altered 12 images of existing targets and distractors which may have been difficult to identify in Experiment 1 (e.g., image of a chocolate bar changed to another one taken from a clearer perspective). For all changes to the stimuli, we sourced and formatted photographic images as per Experiment 1; all items are available in supplemental materials.

Procedure

The procedure was the same as Experiment 1 with the following changes. To prevent participants' knowledge of when articulatory suppression was going to take place from affecting their encoding strategies, we altered our instructions to participants regarding articulatory suppression. Instead of providing verbal instructions at the start of the experiment, participants saw an image of a mouth on the screen to indicate when they should start performing articulatory suppression, and the same image crossed out to indicate that they should not perform articulatory suppression; these cues were visible before the encoding and retrieval stages for each sequence. We then randomised the order of stimulus lists across retrieval conditions, so that participants did not know whether the sequence involved articulatory suppression until encoding was complete. In addition, we prolonged presentation time of each object during encoding to 2000ms in order to give participants more time to process the object. Finally, we changed the “wait” screen so that rather than passively looking at the screen, participants had to click on 4 dots appearing in 4 corners on the screen in a random order to “calibrate the mouse”. This dot-clicking subtask was to eliminate covert rehearsal in the no-suppression condition at encoding (which would give an advantage compared to the suppression condition, where verbal rehearsal was not possible), as well as to reduce the possibility that participants were using visualising strategies and focusing on specific perceptual features from the presented image (e.g. a round thing) instead of relying on memory for the holistic object concept (e.g., a *tomato*).

Design and Analysis

As per Experiment 1.

Results

Based on pre-registered exclusion criteria, no trials were excluded for accuracy analysis. For analysis of correct RTs, no trials were removed as motor errors, but 31 trials (0.012% of data) were removed as RTs were more than 3 SDs above the individual participant’s mean.

Confirmatory Analysis

Accuracy. Bayesian model comparison showed strong evidence *against* Step 2 over Step 1, $BF_{10} = 0.02$; that is, the data were 57 times more likely under the Step 1 model containing only random effects than a model containing articulatory suppression at encoding and retrieval. There was also strong evidence at Step 3 *against* the effect of the encoding*retrieval interaction on accuracy, $BF_{10} = 0.03$: that is, similar to Experiment 1, the data were 40 times more likely under the Step 2 model without the interaction than the Step 3 model with the interaction.

We then used the coefficients in the Step 3 model (Table 3) to estimate the marginal accuracy for each condition of encoding × retrieval articulatory suppression (see Figure 3). Both encoding and retrieval parameters had negative coefficients but – unlike the retrieval effect in Experiment 1 – only encoding had a significant effect in NHST terms, indicating that articulatory suppression during encoding impaired performance accuracy. Accuracy was best in the no-suppression/no-suppression condition (i.e., no articulatory suppression at either encoding or retrieval), with participants correctly recognising on average 11 (SE = 0.2) out of 12 objects per sequence, and was worst in the suppression/suppression condition (9.9 objects remembered, SE = 0.4) rather than in the no-suppression/suppression condition as we had hypothesised. That is, object memory was least accurate when access to language was blocked at both encoding and retrieval, which was in line with the results of Experiment 1.

Table 3: Experiment 2 unstandardized regression coefficients, standard errors, and associated statistics from Step 3 models of Accuracy (logistic mixed-effect regression) and RT (linear mixed-effect regression), for articulatory suppression effects at encoding, retrieval, and their interaction.

DV	Parameter	Coefficient	SE	*df*	*z*	*p*
Accuracy	Intercept	2.437	0.252	-	9.678	<.001
	Encoding	-0.824	0.294	-	-2.808	.005
	Retrieval	-0.220	0.159	-	-1.388	.165
	Encoding*Retrieval	0.171	0.205	-	0.835	.404
					t	
RT	Intercept	2786.07	139.23	30.08	20.011	<.001
	Encoding	68.73	155.66	51.22	0.442	.661
	Retrieval	-150.49	60.87	2429.19	-2.472	.014
	Encoding*Retrieval	-29.11	86.48	2430.96	-0.337	.736

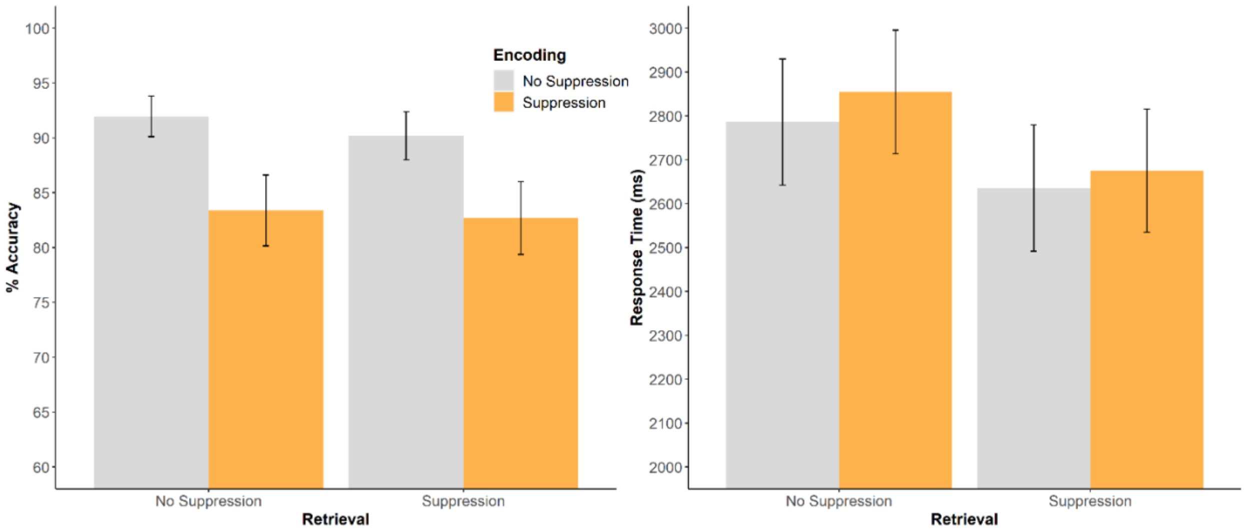


Figure 3: Mean % accuracy and RT per articulatory suppression condition in Experiment 2, calculated as marginal means from the Step 3 models. Error bars represent ± 1 Standard Error.

Response Times. Bayesian model comparison showed equivocal evidence against Step 2 over Step 1, $BF_{10} = 0.61$, that is, the RT data were 1.65 times more likely under a model with only random effects than a model that contained fixed effects of articulatory suppression at encoding and retrieval. As in Experiment 1, there was strong evidence at Step 3 *against* the presence of an encoding-retrieval interaction, $BF_{10} = 0.02$; that is, data were 47 times more likely under the non-interaction model than the interaction model (see Table 3).

As before, we used the coefficients in the Step 3 model (Table 3) to estimate the marginal mean RT for each articulatory suppression condition (see Figure 3). As in Experiment 1, the retrieval coefficient was negative and significant in NHST terms (i.e., unexpectedly faster under articulatory suppression), but this time articulatory suppression at encoding had no effect. Against our predictions, but in line with Experiment 1, recognition of target objects was best (fastest) in the no-suppression/suppression condition (i.e., when language was available at encoding but not at retrieval), and worst (slowest) in the suppression/no-suppression condition (i.e., when language was available at retrieval but not at encoding). That is, participants had most difficulty recognising remembered objects when language was blocked at the point of encoding only but was available at retrieval.

Exploratory Analysis

Because our confirmatory analysis again produced some unexpected results, we ran exploratory analyses to determine the best-fitting model for accuracy and RT. As in Experiment 1, we first explored alternative random effects structures for the null model using Restricted Maximum Likelihood (i.e., random encoding and/or retrieval slopes for participants and/or items) and selected the best model using Bayes Factors calculated via BIC. We then examined (using Maximum Likelihood) whether the data favoured an encoding-only or retrieval-only models in comparison to the null model. As before, we

Bonferroni-corrected the p-values on parameter coefficients by multiplying by 3. Model comparisons results are presented in Table 4.

Table 4: Exploratory analysis of Experiment 2, showing Bayes Factor comparison of each candidate model against the null model (random effects only).

DV	Candidate model	BF_{10}
Accuracy	Encoding only	0.50
	Retrieval only	0.03
	Encoding+Retrieval	0.02
RT	Encoding only	0.02
	Retrieval only	27.66
	Encoding+Retrieval	0.61

Best-Fitting Model of Accuracy. Attempts to model random slopes led to non-convergence in all models (see supplementary materials for full results), as in Experiment 1's exploratory analyses. As a result, we again used models without random slopes (i.e., random intercepts only for participants and items, as per confirmatory analysis) as the null model in explorations of fixed effects on accuracy.

Bayesian model comparisons showed that the evidence similarly favoured both the null model (i.e., random effects only) and the model with articulatory suppression at encoding only (i.e., equivocal evidence for the null), but strongly disfavoured models containing only retrieval and both encoding and retrieval. The encoding-only model was BF_{10} = 16.67 times better than the next-best alternative model with retrieval only, and hence represented a best-fitting model of accuracy alongside the null. In the encoding-only model, articulatory suppression at encoding led to lower accuracy (b = -0.735, *SE* = 0.273, z = -2.691, *p* =.007). That is, as predicted, removing access to language impaired object memory accuracy: people were 109% more likely (i.e., more than twice as likely) to make an error during retrieval if their access to labels had been blocked during encoding. However, although this effect was significant in NHST terms, evidence for the encoding model was very weak in Bayesian terms (i.e., equivocal evidence), and so we treat the effect with caution.

Best-Fitting Model of Response Times. Exploration of random slope structures showed that the best fit emerged from the model without random slopes, as no slope model met the $BF_{10} > 5$ threshold for improving model fit. Therefore, we report models with no slopes. Full statistics can be found in supplemental materials.

In explorations of fixed effects, Bayesian model comparisons showed that, in contrast to the results for accuracy, the data were most likely under a retrieval-only model (positive evidence), followed by a model containing both encoding and retrieval (equivocal evidence), and lastly a model containing only encoding (evidence favoured the null). The retrieval-only model was $BF_{10} = 45.34$ times better than the next-best alternative, and hence represented the best-fitting model of RT. In this model, articulatory suppression at retrieval had a negative effect on RT ($b = -164.89$, $SE = 43.26$, $t(2430) = -3.812$, $p = .001$), indicating that when language access was blocked at retrieval, participants were 165ms *faster* to respond than when they had access to language, which was consistent with the results of Experiment 1. Closer examination of RT and accuracy effects suggested that, in contrast with Experiment 1, this time the retrieval effect was not entirely due to a speed-accuracy tradeoff: when participants were asked to perform articulatory suppression at retrieval, response times were faster, but there was no accompanying drop in accuracy. We discuss this point more below.

Discussion

Experiment 2 examined the role of language in working memory performance having made some methodological improvements in comparison to Experiment 1; in particular, using longer sequences of objects to avoid ceiling effects. We found that articulatory suppression at encoding weakly impaired accuracy of recognition memory, but not speed of response. This encoding effect on accuracy appeared only in NHST coefficient statistics during confirmatory analysis, and Bayesian evidence was equivocal in exploratory analysis of fixed effects; we therefore interpret it cautiously as a somewhat weak effect. When language was blocked at encoding, people's ability to remember objects was impaired to some extent.

We also found that articulatory suppression at retrieval led to *faster* response times in both confirmatory analysis and exploration of best-fitting models (as per Experiment 1), but this time had no effect on accuracy. It is possible that one of the ways people might use language to support object memory at the retrieval stage is by implicitly naming the objects presented in the array, which could introduce a processing delay compared to when language is blocked at retrieval and implicit naming cannot take place (see Phillips et al., 1999, for a similar articulatory suppression finding). We therefore conclude that blocking language access at the point of retrieval does not lead to slower response time (as we originally hypothesised), but instead speeds up response times and tends to increase errors in a pattern consistent with a speed-accuracy tradeoff. Such effects are unlikely to be due to ceiling effects on sequence length or knowledge of the upcoming retrieval condition, because both these issues were addressed in Experient 2's methodology, and may instead simply reflect that performing articulatory suppression during target selection led participants to prioritise response speed over accuracy. We return to this point in the general discussion.

Finally, and importantly, we did not find an interaction of encoding*retrieval on either accuracy or RT. This result was consistent with the findings of Experiment 1, but did not support our hypothesis that people would find it most difficult to remember objects when they had language available during encoding that was later blocked at retrieval. Rather, in such circumstances, it appears that participants were still able to identify targets fairly successfully, potentially on the basis of encoded sensorimotor information.

Based on the results of Experiment 2, we calculated that a sequence of 9.9 object concepts on average can be held in working memory when relying on sensorimotor information only (suppression/suppression condition), but that this capacity increases to 11.0 objects when linguistic labels are available (no-suppression/no-suppression condition). This finding is in line with the hypothesis that working memory capacity is effectively greater when language is available to act as a placeholder for a full sensorimotor representation. These capacity estimates are double those of Experiment 1, indicating that the sequences of 6 objects employed in Experiment 1 were indeed likely to have been subject to ceiling effects on performance. Moreover, the capacity estimates of 10-11 objects are far greater than those previously estimated for the episodic buffer using more artificial stimuli (Allen et al., 2015; Langerock et al., 2014), which suggests that working memory for contextually situated, real-world objects can benefit greatly from conceptual support from long-term memory.

Overall, Experiment 2 partially supported our hypotheses regarding the role of linguistic bootstrapping in working memory for object concepts. In an ostensibly nonverbal recognition memory task, blocking access to linguistic labels via articulatory suppression at the point of encoding led people to remember fewer objects compared to when language was available. The ability of articulatory suppression to impair memory in this experiment (using sequences of 12 objects) when it did not in Experiment 1 (using sequences of 6 objects) suggests that language is most useful in supporting working memory when where are *many* objects to be remembered. That is, consistent with the linguistic bootstrapping hypothesis, words can act as linguistic placeholders when there are insufficient resources to hold all objects in mind as sensorimotor simulations.

Working memory literature shows that performing an articulatory suppression task blocks phonological processing, thus affecting working memory for language stimuli (e.g., numbers, letters, words), but does not affect executive processes in working memory (e.g., Jaroslawska et al., 2018; Murray, 1967). Nonetheless, it could be argued that our findings instead reflect dual task performance, rather than articulatory suppression specifically removing access to the linguistic labels that normally support cognitive processing. We address this possibility in the next experiment.

Experiment 3: Foot-Tapping as Dual-Task Control

In our third experiment (pre-registration, raw data, analysis code, and stimuli are available as supplemental materials on OSF), we had two goals: to replicate the effect of articulatory suppression at encoding that emerged in Experiment 2 in support of the linguistic bootstrapping hypothesis, and to test whether this effect is specifically due to blocking access to language rather than due to performing a secondary task. We therefore compared articulatory suppression during encoding to foot tapping, a secondary task that is unrelated to language use but comparable on a number of other characteristics (i.e., it is a rhythmic motor task, does not involve visual perception or hand action that could interfere with stimulus presentation, and it can be sustained throughout the encoding stage without undue fatigue: see Gaillard et al., 2012; van't Wout & Jarrold, 2020). This time, we also manipulated the secondary task at encoding *within* participants, so that all participants encoded sequences in all three conditions (with no secondary task, while performing articulatory suppression, and while performing foot tapping).

As in previous experiments, we predicted that accuracy and latency of performance would be impaired when access to linguistic labels is blocked via articulatory suppression compared to when language is available (i.e., no secondary task). In addition, we hypothesised that this impairment would not be solely due to performing a secondary task, and that blocking access to language via articulatory suppression would lead to greater impairment of object memory than a secondary task of foot-tapping that does not affect access to language.

Method

Participants

Eighteen native speakers of English (18 female; mean age = 18.6 years, *SD* = 0.8 years) were recruited from Lancaster University, and received course credit or a payment of £3.50 for participation. Data from two originally-recruited participants were replaced due to not being native speakers of English. As pre-registered, we used Bayesian sequential hypothesis testing to determine sample size, and stopped at the minimum sample size Nmin = 18 when our Step 3 models for both RT and accuracy cleared the specified threshold of evidence $BF_{10} > 5$ or its reciprocal $BF_{10} < 0.2$ for three successive participants (see Design and Analysis section for model details; full statistics are reported in the Results section).

Materials

We used the same materials as in Experiment 2 with the following changes. We created 3 additional sequences of 12 objects apiece, bringing the total number of target objects to 108, divided into nine sequences; this number allowed for three sequences to be tested per secondary task condition. As per the existing sequences, each new sequence represented an ecologically-valid order of objects that would be plausibly used in a real-world setting, and were labelled with a brief description that provided a naturalistic, situated context. We also altered 3 target items and 1 distractor item in the existing sequences, to avoid duplicating targets from the new sequences. The order of objects for new sequences was determined by 8 volunteers who did not take part in the study, and was established based on their mean rank as in Experiment 1.

Photographic images of new target and distractor objects were sourced and edited as per Experiment 1. In total, the present experiment utilized 756 object images: 108 target objects presented at encoding, 108 target objects presented at retrieval (i.e., different images to the encoding stage), and 540 distractor objects presented at retrieval.

Procedure

As in the previous experiments, participants were tested individually and consented to publicly share their anonymised data. They sat in front of a computer and were informed that they would be asked to perform a secondary task at some point during the experiment. The experimenter then presented them with three symbols which would be used to signal what they should do for each task condition, and then explained what the tasks involved. A picture of a mouth (the same as in Experiment 2) was used to indicate that participants should repeat the word "the"; a picture of a foot indicated that participants should tap their foot continuously, and a picture of an X was used to indicate that they should stop or not perform either of the tasks (i.e., to indicate the control condition of no secondary task). The experimenter demonstrated both secondary tasks, where articulations of "the" and foot taps were repeated at approximately the same rhythmic rate, and asked the participant to practice them. Once the participant confirmed that they understood and could perform the tasks correctly to the experimenter's satisfaction, they provided demographic information and read the instructions onscreen.

The secondary task was manipulated within-participants at encoding; that is, participants took part in each of the three secondary task conditions, where the order of conditions was rotated in a latin-square design. The object sequences were divided into three sets of three sequences apiece, and the assignment of each set to a secondary task condition was counterbalanced across participants. Within each condition, sequences were presented in a randomised order, and each sequence appeared in each condition an equal number of times.

As before, participants were instructed that they would see a sequence of everyday objects appear one-by-one onscreen, and their task was to remember the objects; later, they would see groups of objects onscreen and they should click on the object that belonged to the sequence they had been asked to remember. Participants first commenced a practice sequence without any secondary task at encoding or retrieval, which provided time-related feedback to habituate them to responding within a time limit ("good job" was displayed on screen if the response was given on time, and "too slow" if they failed to respond within 6000 ms). When the participant confirmed that they understood the task and were happy to continue, they commenced the experimental trials. After the brief description of the sequence, the image indicating the secondary task condition was displayed for 3000ms, and then the 12 objects were presented one by one. The break after encoding was the same as in Experiment 2 (clicking on 4 dots on the screen). Participants were then asked to stop the secondary task (i.e., participants did not perform any secondary task at retrieval). There was no feedback on experimental trials, but the trial timed out and was marked as incorrect if participants failed to respond within 6000ms.

Design and Analysis

We analysed accuracy (incorrect = 0, correct = 1) in a mixed-effects hierarchical logistic regression (binomial, logit link), and response times (RT) for correct responses in a mixed-effects hierarchical linear regression. For both accuracy and RT analysis, participants and items (nested within sequence) were included as crossed random effects. Fixed effects were dummy coded using articulatory suppression as the reference level, which allowed us to test each critical hypothesis (i.e., that articulatory suppression would be worse than no task *and* worse than foot tapping) with a distinct parameter. That is, we included two encoding variables as fixed effects: no task (representing no-task vs. a secondary task at encoding: 1 = no task, 0 = foot tapping or articulatory suppression) and foot tapping (distinguishing foot tapping as a secondary task: 1 = foot tapping, 0 = no task control or articulatory suppression).

In regressions of both accuracy and RT, Step 1 entered random effects, Step 2 added no-task as a fixed effect, and Step 3 added foot-tapping as a fixed effect. We ran Bayesian model comparisons between steps, with Bayes Factors (BF) calculated via BIC as in the previous two experiments. We also report null hypothesis significance testing (NHST) statistics for parameter coefficients in the Step 3 model. Specifically, the no-task coefficient in Step 3 allowed us to test the hypothesis that articulatory suppression produced a greater impairment than no task (i.e., replicating Experiment 2), and the Step 2-3 model comparison along with the foot-tapping coefficient in Step 3 allowed us to test whether articulatory suppression produced a greater impairment than foot tapping (i.e., the critical new hypothesis of the present experiment).

We used the Step 3 coefficients to calculate the marginal mean accuracy and RT per secondary task condition, and used the mean accuracy to obtain an estimated range of working memory capacity.

Results

No trials were excluded for the accuracy analysis. For analysis of correct RTs, 8 trials (0.006% of data) were removed as motor errors or for being more than 3 SDs above the individual participant's mean. All reported results relate to confirmatory analysis.[1]

Accuracy

Bayesian model comparison showed very strong evidence for Step 2 over Step 1, indicating the data were BF_{10} = 2321.57 times more likely under a model that separated a no-task control from some form of secondary task. As predicted, there was also very strong evidence for Step 3 over Step 2, BF_{10} = 1556.20, meaning that the data favoured a model that distinguished between articulatory suppression and foot tapping as secondary tasks.

We then used the coefficients in the Step 3 model (Table 5) to estimate the mean marginal accuracy for each secondary task (see Figure 4). Critically, and as predicted,

[1] As in previous experiments, we attempted to explore different random effects structures in order to select the best-fitting model of accuracy and RT. However, none of the candidate structures involving random slopes improved model fit over the random intercepts of confirmatory analysis, and so we do not report them further. Full details are in supplementary materials.

performance was worst in the articulatory suppression (reference) condition, with participants correctly remembering an average of 8.1 (*SE* = 0.6) out of 12 objects per sequence. Performance was better during foot tapping, where participants were 89% more likely to respond correctly compared to articulatory suppression, and correctly recognized an average of 9.6 (*SE* = 0.5) objects. Finally, participants performed best when there was no secondary task, and, with an average of 10.0 (*SE* = 0.4) objects recognized per sequence, were 146% more likely (i.e., more than twice as likely) to respond correctly compared to when performing articulatory suppression. Overall, as expected, memory performance was impaired when access to language was blocked (replicating Experiment 2), and blocking access to language via articulatory suppression impaired performance more than a secondary task that was unrelated to language (foot tapping).

Table 5: Experiment 3 unstandardized regression coefficients, standard errors, and associated statistics from Step 3 models of Accuracy (logistic mixed-effect regression) and RT (linear mixed-effect regression), for effects of secondary task at encoding with articulatory suppression as the reference level.

DV	Parameter	Coefficient	SE	*df*	*z*	*p*
Accuracy	Intercept	0.736	0.234	-	3.151	<.01
	No task (control)	0.900	0.140	-	6.452	<.001
	Foot tapping	0.637	0.135	-	4.732	<.001
					t	
RT	Intercept	2439.76	87.60	29.55	27.850	<.001
	No task (control)	10.40	48.99	1314.24	0.212	.832
	Foot tapping	-17.89	49.57	1312.81	-0.361	.718

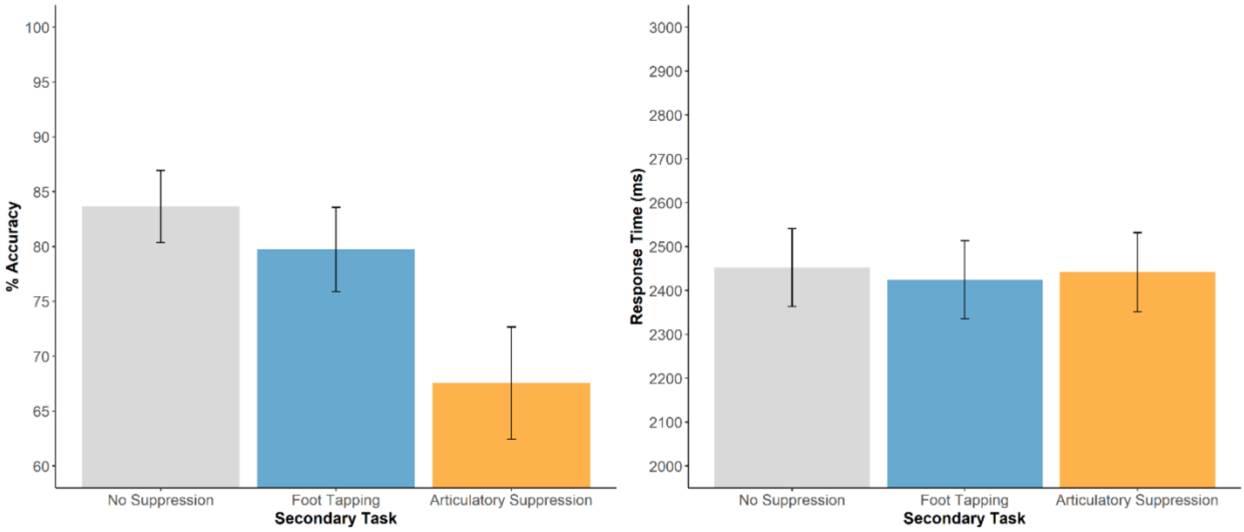


Figure 4: Mean % accuracy and RT per secondary task condition in Experiment 3, calculated as marginal means at Step 3 models. Error bars represent ± 1 Standard Error.

Response Times

Bayesian model comparison showed strong evidence *against* the Step 2 model over Step 1 (BF_{10} = 0.03); that is, the RT data were 33.3 times more likely under a model containing only random effects than a model that distinguished secondary tasks from the no-task control condition. Similarly, there was strong evidence at Step 3 *against* distinguishing between foot tapping and articulatory suppression as secondary tasks, BF_{10} = 0.03, whereby the data were 37.0 times more likely under the Step 2 model than the Step 3 model.

Nonetheless, we used the coefficients of the Step 3 model (Table 5) to estimate the marginal means for each secondary task condition (Figure 4). RT was similar in all conditions, and coefficients indicated no reliable differences. That is, against our predictions but consistent with previous experiments, participants were equally fast to select the correct object during the retrieval stage regardless of which secondary task (if any) was performed during encoding.

Discussion

Experiment 3 determined whether or not the effects of articulatory suppression at encoding that were observed in earlier experiments could be attributed to a dual-task load rather than to blocking access to linguistic labels. The results replicated Experiment 2 in showing that blocking access to language while encoding a sequence of objects impaired accuracy – but not speed – of memory performance. Critically, comparison with a foot-tapping task indicated that this effect was not a mere artifact of a dual-task paradigm. Rather, memory was adversely affected specifically by articulatory suppression, supporting our hypothesis that holding object concepts in working memory normally relies on language (i.e., implicitly-retrieved object labels, used as linguistic placeholders), and that blocking access to language reduces accuracy and capacity of working memory for objects. In other words, access to language enables linguistic bootstrapping, whereby people can use linguistic labels as placeholders for complex sensorimotor representations to increase working memory capacity, and can hence remember a greater number of objects when language is available compared to when it is not.

However, performing a secondary task at encoding had no effect on RT measured at retrieval, and there was also no difference in RT between the two secondary tasks of articulatory suppression and foot tapping. That is, objects which were successfully remembered were processed with the same difficulty regardless of what concurrent task participants performed while encoding them, consistent with the findings of Experiment 2. This lack of effects on RT suggests that – whether an object is held in working memory via its linguistic label, a sensorimotor simulation, or a combination of both – it is relatively easy to match this remembered information with a target onscreen.

In this experiment, we calculated that people can successfully remember on average 8.1 (SE = 0.6) objects when language is blocked, and 10.0 (SE = 0.5) objects when language is available. This capacity is slightly smaller than the estimate in Experiment 2, and we discuss differences between the experiments further in the general discussion. However, the within-participant manipulation of the encoding conditions in the present experiment means that the effect has a much more robust grade of evidence than in Experiment 2, and allows us to conclude that the presence of language allows an additional 2 concepts (approximately) to be held in working memory, which is consistent with the linguistic bootstrapping hypothesis.

Nonetheless, the possibility remains that in both Experiments 2 and 3, the sequence of 12 objects may have been subject to ceiling effects in the no-task control condition. For example, in the current experiment, 65% of participants reached 100% performance on at least one sequence, and 47% of the time participants scored 10 or more on a sequence. It is possible that at least some participants may be capable of remembering more than 12 objects when language is available to support encoding. Experiment 4 therefore addressed this possibility by using a range of sequence lengths to determine the upper limits of working memory capacity when language was fully available.

Experiment 4: Upper Capacity Limit with Language Available

In our final study (pre-registration, data, analysis code, and full results are available as supplemental materials on OSF), we wanted to establish the upper capacity limit of working memory for object concepts; that is, the number of objects that can be remembered when language is fully available and linguistic placeholders may be used to their full extent. Using a similar paradigm to previous experiments but with no secondary task, we asked participants to remember sequences that varied in length between 8 and 14 objects.

We hypothesised that accuracy of recognition memory would start to drop when the number of object concepts in a sequence reaches the maximum capacity for linguistic information. That is, people would remember shorter sequences relatively easily because their representations (sensorimotor simulation and/or linguistic labels) fit within the capacity of working memory. Even some longer sequences may still fit in working memory by the use of linguistic placeholders, where objects are represented via their labels only rather than via sensorimotor simulation. However, once the length of the sequence exceeds the capacity of working memory to hold concepts represented via linguistic labels, participants will not be able to retain them all and accuracy will suffer. As sequence length increases, the tipping point at which accuracy starts to reliably drop would reflect the limit at which capacity has been exceeded.

We also predicted that response times would slow down as working memory capacity is strained and linguistic placeholders are increasingly employed, so that people would be slower to recognise remembered objects when the number of objects in a sequence exceeds the capacity of working memory.

Method

Participants

Twenty native speakers of English took part in the study (17 female; mean age = 18.95; *SD* = 1.07). Data from one participant was replaced due to not being a native speaker of English. As before, the sample size was determined using sequential hypothesis testing with Bayes Factors. We stopped at the sample size *N* = 20 when the Step 4 models for accuracy cleared the specified threshold of evidence $BF_{10} > 5$ for five consecutive participants. (See Design & Analysis section for model details; full statistics are reported in Results).

Materials

Test items comprised 112 target objects, divided into 8 sequences of 14 items each. These sequences were based on materials from Experiment 3, which we extended from 12 to 14 items by adding two extra objects to each sequence. Four of the extra target objects, with their accompanying distractors, were taken from the ninth sequence of Experiment 3 unused in this experiment; two were objects and accompanying distractors used in Experiment 2 that were not used in Experiment 3, while 10 were new objects. Distractor items for the new target objects were selected using the same criteria as Experiment 1. As before, the order of objects for new sequences was determined by 8 volunteers who did not take part in the study, and was established based on their mean rank. We then created subsets of items within each sequence using the first 8, 10, 12 or all 14 items, so that each sequence could be plausibly represented in different lengths while still being ecologically valid (e.g., the context situation of making a cake still applied regardless of whether strawberries or whipped cream were included in the sequence). This subsetting approach allowed us to compare sequences of different lengths without confounding context situation with sequence length.

Photographic images of new target and distractor objects were selected and edited as per Experiment 1, leading to a total of 784 object images: 112 target objects for presentation at encoding, 112 target objects (different images) for presentation at retrieval, and 560 distractor objects for presentation at retrieval.

Procedure

The procedure was the same as in the no-task condition of Experiment 3. After a practice sequence of 8 items, participants completed all eight test sequences in a fixed order of increasing length (i.e., two sequences of 8 objects each, then two sequences of 10 objects each, and so on). We rotated sequences across length conditions so that, across the experiment as a whole, each sequence was presented in each of its possible subsets (i.e., 8, 10, 12, and 14 objects) and therefore in different ordinal positions in the procedure.

Design and Analysis

We analysed accuracy (incorrect = 0, correct = 1) in a mixed-effects hierarchical logistic regression (binomial, logit link), and response times (RT) for correct responses in a mixed-effects hierarchical linear regression. In both analyses, we included participants and items (nested within sequence) as crossed random effects. Sequence length was included as a categorical fixed effect, coded using reverse Helmert coding to compare the effect of each sequence length with the mean of the previous (shorter) sequences, which resulted in 3 coded variables (10 vs. 8 objects, 12 vs. 8-10 objects, 14 vs. 8-10-12 objects). This coding method allowed us to determine the tipping point at which accuracy dropped (or RT slowed down) due to the sequence length surpassing the capacity of working memory.

Hierarchical regressions comprised the following steps: Step 1 entered random effects, Step 2 entered sequence length as 10 vs. 8 objects, Step 3 entered sequence length as 12 vs. 8-10 objects, and Step 4 entered sequence length as 14 vs. 8-10-12 objects. We ran Bayesian model comparisons between steps, with Bayes Factors (BF) calculated via BIC as in earlier experiments. Specifically, the first step to show an improvement in model fit represented the sequence length at which memory performance differed from that of shorter sequences. In addition, the parameter of the first sequence length variable to produce an accuracy effect allowed us to estimate the capacity of working memory for objects when language is fully available to support encoding.

Results

No trials were excluded for the accuracy analysis. For analysis of correct RTs, 14 trials were removed as outliers more than 3 standard deviations from the individual participant's mean (total 0.0096% of data removed). All reported results relate to confirmatory analysis, as the attempt to model random slopes of sequence length on items led to non-convergence in both accuracy and RT analysis (see supplementary materials).

Accuracy

Bayesian model comparison showed evidence *against* any effect of sequence length on accuracy while sequence length remained at 12 objects or fewer: Step 2 did not improve model fit over Step 1 (BF_{10} = 0.02), and Step 3 did not improve model fit over Step 2 (BF_{10} = 0.20). However, there was strong evidence for Step 4 over Step 3 (BF_{10} = 14.95), meaning that the data favoured a model that distinguished 14-object sequences from shorter sequences.

We then used the coefficients in the Step 4 model (Table 6) to estimate the marginal accuracy for each sequence length parameter (see Figure 5). When sequence length reached 14 objects, accuracy decreased compared to shorter sequences, that is, when participants were asked to remember a sequence of 14 objects, they were 76% more likely to make an error in responding than for sequences of 8-12 objects, suggesting that 14 objects exceeded the capacity of working memory. Participants successfully remembered an average of 11.9 (*SE* = 0.5) out of 14 objects.

We noted that in the Step 4 model, the parameter for sequence length of 12 (vs. 8-10 objects) was also significant in NHST terms, reflecting a small drop in accuracy between 8-10 objects and 12 objects. However, since Bayesian model comparisons found evidence *against* the addition of this parameter in Step 3, it indicates that the data were more likely under a model that ignored the distinction between sequence length 12 and sequence lengths 8-10 than under a model that distinguished them. We therefore treat the NHST effect for sequence length 12 with caution, whereas both Bayesian model comparison and NHST coefficient statistics supported a tipping point at sequences of 14 objects. That is, sequences of 14 objects exceeded the capacity of working memory in a way that sequences of 12 objects did not robustly do so, hence the drop in accuracy, and we estimate the capacity limit of working memory for object concepts to be approximately 11.9.

Table 6: Experiment 4 unstandardized regression coefficients, standard errors, and associated statistics from Step 4 models of Accuracy (logistic mixed-effect regression) and RT (linear mixed-effect regression), for Helmert-coded effects of sequence length.

DV	Parameter	Coefficient	SE	*df*	*z*	*p*
Accuracy	Intercept	2.160	0.258	-	8.382	<.001
	10 vs. 8 objects	-0.060	0.241	-	-0.249	.804
	12 vs. 8-10 objects	-0.473	0.179	-	-2.635	.008
	14 vs. 8-10-12 objects	-0.563	0.151	-	-3.725	<.001
					t	
RT	Intercept	2271.09	76.43	24.68	29.715	<.001
	10 vs. 8 objects	100.73	53.23	1346.28	1.892	.059
	12 vs. 8-10 objects	54.51	43.11	1373.06	1.264	.206
	14 vs. 8-10-12 objects	13.82	39.06	1414.61	0.354	.724

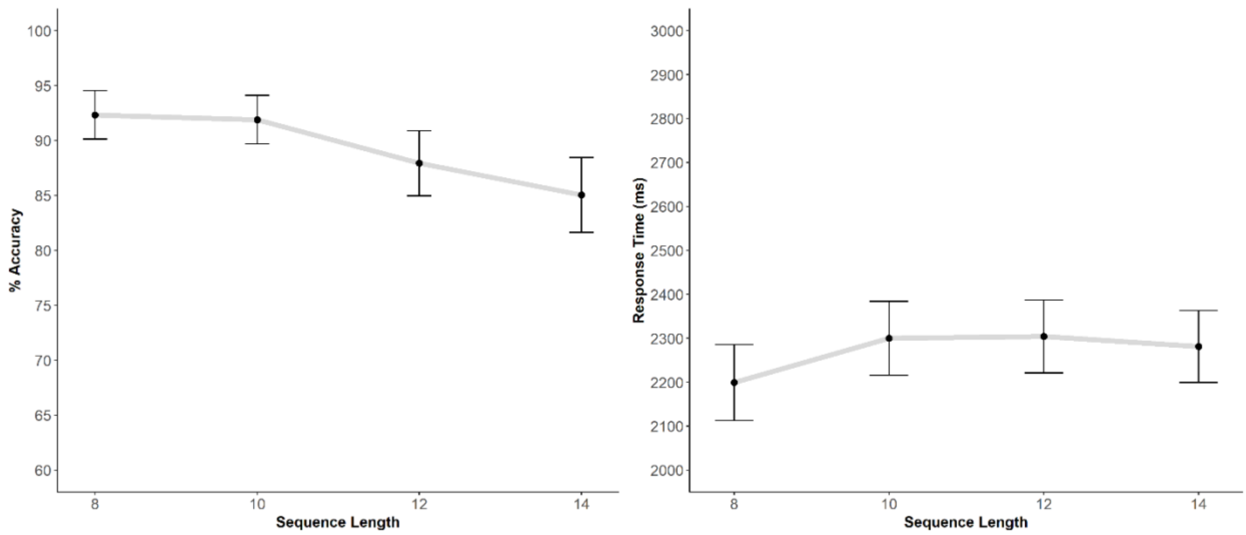


Figure 5: Mean % accuracy and RT for each sequence length condition in Experiment 4, based on marginal means from the Step 4 model. Error bars represent ± 1 Standard Error.

Response Times

Bayesian model comparison showed no effect of sequence length on RT: there was evidence *against* Step 2 over Step 1 (BF_{10} = 0.14), against Step 3 over Step 2 (BF_{10} = 0.05), and against Step 4 over Step 3 (BF_{10} = 0.03). Nonetheless, for the sake of complete reporting, we used the coefficients in the Step 4 model (Table 6) to estimate the marginal mean RT for each sequence length parameter (see Figure 5). RT was similar in all sequence length conditions. That is, against our expectations, people were equally fast to recognize objects regardless of how many objects were being remembered.

Discussion

Experiment 4 aimed to establish the upper limit of working memory capacity for object concepts when language is fully available (i.e., with no articulatory suppression), using longer sequences of objects than previous experiments in this study. We found that sequences of 14 objects caused a drop in accuracy relative to sequences of 8-12 objects, which suggested that 14 objects exceeded working memory capacity in a way that 12 objects did not. On average, participants successfully remembered 11.9 out of 14 objects. That is, when language is available, participants can accurately remember up to a capacity limit of approximately 12 object concepts, which, according to the linguistic bootstrapping hypothesis, is possible because a linguistic label can serve as a placeholder for a full sensorimotor representation of an object when working memory capacity is under strain.

We also found that sequence length had no effect on RT, suggesting that the time required to match an object representation in working memory to a visual stimulus onscreen was not influenced by demands on working memory capacity. The evidence against any effect on RT also suggests that the time required was not influenced by the format of object representation in working memory; that is, people could identify the onscreen target relatively quickly whether an object was represented via sensorimotor simulation (as would be possible for short sequences of 8 objects) or via linguistic labels (as most likely for long sequences of 12-14 objects). We discuss the possible processes involved in the general discussion.

General Discussion

The present study is the first to examine the linguistic bootstrapping hypothesis in the context of working memory; that is, whether word labels can act as placeholders for real-world object concepts when there are insufficient representational resources to maintain a sensorimotor simulation in full (Connell & Lynott, 2014), and thereby allow language to increase the available capacity of working memory. We tested this hypothesis in a series of pre-registered experiments using a nonverbal recognition memory task that asked participants to remember a sequence of pictured objects and then tested their ability to select each remembered object from a distractor array. As predicted, we found that blocking access to language at encoding via articulatory suppression resulted in poorer accuracy in object recognition memory and lower working memory capacity for objects. Participants could remember 8 (Experiment 3) to 10 objects concepts (Experiment 2) when relying on sensorimotor information only (i.e., when language was blocked), but the capacity increased by approximately two items (Experiments 3) to an upper limit of 12 objects (Experiment 4) when linguistic labels were available to act as placeholders and ease the strain on working memory. Critically, this effect was not an artifact of dual task performance, as blocking access to language via articulatory suppression impaired accuracy markedly more than an alternative secondary task of foot tapping that left access to language intact (Experiment 3). This pattern of findings overall supported the linguistic bootstrapping hypothesis that, even in nonverbal paradigms, storage of object concepts in working memory normally relies on language (i.e., implicitly-retrieved object labels). When studying a long sequence of objects for later retrieval, people can drop the sensorimotor representations of the objects from working memory and allow their linguistic labels to deputise as placeholders, in order to maximise the number of objects that can be held in mind. Blocking access to language during this process therefore results in fewer objects being remembered.

However, there was no comparable effect on RT, where participants were just as quick to select a remembered object from a distractor array regardless of whether or not language had been available during encoding (Experiments 2-3). There are two possibilities regarding what happens to sensorimotor representations of objects when the number of objects exceeds working memory capacity *and* language is unavailable to provide linguistic placeholders (i.e., the articulatory suppression condition). Since sensorimotor simulations are flexible and responsive to task demands (Connell & Lynott, 2014), we had originally expected that all sensorimotor representations in working memory would degrade to some extent (i.e., lose some detail, such as information from less-relevant perceptual modalities or action effectors) in an effort to maintain all objects in working memory. This possibility would have led both to greater errors and slower responses at the point of retrieval (compared to the no-task control that allowed linguistic placeholders), due to the difficulty of matching degraded object representations to target object pictures, but the lack of RT effects makes this possibility unlikely. The second possibility was that some objects are dropped entirely from working memory, but the sensorimotor representations of the remaining objects retain their original quality of detail. This possibility would have led to greater errors at retrieval (i.e., because some objects are now absent from working memory) but no effect on RT for those objects successfully recognised (i.e., because sensorimotor objects still in working memory can be easily matched to target object pictures). Results in Experiments 2-3 followed this pattern and therefore suggests that, when working memory capacity for object concepts is strained, sensorimotor representations of individual objects are lost rather than maintained in some degraded form.

In addition, articulatory suppression during the retrieval stage unexpectedly led to *faster* RT (Experiments 1-2) and also greater errors (Experiment 2). As discussed in Experiment 2, this speed-accuracy trade-off suggests that blocking language during retrieval may have led participants to prioritise speed over accuracy, though it is possible it may also reflect a combination of other phenomena (e.g., time saved by not implicitly labelling the object array, plus more errors due to losing linguistic labels from working memory). Regardless, this pattern of findings does not follow our original predictions, and suggests that recognition memory is flexible and robust enough that it can survive losing access to language between encoding and retrieval. If this supposition is correct, it could also explain the absence of interaction between articulatory suppression at encoding and retrieval. When language is not available during encoding, and a concept is represented in working memory via sensorimotor simulation alone, then during a later recognition memory task participants have two options: they can either directly compare their sensorimotor representation to what they see onscreen, or they can implicitly label the representation and the target stimulus and compare the two labels. On the other hand, when a concept is represented in working memory via a linguistic label alone, then in the later recognition memory task, participants have the same two options: they can implicitly label the object onscreen and compare the two labels, or they can retrieve a sensorimotor representation of the label's referent object and then compare that to what they see onscreen. In both cases, blocking access to language during the retrieval process leaves the sensorimotor option available, and so there is no interaction between the format of object representation in working memory and the availability of language during the recognition memory task. In other words, the pattern of findings is consistent with the idea that nonverbal object recognition memory normally relies on language (i.e., implicitly-retrieved object labels), and further suggests that linguistic bootstrapping allows to flexibly adapt the contents of working memory, by accessing linguistic or sensorimotor aspects of a concept's representation, depending on available resources.

It could be argued that the accuracy impairment from articulatory suppression at encoding in Experiments 2 and 3 could also be attributed to a dual-coding advantage (Paivio, 1971), whereby information encoded with both linguistic labels and images is more memorable than information encoded with only one of those modalities (Marschark & Paivio, 1977). If this argument were true, then such a dual coding advantage should have appeared regardless of the number of items in the sequence; in particular, the sequences of 6 objects in Experiment 1 would have shown an accuracy impairment when language was unavailable to dual-code the objects. However, this effect did not occur: articulatory suppression had no effect on participant's ability to remember sequences of 6 objects. The pattern of findings is consistent with linguistic bootstrapping, however, which holds that linguistic placeholders are employed when there are insufficient resources to maintain sensorimotor simulations in full. In Experiment 1, working memory capacity was not under strain and hence there was no need to replace the sensorimotor representation of objects with a linguistic placeholder; but in Experiment 2, trying to remember 12 objects *did* place working memory capacity under strain, and so linguistic placeholders became useful. Linguistic bootstrapping is increasingly employed as the number of objects to be remembered approaches (or exceeds) the capacity of working memory for sensorimotor representations of those objects.

The estimates of working memory capacity for real-world object concepts had a maximum of 12 objects when language was available at encoding (Experiment 4), and this capacity decreased by approximately 2 objects when language was unavailable (Experiment 3). The estimated language-available capacity was slightly lower in Experiments 2-3 (i.e., 10-11 out of 12 objects) when language was available to support linguistic bootstrapping, but since Experiment 4 showed that sequence length of 12 was likely subject to ceiling effects, we conclude that the estimate of 12 objects (out of 14) represents a more accurate upper limit. Additionally, in Experiment 3 the language-unavailable capacity was estimated to be 8 objects, whereas it was 10 objects in Experiment 2; as the relevant conditions are comparable (i.e., articulatory suppression condition in Experiment 3; suppression/no-suppression condition in Experiment 2), this difference is most likely due to inter-participant variability. Taking the within-participant estimate from Experiment 3 of the effect of losing access to language, we conclude that making language unavailable during encoding reduces working memory capacity by 2 objects. Based on these results, we estimate that working memory capacity for object concepts is 12 when language is available to bootstrap cognition, and 10 when it is not available. These capacity estimates are much higher than the traditional working memory capacity estimates of 4-7 objects (Miller, 1956; Allen et al., 2015), or the 3-6 items estimated as the capacity of the episodic buffer (Allen et al., 2015; Langerock et al., 2014). Previous studies focused on the episodic buffer as a storage of multimodal stimuli, but neglected its connection with long-term memory, beyond a few specific sets of information like the layout of a phone keyboard. However, the role of the episodic buffer is to support integrating stimulus information with existing knowledge from long-term-memory about objects and their inter-relationships. In the present experiments, we sought to exploit this ability of long-term memory to support working memory, in order to examine working memory for the kind of information used in real-life, everyday cognition. We therefore asked people to remember contextually-related sequences of objects, embedded in a familiar situation: for example, a realistic recipe for a cake, or equipment for a camping trip. The capacity estimates of 10-12 objects we report here therefore relate to the capacity of working memory for familiar object concepts when conceptually supported by long-term memory, and demonstrate that the episodic buffer may be a much bigger system than previously considered in its ability to support complex conceptual processing. It is possible that working memory capacity for concepts other than objects may differ from our current findings; for instance, recent work in our lab has found that working memory for sequences of action events is limited to approximately three events (Banks & Connell, 2021), and is even lower in the absence of language, which suggests that sensorimotor representations of events are larger (i.e., take up more space) in WM than sensorimotor representations of objects. Future work should examine in more detail how working memory capacity may differ according to the nature of the concept or entity being remembered.

In the present paper, we used real-world object sequences to examine working memory for naturalistic concepts that can draw upon supportive information in long-term memory, and found that language increases the number of familiar objects that people can remember at one time. In line with the linguistic-sensorimotor accounts of the conceptual system, and the linguistic bootstrapping hypothesis, we found that language is a critical part of how people represent, remember, and use their knowledge about concepts: when language is not available, memory capacity decreases on average from 12 to 10 objects. We also found that working memory capacity for familiar objects is much higher than previously estimated from research on artificial stimuli, because drawing on existing knowledge about object concepts is what enables people to implicitly retrieve labels and use them as linguistic placeholders (and, conversely, flesh them out again to sensorimotor representations when required). To date, while the nature of the conceptual system and working memory have both have been studied extensively, their theories of human cognition have tended to develop in parallel, answering different types of questions, and rendering them unable to explain the findings of each other. By bringing together research on linguistic-simulation theories and the episodic buffer of working memory, we hope to develop a more comprehensive understanding of human cognition.

References

Allen, R. J., Havelka, J., Falcon, T., Evans, S., & Darling, S. (2015). Modality specificity and integration in working memory: Insights from visuospatial bootstrapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 820–830. https://doi.org/10.1037/xlm0000058

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

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Sciences*, *4*(11), 417-423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*(3), 438-456. https://doi.org/10.1016/j.jml.2009.05.004

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Banks, B. & Connell, L. (2021). Language enhances working memory for action events. [Manuscript in preparation]. *Department of Psychology, Lancaster University.*

Banks, B., Wingfield, C., & Connell, L. (2020). Linguistic distributional knowledge and sensorimotor grounding both contribute to semantic category production. Retrieved from PsyArXiv database. https://doi:10.31234/osf.io/7eucw

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, and A. C. Graesser (Eds), *Symbols, Embodiment, and Meaning.* Oxford: Oxford University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bidet-Ildei, C., Meugnot, A., Beauprez, S. A., Gimenes, M., & Toussaint, L. (2017). Short-term upper limb immobilization affects action-word understanding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1129-1139. https://doi.org/10.1037/xlm0000373

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2018). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*, *29*, 120-153

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743-756. https://doi.org/10.1016/j.neuropsychologia.2007.10.007

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences*, *113*(27), 7459-7464. https://doi.org/10.1073/pnas.1520027113

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 157–165. https://doi.org/10.1037/0278-7393.18.1.157

Connell, L. (2018). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, *34*(10), 1308-1318. https://doi.org/10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin and Review*, *20*(3), 542–550. https://doi.org/10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–406. https://doi.org/10.1111/tops.12097

Connell, L., Lynott, D., & Dreyer, F. (2012). A functional role for modality-specific perceptual systems in conceptual representations. *PLoS ONE*, *7*(3), e33321. https://doi.org/10.1371/journal.pone.0033321

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922

Davis, C. P., Joergensen, G. H., Boddy, P., Dowling, C., & Yee, E. (2020). Making it harder to "see" meaning: The more you see something, the more its conceptual representation is susceptible to visual interference. *Psychological Science*, *31*(5), 505-517. https://doi.org/10.1177/0956797620910748

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved?. *Memory & Cognition*, *29*(2), 267-273. https://doi.org/10.3758/BF03194920

Dutriaux, L., Dahiez, X., & Gyselinck, V. (2018). How to change your memory of an object with a posture and a verb. *Quarterly Journal of Experimental Psychology*, *72*(5), 1112-1118. https://doi.org/10.1177/1747021818785096

Dymarska, A., Connell, L., & Banks, B. (2021). Linguistic bootstrapping in object memory [Data set]. Open Science Framework. http://osf.io/mwzfh

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211-245. https://doi.org/10.1037/0033-295X.102.2.211

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013). Where is the action? Action sentence processing in Parkinson's disease. *Neuropsychologia*, *51*(8), 1510-1517. https://doi.org/10.1016/j.neuropsychologia.2013.04.008

Gaillard, V., Destrebecqz, A., & Cleeremans, A. (2012). The influence of articulatory suppression on the control of implicit sequence knowledge. *Frontiers in Human Neuroscience*, *6*(208), 1-9. https://doi.org/10.3389/fnhum.2012.00208

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *Journal of Neuroscience, 26*(18), 4917-4921.

Goodhew, S. C., McGaw, B., & Kidd, E. (2014). Why is the sunny side always up? Explaining the spatial mapping of concepts by language use. *Psychonomic Bulletin & Review*, *21*(5), 1287-1293. https://doi.org/10.3758/s13423-014-0593-6

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Hitch, G. J., Brandimonte, M. A., & Walker, P. (1995). Two types of representation in visual memory: Evidence from the effects of stimulus contrast on image combination. *Memory & Cognition*, *23*(2), 147-154. https://doi.org/10.3758/BF03197217

Jaroslawska, A. J., Gathercole, S. E., & Holmes, J. (2018). Following instructions in a dual-task paradigm: Evidence for a temporary motor store in working memory. *Quarterly Journal of Experimental Psychology*, *71*(11), 2439-2449. https://doi.org/10.1177/1747021817743492

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Langerock, N., Vergauwe, E., & Barrouillet, P. (2014). The maintenance of cross-domain associations in the episodic buffer. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(4), 1096–1109. https://doi.org/10.1037/a0035783

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source?. *The Quarterly Journal of Experimental Psychology*, *56*(8), 1249-1268. https://doi.org/10.1080/02724980244000765

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Louwerse, M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*(2), 381–398. https://doi.org/10.1111/j.1551-6709.2010.01157.x

Louwerse, M. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In M. deVega, A. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning: A debate* (pp. 309–326).Oxford, England: Oxford University Press

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. https://doi.org/10.1016/j.cognition.2009.09.002

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology*, *3*(54). https://doi.org/10.3389/fpsyg.2012.00054

Marschark, M., & Paivio, A. (1977). Integrative processing of concrete and abstract sentences. *Journal of Verbal Learning and Verbal Behavior, 16*(2), 217-231. https://doi.org/10.1016/S0022-5371(77)80048-0

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25-45. https://doi.org/10.1146/annurev.psych.57.102904.190143

McRae, K., Nedjadrasul, D., Pau, R., Lo, B. P. H., & King, L. (2018). Abstract concepts and pictures of real-world situations activate one another. *Topics in Cognitive Science*, *10*(3), 518-532. https://doi.org/10.1111/tops.12328

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97. https://doi.org/10.1037/h0043158

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, *21*, 263-267 https://doi.org/10.1037/h0082978

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

Paivio, A. (1971). Imagery and language. In Segal, S.J., (Ed.), *Imagery: Current cognitive approaches.* (pp. 7-32). Academic Press

Peirce J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*(10), 1-8. https://doi.org/10.3389/neuro.11.010.2008

Phillips, L. H., Wynn, V. E., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory*, 7, 209–231. https://doi.org/10.1080/741944066

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303-345. https://doi.org/10.1111/j.1756-8765.2010.01111.x

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roome, H. E., Towse, J. N., & Crespo-Llado, M. M. (2019). Contextual support for children's recall within working memory. *Quarterly Journal of Experimental Psychology*, *72*(6), 1364-1378. https://doi.org/10.1177%2F1747021818804440

Rudner, M., Orfanidou, E., Cardin, V., Capek, C. M., Woll, B., & Rönnberg, J. (2016). Preexisting semantic representation improves working memory performance in the visuospatial domain. *Memory & Cognition*, *44*(4), 608-620. https://doi.org/10.3758/s13421-016-0585-z

Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, *3*(1), 83-119. https://doi.org/10.1515/langcog.2011.004

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322-339. https://doi.org/10.1037/met0000061

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive psychology*, *22*(1), 36-71. https://doi.org/10.1016/0010-0285(90)90003-M

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex*, *49*(1), 222–231. https://doi.org/10.1016/j.cortex.2011.10.005

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology*, *102*(1-3), 106-119. https://doi.org/10.1016/j.jphysparis.2008.03.014

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Suggate, S., Lehmann, J., Stoeger, H., & Jansen, P. (2019). Cognition embodied: mental rotation is faster for objects that imply a greater body–object interaction. *Journal of Cognitive Psychology*, *31*(8), 876-890. https://doi.org/10.1080/20445911.2019.1678627

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological review*, *114*(1), 104. https://psycnet.apa.org/doi/10.1037/0033-295X.114.1.104

Vannuscorps, G., & Caramazza, A. (2016). Impaired short-term memory for hand postures in individuals born without hands. *Cortex, 83,* 136–138. https://doi.org/10.1016/j.cortex.2016.07.019

Van't Wout, F., & Jarrold, C. (2020). The role of language in novel task learning. *Cognition, 194*(104036). https://doi.org/10.1016/j.cognition.2019.104036

Vermeulen, N., Chang, B., Mermillod, M., Pleyers, G., & Corneille, O. (2013). Memory for words representing modal concepts. *Experimental Psychology*, *60*(4), 293-301. https://doi.org/10.1027/1618-3169/a000199

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(02), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: MIT Press. (Original work published 1934).

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review, 14*(5), 779–804. https://doi.org/10.3758/BF03194105

Walker, P., Hitch, G. J., Dewhurst, S. A., Whiteley, H. E., & Brandimonte, M. A. (1997). The representation of nonstructural information in visual memory: Evidence from image combination. *Memory & cognition*, *25*(4), 484-491. https://doi.org/10.3758/BF03201124

Wingfield, C., & Connell, L. (2019). Understanding the role of linguistic distributional knowledge in cognition. Retrieved from PsyArXiv database. https://doi.org/10.31234/osf.io/hpm4z

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2018). The production effect and the generation effect improve memory in picture naming. *Memory*, *27*(3), 1–13. https://doi.org/10.1080/09658211.2018.1510966

5 Imageability Effects and Variance in Lexical Decision and Word Naming Tasks

In Chapter 4 I demonstrated that working memory capacity for object concepts is higher when language is available to use as a placeholder for complex sensorimotor information. Further insight into memory performance and what influences how well information is encoded can be provided by the study of sensorimotor and linguistic information in long-term memory. Before turning to long-term memory, I want to consider the type of information which gets temporarily activated in working memory when a decision about a word has to be made.

Much of the work on sensorimotor grounding focuses on language processing and whether sensorimotor information is useful in predicting, for example, lexical decision performance. However, very little work has focused on what kind of sensorimotor information is used to support this kind of processing, and how it relates to other variables often used as predictors of word processing. More specifically, extensive literature on imageability suggests that ease of consciously generating mental imagery supports lexical decision performance. Imageability is thought to reflect the sensory information associated with a concept, and in this respect it resembles sensorimotor strength, but whether imageability and sensorimotor strength provide the same kind of semantic information needed for word recognition is not conclusive. In the next chapter I will examine the nature of imageability ratings and how well they relate to sensorimotor strength, as well as evaluate how well conscious imagery performs as a predictor of word recognition across different sources of ratings, once sensorimotor information has been accounted for. By examining what kind of sensorimotor information is activated in working memory upon reading a word, it will help us to understand in subsequent work how sensorimotor information impacts on people’s ability to remember such words in a memory task.

Imageability Effects and Variance in Lexical Decision and Word Naming Tasks

Agata Dymarska, Louise Connell, and Briony Banks

Department of Psychology, Lancaster University

Author Note

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 682848) to LC.

Correspondence concerning this article should be addressed to Agata Dymarska, Department of Psychology, Fylde College, Lancaster University, Lancaster LA1 4YF, UK. Email: a.dymarska@lancaster.ac.uk

 Stimuli, data, and analysis code can be found at: https://osf.io/hrj3c/?view_only=289c468c6ae4496f9f8771f9ea5a0ec3

Abstract

Imageability, defined as ease of generating a mental image associated with a word, has been commonly used as a predictor of word recognition. However, imageability effects are highly variable across the literature, which raises questions about the robustness and stability of imageability as a construct. We compared six existing imageability norms in their ability to predict RT and accuracy in lexical decision and word naming. We found that different sources of imageability norms varied in their ability to predict word recognition performance, even when other lexical and sensorimotor sources of variance were partialled out. We conclude that the variance is not due to differences in word characteristics, but can be attributed to the way participants in different norming studies interpret and rate ease of generating a mental image. Our findings are not consistent with the idea that the ease of generating a mental image is a robust facilitator of word recognition. Rather, they suggest that imageability ratings are subject to a high level of random variability between sets of norms that impacts on their ability to predict behavioural results, and that imageability itself cannot be reliably rated as a stable construct.

Keywords: imageability; sensorimotor information; word recognition

Imageability Effects and Variance in Lexical Decision and Word Naming Tasks

Some concepts are easier to imagine perceptually than others; for example, “dog” versus “loyalty” are experienced and visualised very differently. This distinction is captured by imageability, which is the ease of generating a mental image associated with a word. According to dual-coding theory (Paivio, 1971; 1990; Sadoski & Paivio, 2004), words which can be directly linked with a mental image are considered more concrete, while abstract words, which are more difficult to imagine, likely rely more on word associations (e.g., Connell & Lynott, 2013; Lenci et al., 2018; Vigliocco et al., 2009). For example, “courthouse” refers to a building, or “teacher” refers to a person, both strongly visual concepts. But concepts such as “justice”, or “knowledge” do not have a direct, tangible referent. Imageability is thus believed to represent semantic information associated with a concrete concept (e.g., Bakhtiar & Weekes, 2015; Cortese et al., 1997), and the two measures (concreteness and imageability) are highly correlated (Kousta et al., 2011; Vergallito et al., 2020; Westbury & Moroschan, 2009) and often used interchangeably (Fernandino et al., 2015; Westbury & Moroschan, 2009).

This simple distinction between concepts has led to the idea that our ability to imagine things affects a number of aspects of our cognition, and particularly it has implications for processing different types of concepts. For example, concrete words are faster to recognise or to read out loud (Balota et al., 2004; Cortese & Schock, 2013; de Groot, 1989; Reilly & Desai, 2017). Imageability has been found to offer a large contribution to word recognition when other, traditionally used linguistic variables were accounted for. High frequency, early acquired, and short words are processed quickly and accurately (see Brysbaert et al., 2018 for review) due to our frequent and prolonged experience with them. But when a word is not recognised automatically based on its lexical features, its semantic representation has to be accessed, and generating a mental image could help with this; for example, words higher in imageability are recognised faster (than low imageability words) when they are low frequency (Cortese & Schock, 2013; González-Nosti et al., 2014), acquired later (Cortese & Khanna, 2007; Cortese & Schock, 2013; González-Nosti et al., 2014), or have atypical spelling (Evans et al., 2012; Woollams, 2015), and low phonological neighbourhood (Westbury & Moroschan, 2009). The facilitation of word processing by imageability is in line with the theory of semantic richness, where the stronger a semantic variable (such as semantic neighbourhood density, number of features, emotional valence etc.) is for a particular word, the easier it is to recognise or remember that word (Buchanan et al., 2001; Pexman et al., 2008; Yap et al., 2012).

However, whether imageability indeed measures the ease of consciously generating a mental image is not clear, as it has a different relationship with different types of words. For example, verbs are rated as lower in imageability than nouns, even when they are homonyms of highly imageable nouns (e.g., “post”, “iron”, “attack”; Bird et al., 2000; Simonsen et al., 2013). Since nouns often denote manipulable objects which we interact with on a daily basis, or physical entities which we can interact with using our senses, it is not unexpected that they are easier to visualise. Nouns therefore have a somewhat different relationship with imageability than other parts of speech. Bedny and Thompson-Schill (2006) found that brain areas associated with visual semantic information are activated more when processing words with higher imageability. They also found an interaction between imageability and grammatical class: two brain areas were more active when processing lower imageability nouns, but not verbs, suggesting that highly imageable nouns are processed more easily. This raises the question of reliability: if only nouns can be reliably rated for imageability (and are perhaps rated higher than other parts of speech) then imageability is only useful in processing of nouns, but not other parts of speech. The overall effects of imageability found in the literature might therefore be overestimated, and in reality imageability effects may be an artefact of our experience with nouns.

Additionally, the visual focus of imageability ratings limits our understanding about the nature of semantic information related to a concept; indeed, conceptual representations rely on other senses as well as vision, and also motor experience. Asking participants to rate the extent of their experience with different concept via their senses (vision, smell, touch, hearing, and interoception) as well as action effectors (hand/arm, foot/leg, torso, head, mouth) generates a measure of sensorimotor strength, which has been found to predict lexical decision (Lynott et al., 2019). Contrary to imageability, this is activated automatically and not always fully available to conscious awareness (Connell & Lynott, 2016a). Indeed, Connell and Lynott (2012a) found that perceptual strength (e.g., information about vision and touch) outperforms imageability in lexical decision tasks (see also: Juhasz et al., 2011), suggesting that sensorimotor strength represents the semantic content of a word needed for word recognition well, and that imageability does not capture semantic information about a concept to the same extent. Although Paivio stated in his later work that all senses can support a representation of a concrete concept (Sadoski & Paivio, 2012), research on imageability, as well as the lay meaning of imageability, has focused on vision. Imageability norming studies tend to instruct participants to focus on their ability to generate a mental image ("a mental picture, sound, or other sensory experience", Gillhooly & Logie, 1980). Even when the instructions allow for rating multisensory experience, the word "image" or "imagery" itself is highly associated with visual experience (Lynott et al., 2019), and it is hard not to be biased towards visual experiences when generating an image associated with a word. Because of that, mental imagery may refer to the ability to visualise what something looks like, not so much what it tastes, feels or sounds like, which means it does not capture our full semantic experience of the concept, such as the automatic, effortless simulation of objects and situations. More specifically, the consciously available information provided as a single rating does not correspond to the sum of multiple ratings of individual modalities (Connell & Lynott, 2016a; Pecher et al., 2009). In fact, if the imageability effect is mainly driven by visual information, its effect may be overestimated by the fact that not only are concepts which can be perceived with vision processed faster (Connell & Lynott, 2014; Chedid et al., 2019), but also directing one's attention to a particular modality, such as vision, can influence how well the other modalities process information (Connell & Lynott, 2012b). Additionally, the importance of the effect is overestimated because it does not reflect the full sensorimotor experience which supports conceptual representation.

While there is plenty of evidence of the effect of imageability on word recognition[2], there is also variability between studies, and some evidence suggests that imageability ratings are not the most robust predictor of word processing, raising further questions about its reliability. For example, Cortese and Fugett (2004) report that the average imageability ratings in their study were significantly lower than the average ratings from the Toglia and Battig (1978) norms on the same set of words (e.g. the word "norm" – rated 2.5 vs 6 out of 7 in each study, respectively), while Clark and Paivio (2004) report the correlation between their norms and Paivio et al.'s (1968) norms to be only r = .67 (i.e. 45% of shared variance) for adjectives. Effect sizes of imageability on lexical tasks also vary between studies: The zero-order correlation between imageability and response times (RTs) can range from r = -.22 (on lexical decision, Bird et al., 2001), to r = -.19 (on lexical decision, Bennett et al., 2011), to r = -.13 (on word naming, Bird et al., 2001). Comparing imageability effects on RTs in a lexical decision task, Balota et al. (2004) report standardized regression coefficients ranging from β = -.16 (for the Toglia and Battig ratings) to β = -.27 (Cortese & Fugett, 2004), while in

[2] It is possible that studies which do not find evidence supporting an imageability effect are subject to a publication bias, and are not published as often as studies which do find evidence.

other studies the effects of imageability on RT varied from β = -.28 (Bennett et al., 2011) to β = -.09 (Connell & Lynott, 2012a). A thorough comparison of the stimuli and methodology of these studies could shed light on whether imageability effects can be considered reliable.

In addition to the variability in effect sizes, some studies find no effect or contradictory effects of imageability on word processing, particularly when other linguistic and semantic variables are included. Imageability did not contribute to word recognition performance over frequency and age of acquisition (Brysbaert et al., 2000), or when length, frequency, emotionality and context were accounted for (Westbury et al., 2013). In another study, imageability explained speed of response, but not accuracy in a lexical decision task (Yap et al., 2015). Additionally, it did not contribute to speed of word recognition when included as a moderator alongside phonological neighbourhood (Yates, 2005). Some results simply negate each other: In word naming, Bakhtiar and Weekes (2015) found that the facilitatory effect of early acquired words is stronger for low imageability words, while Wilson et al. (2013) found that this effect appears only with high imageability words. Even within the same study, effects of imageability can be unclear. Vergallito et al. (2020) found that a model with imageability was 31 times more likely to predict lexical decision RT in Italian than its nearest competitor (which included strength ratings of the five perceptual modalities). However, when words taken from Connell and Lynott (2012a) were analysed in this study using the same methods, Vergallito and colleagues found that perceptual strength predicted more variance than imageability (0.54% more for accuracy, 0.63% more for RT), indicating that different word samples show different sensitivity to imageability ratings. Therefore, the effects of imageability seem to vary depending on the lexical variables that are controlled for, as well as the tasks and measures used in the analysis.

Such variability in imageability effects is surprising for a construct which is thought to measure a fundamental process in mental representations. However, it might be explained by methodological differences; for example, different studies often analyse imageability ratings for different words and control for other variables using different lexical models. Additionally, a lot of studies only analyse a small sample of words (e.g., 80 words in Evans et al., 2012, or 240 words in Yap et al., 2015), usually from extreme ends on the imageability scale to make a clear comparison. Research on the contribution of imageability to word recognition has tended to examine only one individual set of norms at a time, and there is no comprehensive analysis of imageability across norms from different sources. While newer imageability norms claim consistency with existing ones in method and ratings (e.g. Bird et al. 2001; Cortese & Schock, 2013; Scott et al., 2018), they vary in terms of participants' age and country of testing, word samples and rating distributions, as well as testing methods (see Table 1 for an overview of the norms used in the current study).

Current Study

In the present paper, our goal was to shed more light on imageability effects in word recognition by comparing the effects of a number of different imageability norms in a series of exploratory studies. One possible reason for the variability in published imageability results is low consistency in the baseline lexical variables examined or controlled alongside imageability. In Study 1, we therefore examined whether different sources of imageability norms produce consistent effects in word processing when the baseline (lexical) model was held constant. Results remained highly variable, with imageability ratings ranging from non-existent to large depending on the source of the imageability norms used.

Hence, in Study 2, we explored what possible aspects of the imageability norms might explain their variability. We obtained 6 lexical and sensorimotor components through a PCA, which we used as predictors of existing imageability ratings. We found that variance in imageability ratings was explained by sensorimotor variables such as visual and haptic strength, and object manipulability, but some of the ratings could be explained by sensorimotor variables more than others. This indicates that some of the imageability effects in word recognition could be attributed to sensorimotor grounding of the word, rather than ease of consciously generating a mental image. Nonetheless, the variability between different norms remained. The PCA results also allowed us to partial out variability which stems from the fact that different imageability norms use different word samples with different characteristics, and to investigate whether this could have been the cause of variability in the imageability effect sizes.

Thus, in Study 3 we used the lexical and sensorimotor variables in a baseline model, and examined the effects of imageability over and above these variables. We found that the results were still quite inconsistent, with some norms producing large effects, but others not eliciting effects on word recognition at all. These results suggested that imageability is not, in fact, a stable construct that acts as the single best predictor of *semantic effects* in word processing. Finally, we conducted an internal meta-analysis to determine whether there is an overall effect of imageability on word recognition. As before, we found that imageability did not produce consistent effects, even when the analysis included moderating variables.

Study 1 - Does imageability consistently contribute to word recognition?

In this study, we analysed 6 different sources of imageability norms in their ability to predict word recognition. We controlled for variance that can be attributed to word features with an objectively selected lexical baseline model, and analysed lexical decision and word naming performance from the English Lexicon Project (Balota et al., 2007) and lexical decision data from the British Lexical Project (Keuleers et al., 2012), using hierarchical regression analyses, to compare how well different sources of imageability predicted word recognition performance. We expected imageability to facilitate accuracy and RT of word recognition, and the effects to be stronger for low frequency words.

Method

Materials

Data selection. We used imageability ratings from 7 sources (Bird norms - Bird et al., 2001; Bristol norms - Davis & Stadthagen-Gonzalez, 2006; Chiarello norms - Chiarello et al., 1999; Cortese norms - Cortese & Fugett, 2004; Schock et al., 2012; Glasgow norms - Scott et al., 2018; MRC norms - Gilhooly & Logie, 1980; Paivio et al., 1968 and other studies). We combined Cortese and Fugett (2004) and Schock et al. (2012) into one larger set of norms, because they were obtained by the same research group, thus method and population tested were sufficiently similar, and they covered monosyllabic and di-syllabic words respectively. Table 1 shows an overview of the norming studies.

Table 1: Overview of the imageability norms used in the current study

Authors	Publication Year	Country	Participant Age	Number of Words	Type of Words
Bird et al.	2001	UK	M=65.0	2645	All Types
Stadthagen-Gonzales & Davis, ("Bristol norms")	2006	UK	M=19.7	1526	Nouns & Verbs
Chiarello et al.	1999	USA	M=20.7	1197	Nouns & Verbs
Cortese et al.; Schock et al.	2004&2012	USA	M=23.8	6000	All Types
Scott et al. ("Glasgow norms")	2018	UK	M=21.7	5500	content words (N, V, Adj, Adv)
MRC (Gilhooly & Logie; Paivio et al. and more)	1960s-1980s	UK+USA+ Canada	NA	9240	Nouns & Verbs

We collated corresponding lexical decision accuracy and response time (RT) data from the British Lexicon Project (BLP, Keuleers et al., 2012) and the English Lexicon Project (ELP, Balota et al., 2007), as well as word naming accuracy and RT from the ELP. These would be used as our dependent variables, as a measure of performance on lexical tasks. This resulted in a sample of 9560 words with Lexical Decision and Word Naming Accuracy and RT data from the ELP, and a sample of 7550 words with Lexical Decision Accuracy and RT from the BLP[3]. A breakdown of word subsamples for each set of imageability norms is available in supplemental materials. R tidyverse package (Wickham et al., 2019) was used to combine datasets and calculate average values of different variables per dataset. All other analyses were carried out in JASP (0.14.1: JASP Team, 2020).

Null lexical model selection. While variables like word frequency, length or phonological neighbourhood are often included in null (i.e., baseline) lexical decision models (Connell & Lynott, 2012a, Cortese et al., 2010, 2015; Yap et al., 2012), we wanted to select a model that provided optimal fit to the current data. We did not include age of acquisition, as there is some disagreement as to whether it is a lexical or a lexical–semantic variable (e.g., Cortese & Khanna, 2007; Zevin & Seidenberg, 2002). Our candidate lexical predictors were obtained from the ELP (word length, log subtitle frequency [LgSUBTLWF], log subtitle contextual diversity [LgSUBTLCD], Orthographic Levenshtein Distance [OLD], Phonological Levenshtein Distance [PLD], number of syllables). We also included prevalence, which is another way of estimating frequency by measuring how many participants know a particular word (Brysbaert et al., 2018), and Zipf frequency (van Heuven et al., 2014), which is a standardised measure of word frequency that can outperform log frequency. From this set of lexical variables, we objectively selected a baseline lexical model which provided the best fit to the word recognition data.

We ran a Bayesian regression model with 8 possible candidate variables (word length, number of syllables, OLD, PLD, LgSUBTLWF, LgSUBTLCD, Zipf frequency, prevalence) as predictor variables, and accuracy and RT from each BLP and ELP dataset as dependent variables. For each dependent variable, we selected those predictors where evidence for their inclusion was $BF_{10} > 3$ (detailed analysis is available in supplemental materials), that is, the

[3] For the regression analyses we removed words which did not have sensorimotor strength ratings to make potential future comparison more balanced, N=691 for BLP and N=721 for ELP data.

model explained more variance in word recognition performance when the variable was included, than when it was not. After selecting the best candidate model for each analysis, we combined the best predictors from RT and accuracy, and made sure that their inclusion offered the highest possible R^2 value. We then combined the predictors from all three tasks (BLP lexical decision, as well as ELP lexical decision and word naming), and again made sure that for each DV, the inclusion of all predictors offered the highest R^2 value. The final null model consisted of: word length, number of syllables, LgSUBTLCD, OLD, and prevalence. LgSUBTLCD, which is a measure of how many different contexts a word appears in, was considered a measure of frequency, and was expected to interact with imageability in the same way as other frequency measures have been reported to interact. These 5 variables were used as a null model in the subsequent analysis.

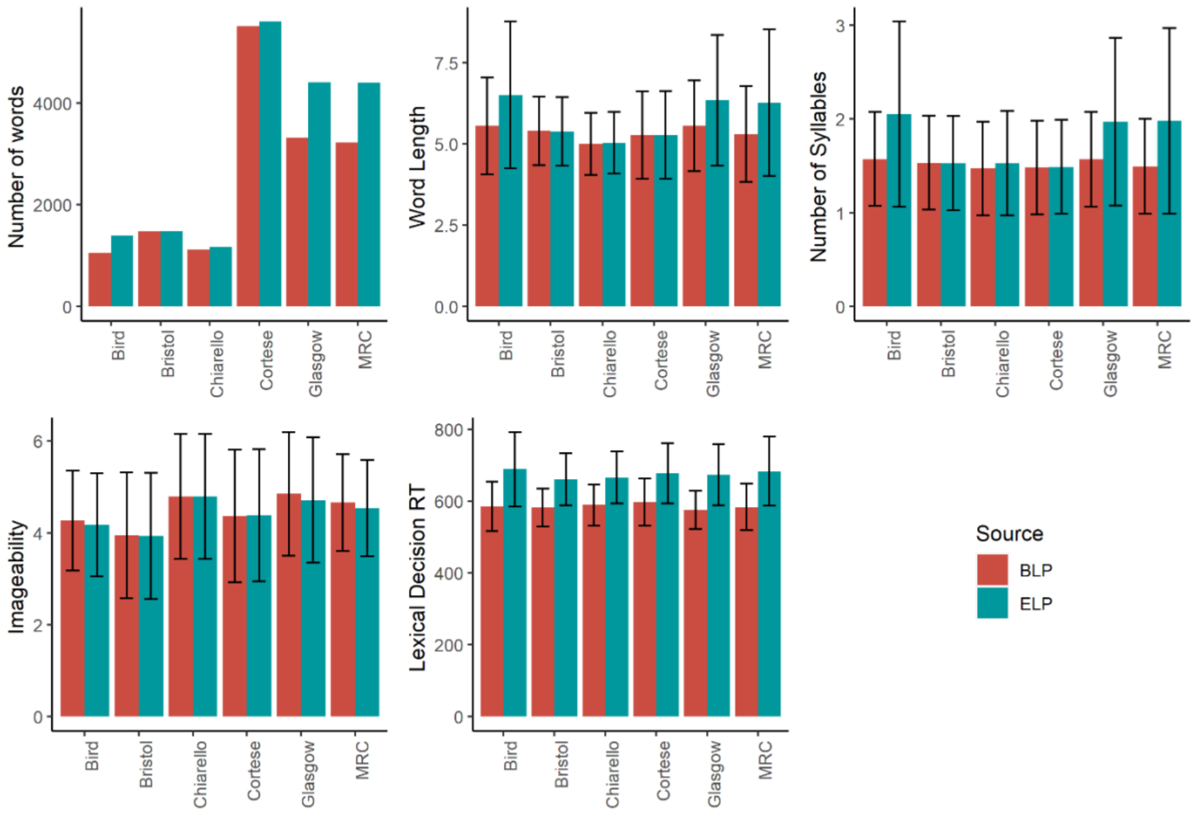
Design and analysis

For each set of imageability rating norms, we ran a hierarchical regression model to see how well the imageability norms predicted accuracy and zRT for lexical decision and word naming tasks from BLP and ELP. In Step 1 we entered all lexical predictors selected above, in Step 2 we added imageability (each set of norms was entered in a separate analysis), and in Step 3 we added the imageability*frequency interaction using contextual diversity as a proxy for frequency because it was the best frequency measure in the null model selection. This allowed to obtain the contribution of imageability for words of varying frequency. The step was repeated for every lexical DV and every imageability dataset. Based on previous literature (e.g., Cortese & Schock, 2013), we expected imageability to have a positive effect on accuracy and a negative effect on zRT (higher imageability=faster responses). We also expected the interaction to follow the same direction (low contextual diversity=stronger imageability effects). We ran Bayesian linear regressions with default JZS priors ($r = .354$) and a Bernoulli distribution ($p = 0.5$), from which we report Bayes Factors for model comparisons between hierarchical steps and posterior inclusion Bayes Factors of imageability coefficients (i.e., indicating the relative likelihood of models including a particular predictor compared to models excluding it). In addition, to calculate part correlation coefficients for each predictor (i.e., the unique contribution each predictor makes to the dependent measure in question), we ran NHST linear regression analyses using the same structure as the Bayesian linear regression. When the imageability effect is positive, negative values of part correlation for the imageability*frequency interaction indicate that the effect of imageability is lower for high frequency words (+1 SD), and higher for low frequency words (-1 SD). Conversely, positive values indicate that the effect is stronger for high frequency words (+1 SD), and weaker for low frequency words. However, it is the opposite for negative imageability effects, because the interaction is multipled by a negative value.

We expect that imageability will facilitate word processing, thus a positive imageability effect on accuracy and a negative interaction would indicate that, as predicted, higher imageability leads to more accurate responses, in particular for low frequency words. Similarly, a negative imageability effect on RT and a positive interaction would mean that, as predicted, higher imageability leads to shorter (faster) RT, in particular for low frequency words.

Results & Discussion

Figure 1 shows that word characteristics were mostly consistent across norms, except for the number of words which inevitably varied. However, there were some differences between tasks. First, there were slightly more words in the ELP dataset than in the BLP dataset. Compared to the BLP dataset, the ELP words were on average 1 letter and 0.5 syllable longer for the Bird, Glasgow and MRC norms. Participants also responded to the words from the ELP around 100ms slower than to the words from the BLP.



Note: the key variables displayed here are the ones where the norms differ the most, descriptive statistics for all variables are available in supplemental materials

Figure 1: Mean characteristics of the key variables of the imageability norms. Error bars represent ±1 Standard Deviation.

Table 2: Variance in word recognition accuracy explained by each step of the regression model (change in R^2, with levels of Bayesian evidence) and uniquely explained by imageability in the Step 2 and 3 models (squared part correlations).

Statistic	Bird	Bristol	Chiarello	Cortese	Glasgow	MRC
	BLP Lexical Decision					
Null model R^2	0.532***	0.468***	0.552***	0.619***	0.403***	0.569***
Imageability ΔR^2	0.019***	0.008***	0.001	0.002***	0.005***	0.009***
Imageability*frequency ΔR^2	0.019***	0.006***	0.001	0.001*	0.002**	0.004***
Imageability sr^2	0.032	0.013	0.002	0.003	0.008	0.013
Imageability*frequency sr^2	0.019	0.006	0.001	0.001	0.003	0.004
	ELP Lexical Decision					
Null model R^2	0.523***	0.492***	0.515***	0.591***	0.435***	0.514***
Imageability ΔR^2	0.006***	0.000	0.007***	0.011***	0.000	0.002***
Imageability*frequency ΔR^2	0.002	0.001	0.005**	0.006***	0.002**	0.000
Imageability sr^2	0.007	0.000	0.011	0.016	0.000	0.002
Imageability*frequency sr^2	0.002	0.001	0.005	0.005	0.002	0.000
	ELP Word Naming					
Null model R^2	0.285***	0.178***	0.171***	0.228***	0.181***	0.244***
Imageability ΔR^2	0.003	0.000	0.001	0.002***	0.001	0.004***
Imageability*frequency ΔR^2	0.013***	0.000	0.000	0.001	0.001	0.006***
Imageability sr^2	0.006	0.000	0.001	0.004*	0.002	0.008
Imageability*frequency sr^2	0.013	0.000	0.000	0.001	0.001	0.006

* $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence

Table 3: Variance in word recognition RT explained by each step of the regression model (change in R^2, with levels of Bayesian evidence) and uniquely explained by imageability in the Step 2 and 3 models (squared part correlations).

Statistic	Bird	Bristol	Chiarello	Cortese	Glasgow	MRC
	BLP lexical Decision					
Null model R^2	0.630***	0.555***	0.620***	0.645***	0.559***	0.617***
Imageability ΔR^2	0.014***	0.016***	0.009***	0.008***	0.013***	0.013***
Imageability*frequency ΔR^2	0.000	0.000	0.000	0.002***	0.000	0.003***
Imageability sr^2	0.018	0.016	0.007	0.005	0.012	0.005
Imageability*frequency sr^2	0.000	0.001	0.000	0.002	0.000	0.003
	ELP Lexical Decision					
Null model R^2	0.699***	0.570***	0.593***	0.627***	0.645***	0.680***
Imageability ΔR^2	0.009***	0.005***	0.011***	0.013***	0.003***	0.007***
Imageability*frequency ΔR^2	0.001	0.003*	0.001	0.004***	0.001***	0.003***
Imageability sr^2	0.008	0.002	0.008	0.007	0.001	0.003
Imageability*frequency sr^2	0.001	0.003	0.001	0.004	0.002	0.003
	ELP Word Naming					
Null model R^2	0.513***	0.284***	0.225***	0.373***	0.444***	0.501***
Imageability ΔR^2	0.007***	0.000	0.000	0.003***	0.001	0.002***
Imageability*frequency ΔR^2	0.000	0.000	0.001	0.003***	0.000	0.000
Imageability sr^2	0.007	0.000	0.000	0.000	0.001	0.002
Imageability*frequency sr^2	0.001	0.000	0.000	0.003	0.000	0.000

* $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence

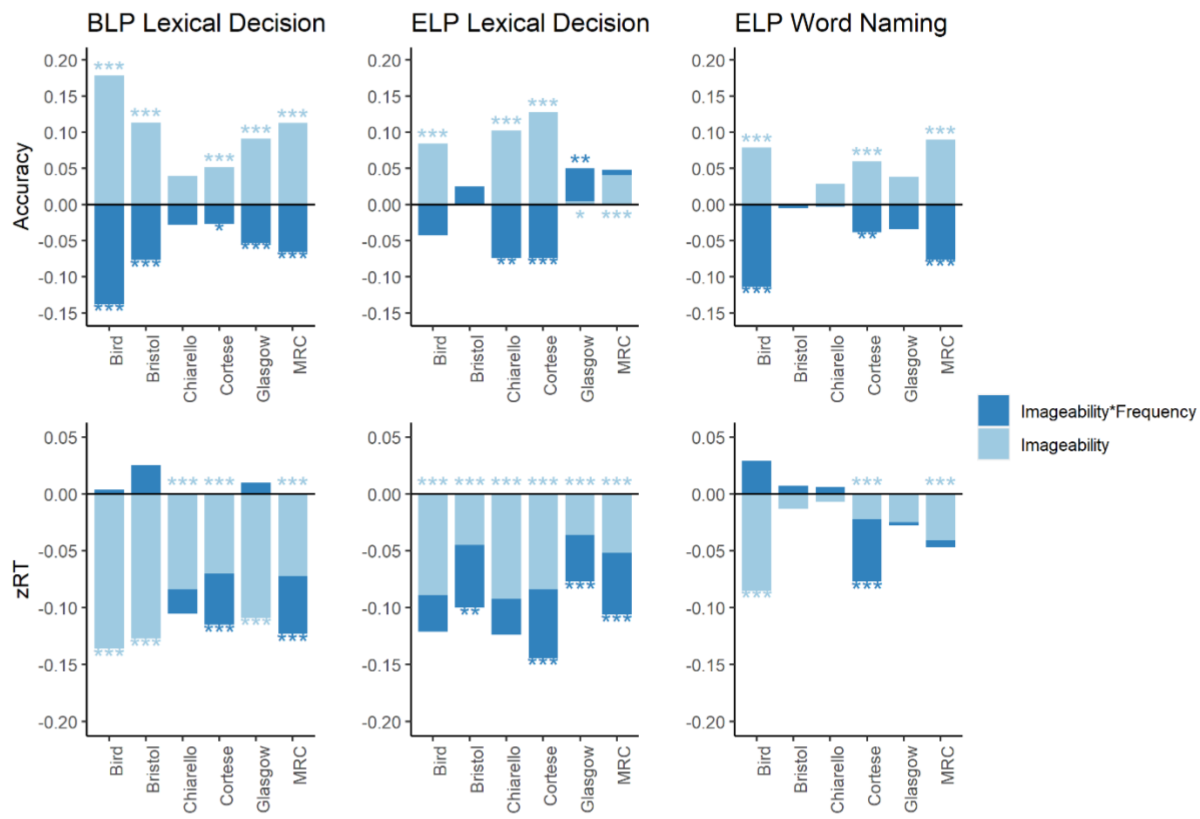


Figure 2: Part correlations in the Step 3 (final) model. Stacked bars represent the combined effect of imageability (lighter shade) and the imageability*frequency interaction (lighter shade) in the Step 3 (final) regression model. The light blue set of asterisks per bar refer to the Bayes Factor (BF) of imageability coefficient, while the dark blue set of asterisks refer to the Bayes Factor (BF) of imageability*frequency coefficient, * $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence.

Lexical (null) model

As shown in Table 2, R^2 for lexical decision accuracy varied across different sets of imageability norms, and even more so between different lexical tasks. Depending on the task and imageability dataset, lexical models explained between 17.1% to 69.9% variability in word processing. The lexical model's contribution was lowest for word naming – 17.1% to 28.5% variance for accuracy, and between 22.5% and 51.3% for RT. For lexical decision, the model performed similarly for BLP and ELP data, with accuracy having a wider range (40.3% - 61.9%) than RT (55.5% - 69.9%).

Imageability

As shown in Tables 2 and 3, change in R^2 for the Step 2 model with imageability shows that imageability contributed up to 1.9% of additional variability over and above the lexical variables included in the null model, depending on the task and dataset.

The imageability effect size was highly variable across different sets of norms for both accuracy and RT. The Bird, Cortese and MRC norms were the most reliable in terms of eliciting imageability effects in the expected direction for both accuracy and RT, in both lexical decision and word naming. Other norms performed inconsistently. The Glasgow norms produced effects in four of the six DVs, while the Bristol and Chiarello norms performed most poorly overall, producing effects in only three DVs.

There were also systematic differences between tasks and between dataset sources. Imageability effect sizes were generally smaller for word naming than for lexical decision, consistent with previous findings in the literature (Ferrand et al. 2011); indeed, half of the norms (Bristol, Chiarello and Glasgow norms) failed to produce any effects on naming at all. Within lexical decision, imageability effects were generally larger for data from the BLP than the ELP, regardless of whether the norms originated from UK participants (Bird, Bristol and Glasgow norms) or US participants (Chiarello or Cortese norms).

Imageability*Frequency

The model with an interaction revealed that the effect sizes of imageability for low frequency words varied between the norms, more than for average frequency words (when no interaction was included in the model). No set of norms produced a consistent facilitation effect on both accuracy and RT. While the Cortese norms had an effect on all tasks, for low frequency words its effects on RT were reduced, that is, imageability had the largest facilitatory effect on high frequency words, contrary to our predictions. Other norms also had inconsistent effects, typically facilitating accuracy, but not RT (Bird and MRC norms). The

Bristol, Chiarello and Glasgow norms performed the worst again, producing an effect in the expected direction in only 1 of the 6 DVs.

While there were some systematic patterns in the effects of imageability across different word recognition tasks, the effects are overall weaker and more inconsistent than for average frequency words (i.e., when the unique contribution of imageability was calculated with frequency held constant at the variable mean). The effects of imageability on accuracy were strong for the BLP dataset where 5 out of 6 norms interacted with frequency, but in ELP lexical decision only 2 sets of norms had an expected effect on accuracy, (and one, the Glasgow norms, produced an effect in the opposite direction). In ELP word naming 3 sets of norms elicited an expected effect on accuracy. None of the norms predicted facilitation of RT for lower frequency words in lexical decision, or in word naming. This did not support previous findings that imageability contributes reliably to word recognition, and that it has a stronger facilitatory effect on low frequency words (Cortese & Schock, 2013; González-Nosti et al., 2014).

Conclusion

We found that imageability did not consistently predict word recognition performance even when the same lexical baseline model was used in the analysis. There were large differences in effect sizes depending on the norms used. This pattern of results could be due to a number of reasons. Each dataset used different words and participants, which could differ in some characteristics (although Figure 1 shows that the norms were consistent across many lexical variables). Some participants might also be more accurate at judging the ease of generating a mental image, and these differences could affect how successful imageability is at predicting word recognition. It is also possible that ease of generating mental imagery is simply not a reliable predictor of word recognition, and any effects found in the literature so far cannot be attributed to ease of generating a mental image of a word. Instead, imageability ratings might capture partial aspects of sensorimotor information, which is used to represent word meaning. However, since imageability is primarily a vision-associated concept, it does not explain full sensorimotor processing, which leads to variation of its effects depending on the type of words that are analysed and their sensorimotor content.

In the next study, we aim to address this issue by analysing what lexical and semantic variables may contribute to variance in imageability not otherwise explained by ease of generating a mental image.

Study 2 – Which aspects of the imageability norms might explain their variability?

Study 1 showed that imageability ratings can explain some differences in word recognition performance, but there is variability in the results depending on the dataset. We wanted to examine potential sources of lexical and sensorimotor variance in different imageability ratings. One possibility is that the variance stems from different words used in different norming studies. For example, imageability may contribute more to processing of nouns, or low frequency words, and this may cause discrepancies in how different norms explain word recognition. Alternatively, it may be due to participant variability. Although each set of norms used the same instructions, they were was collected from different participants. It is possible that some of them rated a sensorimotor representation of the word which was not directly the ease of generating a mental image, but for example, the visual strength of the word. This is still useful for word recognition but does not have the same effect as ease of generating a mental image.

To assess whether imageability ratings were associated with any other variables, we conducted a Principal Component Analysis (PCA) of a number of lexical and sensorimotor variables. The aim was to identify which aspects of lexical and sensorimotor experience might contribute to differences between imageability ratings.

Method

Materials

We used the same words and norms which were collated for Study 1 and added further lexical and sensorimotor variables. All variables are listed in Table 4.

Table 4: Variables used in Principal Components Analysis in Study 2 and the rotated components to which they most strongly contributed with positive or negative weighting (*r* > .3 or < -.3)

Original variable	Source	Definition	PCA component
LgSUBTLWF	ELP	Log word frequency	+Frequency
LgSUBTLCD	ELP	Log contextual diversity (how many contexts a word appears in)	+Frequency
Zipf Frequency	Van Heuven et al. (2014)	Word frequency on Zipf scale	+Frequency
Prevalence	Brysbaert et al. (2018)	How many people know the word	+Frequency
Familiarity	Stadthagen-Gonzales & Davis (2006); Scott et al. (2018); Wilson (1988)	How subjectively familiar a word seems (ratings)	+Frequency
Age of Acquisition	Kuperman et al. (2012)[a]	Approximate age that the word was learned	–Frequency
Linguistic distributional distance (LDD20)	Generated for the purpose of this analysis	Distributional neighbourhood (mean cosine distance to closest 20 neighbours, based on vectors of log co-occurrence frequency)	–Frequency
Word length	ELP	Word length in letters	+Length
Number of syllables	ELP	Word length in syllables	+Length
Orthographic Levenshtein Distance (OLD20)	ELP	Orthographic neighbourhood (mean letter Levenshtein distance to closest 20 neighbours)	+Length
Phonological Levenshtein Distance (PLD20)	ELP	Phonological neighbourhood (mean phoneme Levenshtein distance to closest 20 neighbours)	+Length
Noun (part of speech)	ELP	Whether or not word is a noun (binary coded: noun=1, non-noun=0)	+Body
Torso action strength	LSN	Motor strength in torso effector	+Body
Foot/leg action strength	LSN	Motor strength in foot/leg effector	+Body, +Object
Hand/arm action strength	LSN	Motor strength in hand/arm effector	+Body, +Object, +Communication, +Food
Composite sensorimotor strength	LSN	Aggregated sensorimotor strength in all dimensions (Minkowski-3 distance of 11-dimension vector from the origin)	+Communication
Head action strength	LSN	Motor strength in head effector	+Communication
Auditory strength	LSN	Perceptual strength in hearing modality	+Communication, +Food
Mouth action strength	LSN	Motor strength in mouth effector	+Food
Gustatory strength	LSN	Perceptual strength in taste modality	+Food
Olfactory strength	LSN	Perceptual strength in smell modality	+Object
Visual strength	LSN	Perceptual strength in sight modality	+Object
Haptic strength	LSN	Perceptual strength in touch modality	+Object, +Body, –Communication
Interoceptive strength	LSN	Perceptual strength in interoceptive (sensations inside the body) modality	–Object, +Body, +Communication

[a] With extended norms from http://crr.ugent.be/archives/806

Note: ELP = English Lexicon project (Balota et al., 2007); LSN = Lancaster Sensorimotor Norms (Lynott et al., 2019).

We used the zipf, prevalence and familiarity variables, in addition to the standard subtitle frequency from ELP, to capture different types of word frequency. We also created the noun/non-noun distinction because of the previously discussed link between nouns and imageability (see Introduction) which we wanted to control for. Lexical distributional distance variable (LDD20) was generated for the purpose of this analysis. This variable was similar to Hoffman et al.'s (2013) and Shaoul and Westbury's (2010) measures of diversity of words that tend to occur in a similar context, and we included it because context diversity tends to affect word recognition (Adelman et al., 2006).

Design and Analysis

We conducted a PCA of the 24 lexical and sensorimotor variables in order to discover how the variables cluster into larger components that represent common lexical-semantic factors, which may contribute to variance in imageability, unrelated to ease of generating a mental image.

We used JASP to conduct a parallel analysis (95th percentile), with orthogonal varimax rotation and pairwise exclusion, using the correlation matrix. This reduced the original 24 dimensions to an optimal 6 principal components that captured 75.1% of the original variance, with a loading value of each variable per component. We then used the Principal function[4] in R to calculate and save rotated component scores for each word (a facility not currently available in JASP). We obtained 6 orthogonal components with intercorrelations of zero or near-zero. We used these new variables in a Bayesian linear regression analysis to see how much the component scores contributed to imageability ratings. We entered all 6 variables as predictors and each set of imageability norms as a DV in a separate regression analysis. We report posterior coefficients and BFs of variable inclusion from the most complex model from each analysis.

[4] Which is also used to perform PCA in JASP

Results and Discussion

The PCA of lexical and sensorimotor variables resulted in six rotated components, which reflected different lexical and sensorimotor characteristics of the words. We assigned names to the components based on what types of variables they loaded on, and what type of words were representative of the components. Two components represented lexical word features (RC1 – Frequency, RC2 – Length) and four components represented sensorimotor word features (RC3 - Body, RC4 - Food, RC5 - Objects, RC6 - Communication). The Body component was most strongly associated with Torso, as well as foot/leg and hand/arm experience. The Food component was strongly associated with gustatory and olfactory experience. The Object component was mostly dominated by visual experience, which makes is likely that this component would account for a large portion of the imageability effect. Finally, the Communication component was most strongly associated with auditory and head strength. Variables contributing to each component are presented in Figure 3. Example words which are rated high on each component are presented in Table 5.

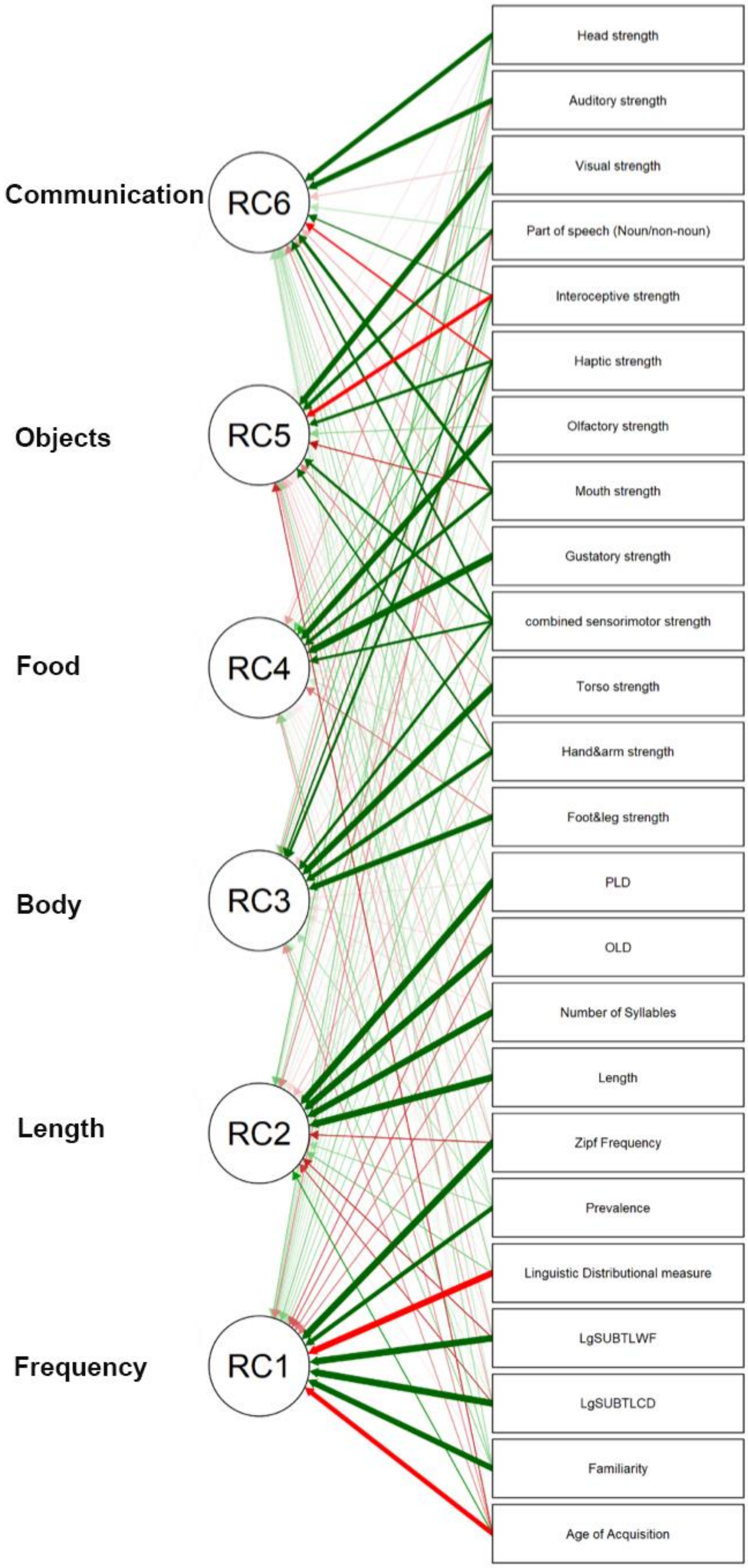


Figure 3: Rotated components obtained through the PCA, with the variables that loaded positively (green arrows) and negatively (red arrows) on the components.

Table 5: Example words rated high on each component

Component	Example words
Frequency (RC1)	“know”, “man”, “good”
Length (RC2)	“intercontinental”, “microbiology”
Body (RC3)	“clothes”, “bath”, “strong”
Food (RC4)	“pastry”, “supper”, “buffet”
Object (RC5)	“pen”, “basket”, “dog”
Communication (RC6)	“outrage”, “comedy”, “crying”

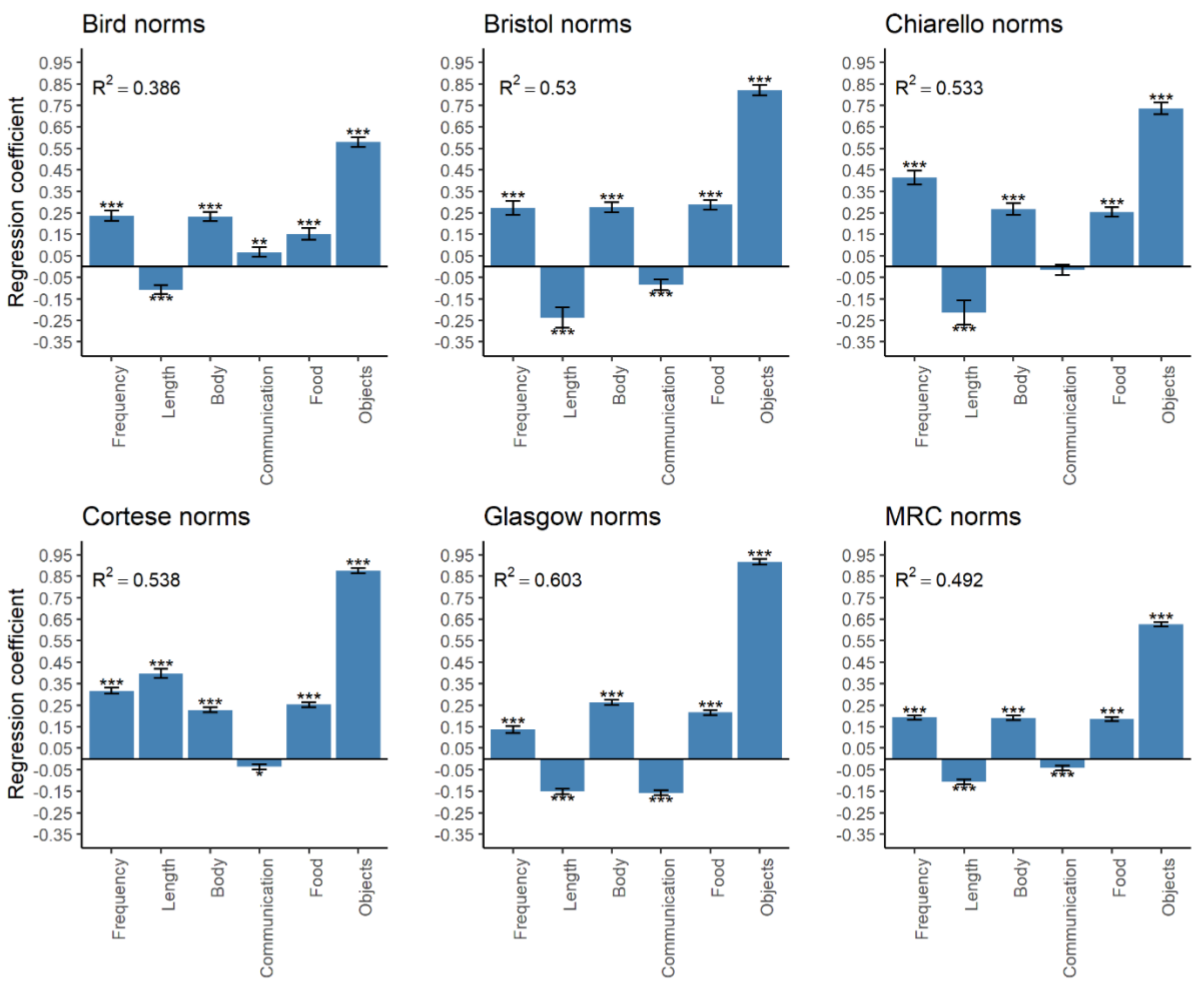


Figure 4: Regression coefficients of each component as a predictor of each set of norms, using the most complex model with all components. Error bars indicate ±1 Standard Deviation. The asterisks per bar refer to the inclusion Bayes Factor (BF) of the coefficient, * $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence. R^2 of the most complex model.

The effect of the components on imageability ratings varied across the norms, predicting 60.3% of variability in the Glasgow norms (highest contribution of components), but only 38.6% of variability in the Bird norms (lowest contribution of components). This difference suggests that imageability ratings from different sources differ in what they actually measure. While the components did not capture the contents of the Bird ratings well, the other ratings can be predicted by lexical and sensorimotor variables to a large extent, around 50% of variance. This indicates that only 50% of the variance in imageability ratings can be attributed to the ease of consciously generating a mental image.

Most components contributed to all sets of ratings. All norms were most strongly predicted by the Object component – that is, words were rated higher in imageability, when they related strongly to external object concepts (referents of nouns which can be seen, touched and manipulated). However, the effect sizes varied substantially between 33.5%-83.1%, depending on the imageability dataset. Imageability ratings were also consistently predicted, although with varying contribution, by the Frequency (1.8%-17.1%,), Body (3.6%-7.6%), and Food (2.3%-8.3%) components. Words were rated higher in imageability when they were to some extent more common (i.e., highly frequent, prevalent and familiar), higher in bodily experience (strongly visual, haptic and interoceptive) and related to food and eating (denoting the experience of smell, taste and mouth action).

However, there were inconsistencies in the effects of other components on imageability ratings. The Length component had a positive effect on the Cortese norms, where longer words were more likely to be rated higher in imageability, but this was in contrast to all the other norms, where Length elicited a negative effect on imageability ratings. The Communication component was the most inconsistent as a predictor, producing a small negative effect in most norms (Bristol, Chiarello, Glasgow and MRC), but a positive effect in the Bird norms, and no effect in the Cortese norms.

Overall, imageability ratings tend to reflect whether a word referent is a manipulable object (or another noun that are easy to interact with, such as an animal), and to some extent, whether it is a frequent word which relates to olfactory, gustatory and auditory experience, or words experienced interoceptively and with the body (hands, arms, legs, torso). However, high imageability does not consistently reflect whether word referents are linked to communication experience (through sound and mouth action), and is negatively related to the length of the word in some, but not all sets of ratings.

Conclusion

In Study 2 we identified several lexical and sensorimotor sources of variance between different imageability norms, which could explain why imageability was not a reliable predictor in Study 1. We obtained 6 rotated component scores which represent lexical and sensorimotor information about word referents, and we used them to investigate the nature of imageability ratings. The analysis showed that there was a range of lexical and sensorimotor information captured by imageability ratings, some of which was systematic and some of which differed from one set of norms to the next. This variability in lexico-semantic factors underlying imageability ratings in different norms might explain the variability in imageability effect sizes found in Study 1.

We examine this possibility in Study 3 by using the lexico-semantic components as baseline predictors of word recognition performance, and examining the effect of imageability above and beyond them, as with the lexical baseline model in Study 1. The remaining imageability effect would thus be a “pure” imageability contribution which includes only participants’ ease of generating a mental image, with other sources of variance in imageability, including sensorimotor grounding of concepts, accounted for. In the next study, we revisit our initial research question of whether imageability is a stable predictor of performance in lexical decision and word naming tasks, while controlling for variation in lexical and semantic variables.

Study 3 - Do word characteristics explain variability in imageability norms' word recognition performance?

In Study 1 and 2 we found that imageability has very variable effects on word recognition depending on what norms are used in the analysis, and the norms themselves show a lot of participant and item variability. We wanted to investigate whether imageability norms could predict word recognition more consistently when the variability of lexical and semantic characteristics of the words is partialled out using the rotated components. The differences in results across analyses could be due to differences in how participants rated imageability across different norming studies. That is, when participants are asked to rate imageability, they might be evaluating their ability to consciously image the concept, or they might actually be rating some aspect of the underlying sensorimotor grounding (e.g., visual strength), which may or may not be available to conscious awareness. If participants indeed tend to rate sensorimotor grounding of the concept and not imageability, we expect the variability in the norms' performance to persist, because they do not capture any additional information reliably. This would mean that imageability, as a measure of how easy it is to generate imagery, is not a reliable construct. If, on the other hand, the effect variability was due to variability of lexical models used in previous research, or differences between lexical and semantic characteristics of words used in imageability datasets, we expect the effects of imageability in the current analysis to be more consistent, reflecting the ease of generating mental imagery, as it is meant to.

Method

Materials

We used the 6 Rotated Component scores from Study 2, and the imageability norms and word recognition (lexical decision and word naming) data from Study 1. Data from the BLP dataset covered 6859 words, and the ELP dataset covered 8839 words.

Design and Analysis

We ran hierarchical linear regression with 6 lexical-semantic components. In Step 1 (null model) we entered the lexical components (Frequency and Length), in Step 2 we entered the 4 sensorimotor components (Body, Communication, Food, Object), and in Step 3 we entered imageability norms and their interaction with the Frequency component. As in Study 1, the dependent variables were accuracy and response time on lexical decision and word naming tasks, and we ran Bayesian linear regressions in JASP with default JZS priors ($r = .354$) and a Bernoulli distribution ($p = 0.5$)), from which we report Bayes Factors for model comparisons between hierarchical steps and inclusion Bayes Factors of coefficients (i.e., relative likelihood of models including a particular predictor compared to models excluding it). Again, to calculate part correlation coefficients for each predictor (i.e., the unique contribution each predictor makes to the dependent measure in question), we ran NHST linear regression analyses using the same structure as the Bayesian linear regression. This way we could test whether imageability explains language processing better after variability in lexical-semantic charcteristics is partialled out through PCA, when any suppression effects of other variables are eliminated. Each of the 6 imageability rating datasets was analysed separately, with all lexical tasks as DV (from both ELP and BLP datasets, as in Study 1).

Results and Discussion

Part-correlation coefficients for each set of imageability norms and each task are presented in Table 6, Table 7, and Figure 5.

Table 6: Variance in word recognition accuracy explained by each step of the regression model (change in R^2, with levels of Bayesian evidence) and uniquely explained by imageability in the Step 3 model (squared part correlations).

Statistic	Bird	Bristol	Chiarello	Cortese	Glasgow	MRC
			BLP Lexical Decision			
Null model R^2	0.403***	0.331***	0.382***	0.404***	0.299***	0.512***
Total sensorimotor (ΔR^2)	0.010	0.014**	0.008	0.015***	0.008***	0.013***
Total imageability (ΔR^2)	0.127***	0.027***	0.070***	0.065***	0.008***	0.100***
Imageability sr^2	0.084	0.019	0.034	0.052	0.004	0.071
Imageability*Frequency sr^2	0.057	0.014	0.04	0.022	0.007	0.036
			ELP Lexical Decision			
Null model R^2	0.343***	0.366***	0.364***	0.406***	0.301***	0.438***
Total sensorimotor (ΔR^2)	0.014**	0.009	0.006	0.014***	0.006***	0.008***
Total imageability (ΔR^2)	0.047***	0.000	0.088***	0.090***	0.003	0.030***
Imageability sr^2	0.031	0.001	0.045	0.074	0.001	0.024
Imageability*Frequency sr^2	0.018	0.000	0.047	0.028	0.000	0.008
			ELP Word Naming			
Null model R^2	0.216***	0.114***	0.103***	0.149***	0.141***	0.243***
Total sensorimotor (ΔR^2)	0.004	0.003	0.007	0.005***	0.002	0.005*
Total imageability (ΔR^2)	0.057***	0.000	0.039***	0.044***	0.006***	0.059***
Imageability sr^2	0.033	0.000	0.024	0.035	0.002	0.035
Imageability*Frequency sr^2	0.026	0.000	0.018	0.015	0.005	0.031

Note: * $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence. Note that the summed sr^2 for imageability and imageability*Frequency in the final model can be greater than the corresponding ΔR^2 due to redistribution of variance across all parameters in the final model.

Table 7: Variance in word recognition RT explained by each step of the regression model (change in R^2, with levels of Bayesian evidence) and uniquely explained by imageability in the Step 3 model (squared part correlations).

Statistic	Bird	Bristol	Chiarello	Cortese	Glasgow	MRC
			BLP Lexical Decision			
Null model R^2	0.555***	0.538***	0.585***	0.544***	0.557***	0.606***
Total sensorimotor (ΔR^2)	0.019***	0.021***	0.020***	0.029***	0.015***	0.026***
Total imageability (ΔR^2)	0.038***	0.007***	0.012***	0.016***	0.000	0.014***
Imageability sr^2	0.036	0.006	0.010	0.016	0.000	0.013
Imageability*Frequency sr^2	0.005	0.002	0.002	0.000	0.000	0.000
			ELP Lexical Decision			
Null model R^2	0.641***	0.525***	0.558***	0.531***	0.606***	0.649***
Total sensorimotor (ΔR^2)	0.012***	0.020***	0.015***	0.026***	0.011***	0.018***
Total imageability (ΔR^2)	0.012***	0.002	0.010***	0.022***	0.002***	0.006***
Imageability sr^2	0.015	0.000	0.008	0.02	0.001	0.006
Imageability*Frequency sr^2	0.000	0.002	0.003	0.001	0.002	0.001
			ELP Word Naming			
Null model R^2	0.511***	0.240***	0.215***	0.314***	0.428***	0.502***
Total sensorimotor (ΔR^2)	0.003	0.002	0.005	0.008***	0.002	0.007***
Total imageability (ΔR^2)	0.024***	0.001	0.003	0.010***	0.000	0.006***
Imageability sr^2	0.022	0.001	0.000	0.009	0.000	0.006
Imageability*Frequency sr^2	0.003	0.000	0.003	0.000	0.000	0.002

Note: * $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence. Note that the summed sr^2 for imageability and imageability*Frequency in the final model can be greater than the corresponding ΔR^2 due to redistribution of variance across all parameters in the final model.

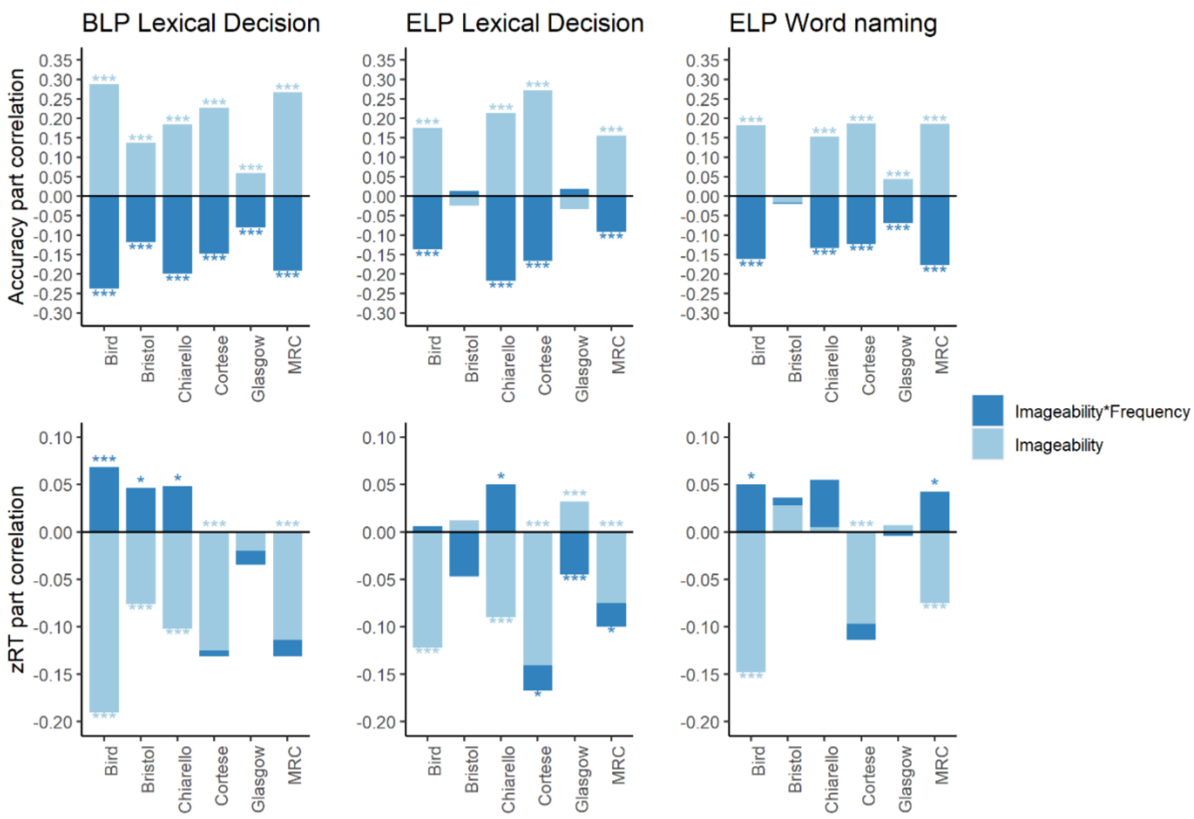


Figure 5: Part correlation in the Step 3 (final) model. Stacked bars represent the combined effect of imageability (light blue) and the imageability*Frequency component interaction (dark blue) in the Step 3 (final) regression model. The light blue set of asterisks per bar refer to the Bayes Factor (BF) of imageability coefficient, while the dark blue set of asterisks refer to the Bayes Factor (BF) of imageability*Frequency coefficient, * $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence.

Lexico-semantic model

As shown in Tables 6 and 7, the contribution of the lexical and sensorimotor components to lexical decision accuracy and RT was fairly consistent across norms, while for word naming accuracy and RT the differences were larger, and the effects varied between tasks. Depending on the DV and imageability dataset, lexical components explained between 10.3% to 64.9% of variability in word processing, and the sensorimotor components predicted an additional 0.2% - 2.9% of variability (a plot of how each sensorimotor component contributed to accuracy and RT and per each set of norms is available in supplemental materials for this chapter).

Imageability

Part correlation results showed that imageability explained up to 8.4% of unique variance in word recognition over lexical and sensorimotor information (Bird norms, BLP lexical decision accuracy). However, this was rather the exception, and mostly the contribution of imageability to word recognition was very low, often explaining less than 1% of the variance. Moreover, the effects still varied between the norms. The Bird, Cortese and MRC norms were the most consistent in facilitating accuracy and RT for all 6 DVs. The Chiarello norms were consistent predictors of accuracy, but did not always facilitate RT. The Bristol and Glasgow norms performed the worst overall, with only a small effect on accuracy in 2 out of 6 DVs in the expected direction.

There were also systematic differences between imageability effects on different tasks and datasets. Imageability effects were somewhat smaller for word naming than lexical decision, and were more variable for RT than for accuracy. As before, effects were more consistent for the BLP dataset, where most norms predicted facilitation of accuracy and RT. The effects on the ELP dataset were more variable. Two of the imageability norms (Bristol and Glasgow) did not elicit a facilitation effect on lexical decision, and while the Bristol norms had no effect on word naming, the Glasgow norms elicited an effect which was smaller than the effects of other, better performing norms. The other four sets of norms had a positive effect on word naming accuracy, but only 3 out of 6 norms (Bird, Cortese, MRC) facilitated word naming RT. Overall, the effects were larger for accuracy than RT, regardless of the task (see standardised regression coefficients in supplementals).

It is possible that the variation stems from how well participants were able to rate ease of generating a mental image in different norming studies. This may be due to the difficulty of rating words which do not refer to physical objects, for example “scope” (1.9 rating in Bristol norms, 4.8 in Cortese norms) or “truce” (2.3 rating in Cortese norms, 4.8 in MRC norms), where the rating is more dependent on the participants’ experience with the word, or their interpretation of the instructions. However, if only some types of words, like those where the referents are most tangible, can be reliably rated for imageability, then the concept is only useful for that specific subset of words, and cannot be used to study processing of other types of concepts.

Imageability*Frequency interaction

The effects of imageability on low frequency words were even more inconsistent than on average frequency words. Four sets of ratings (Bird, Chiarello, Cortese, MRC) elicited an effect on all accuracy measures, where the imageability effects were stronger for lower frequency words, while the other two sets of norms produced the expected effect in 1 (Bristol) or 2 (Glasgow) of the 3 DVs.

Some norms also elicited an effect on RT, but notably, the interaction was not always in the expected direction, that is, the effects were stronger for high frequency words (the Cortese, Glasgow and MRC norms in ELP lexical decision). As predicted, the Bristol and MRC norms interacted with frequency so that the imageability effect was stronger for low frequency words in 1 out of 3 DVs, while the Bird and Chiarello norms elicited a facilitation effect on RT 2 out of 3 times.

As before, effects varied between different analyses. For BLP lexical decision, all norms facilitated accuracy on low frequency words. However, for the ELP lexical decision and word naming, four and five sets of norms, respectively, interacted with frequency in the predicted direction. Further, only half of the norms an effect on RT for low frequency words in the expected direction for lexical decision, and 2 out of 6 for word naming, while the effects of the Cortese, Glasgow and MRC norms on ELP lexical decision were stronger for high frequency words, contrary to predictions and previous findings (Brysbaert et al., 2018; Raman & Baluch, 2001), and no set of norms produced an enhanced imageability effect on word naming RT for low frequency words.

Conclusion

In a series of hierarchical regression analyses, and with the use of part correlations, we compared the contribution of the six sets of imageability norms to word recognition performance. We used variables obtained through Principal Components Analysis in order to extract common lexical-sensorimotor factors that may provide extraneous sources of variance in imageability, which allowed us to remove variability associated with different word samples in different studies; something that we could not control for with the raw lexical variables in Study 1. A clear pattern emerged where some imageability norms were better predictors of lexical tasks than others, and furthermore, the results were not fully in line with the theories of how imageability influences word processing. Specifically, some of the imageability ratings did not elicit an effect on word recognition once lexical and sensorimotor characteristics of the word had been accounted for. This suggests that there is variability in how sensorimotor information contributes to imageability and its performance as a word recognition predictor. This is especially the case for information related to "Object" concepts – whether the word referent can be touched, seen and labelled as a noun – as indicated by the PCA and regression analyses in Study 2 (see also: supplemental materials for the plot of contribution of sensorimotor components to word recognition performance). More specifically, the results indicate that participants vary in how well they rate this information on the imageability scale. Thus, it appears that in the norming studies of imageability, which produced no effect beyond the lexical and sensorimotor components, participants' ratings were influenced by lexical artefacts of word frequency, or sensorimotor grounding in terms of eating food, using the body, communicating, and interacting with objects. On the other hand, some norms seemed to have captured useful information that goes beyond lexical and sensorimotor information about words and their referents. In these well-performing norms participants rated useful information which can predict word processing even once lexical and sensorimotor information is taken into account. This variability in our results indicates that despite using the same instructions and scales in every study, the construct of imageability itself is not stable. There is so much variability in the kind of information participants rate when providing the ratings, that the notion of imageability as the ease of generating a mental image is not a psychological construct which can be reliably rated. Given this level of inconsistency between different norms, tasks, and dependent variables, it is not clear how large the effect of imageability on word recognition is in general. In order to assess that, we conducted a meta-analysis of the imageability effects on word recognition.

Study 4 – Meta-analysis of imageability effects on word recognition

In order to further test the reliability of imageability as a construct, we conducted an internal meta-analysis of the part correlations from Study 3 after lexical and sensorimotor variability had been partialled out. We generated a meta-effect size for the overall imageability effect, as well as per norms, to compare the ability of different norms to predict word recognition performance. We also investigated which factors moderated the relationship between imageability and word recognition. If imageability norming studies all measure the same thing (i.e., the ease of generating conscious mental imagery for a given concept), then once task and source variation is accounted for, all effect sizes should be within the same range, showing an overall meta effect. If, however, imageability norms do not measure ease of imageability, then there will be a large variation between the effect sizes of different studies, and potentially no reliable meta effect, which will show that imageability norms are not reliable in predicting word recognition.

4a) Meta-analysis of overall imageability effect

So far, we have found that imageability effect sizes differ between sets of norms and tasks. We therefore wanted to investigate the overall effect of different norms by running a separate meta-analysis for each set. This way we could further determine the variability of effect sizes for different tasks within each set of norms, which would tell us which imageability norms are most consistent, useful, and reliable as predictors of word recognition.

Method

Materials. We used effect sizes from Study 3, which were the part correlation coefficients of the imageability effect on word recognition, with lexical and sensorimotor variables partialled out. There were 36 data points, 6 per each set of norms (accuracy and RT effects on the BLP lexical decision data and the ELP lexical decision & word naming data). In order to obtain standardised effect sizes and their standard errors, we calculated[5] Fisher's *z* to use as a normalised effect size, and standard error based on word sample size. Because accuracy and RT are expected to have opposite effect sizes (better accuracy means a positive effect size, while faster RT means a negative effect size), we also calculated a 'facilitation' effect size. That is, we changed the sign of each RT effect from negative to positive and vice versa. That way, effect sizes for both accuracy and RT were predicted to be positive. Positive effects sizes thus represent facilitation of word recognition (faster or more accurate), while negative effect sizes represent inhibited (slower or less accurate) word recognition.

Design and Analysis. We ran a meta-analysis of imageability effect sizes in JASP to see whether an overall effect of imageability was present across different studies. We first ran an overall analysis with no moderators on all data points. We also wanted to investigate the overall effect of different imageability norms across tasks, by running a separate meta-

[5] Using the formula: $F_z=1/2\times((\ln(1+r))-(\ln(1-r)))$, (Silver&Dunlap, 1987) in R; code available in supplementals

analysis for each set of norms. That is, we ran 6 meta-analyses using 6 data-points each. We used a random effects model with Restricted Maximum Likelihood, which is considered a good method to estimate between study variance (Veroniki et al., 2015) when there is significant heterogeneity (Cumming, 2012).

Results and Discussion

We observed that 96% of the variance was due to heterogeneity between different analyses (I^2=96.27), which indicated that imageability effect sizes varied significantly across different word recognition studies. This suggests that there is more variance in effect size than would be expected because of measurement error if imageability had a consistent effect on word recognition. The overall expected effect size of imageability was r = 0.11, 95% CI [0.08, 0.14]; r^2 = 0.012. That is, 1.2% of variance in word recognition performance could be explained by imageability. Based on the forest plot (Figure 6), which includes expected versus observed effect sizes, the Bristol and Glasgow norms performed worse than expected, with effect sizes around or below 0 (which means that there was either no effect, or the effect was in the wrong direction: higher imageability impaired accuracy or speed of response). The other norms had mostly positive but small and inconsistent facilitation effects on all tasks.

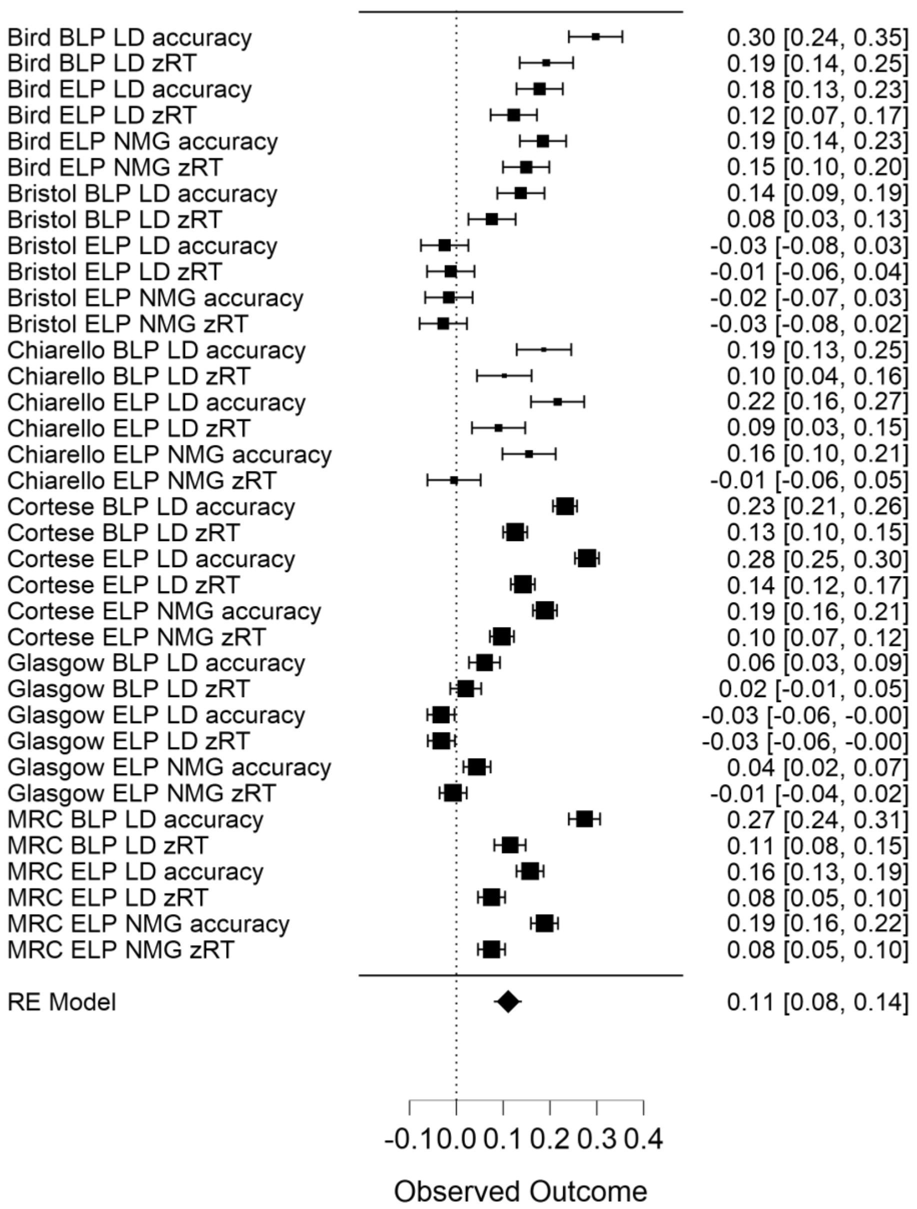


Figure 6: Forest plot of overall meta-analysis. Expected effect size calculated from the random effects model is illustrated by the diamond at the bottom of the graph.

In the per-norms analysis, there was a lot of variability in how much each set of imageability norms contributed to word recognition, with 80%-96% of heterogeneity between tasks for all norms. This suggests high variability between effect sizes, and indicates that imageability does not have a stable effect across analyses. As before, the MRC, Bird, Chiarello, and Cortese norms had some positive effects on word recognition, while the Bristol and Glasgow norms did not produce any effects. The Bird norms had the strongest effect, explaining 3.46% of word recognition, followed by the Cortese norms which explained 3.17%, but had very high heterogeneity. The Chiarello and MRC norms explained less variance, but facilitated word recognition (apart from a negative effect for the Chiarello norms on word naming RT). On the other hand, the Bristol and the Glasgow norms explained only up to 0.05% of the variance in word recognition, and their effect sizes were variable but mostly below or around 0. The meta-analysis results for each set of norms are presented in Table 8 and Figure 7.

Table 8: Between-task variance in performance calculated for each set of norms

Norms	I^2 (% between-task variance)	Meta-effect of imageability (% variance explained)
Bird	79.65	3.46
Bristol	85.99	0.05
Chiarello	86.57	1.53
Cortese	96.39	3.17
Glasgow	84.41	0.01
MRC	95.90	2.16

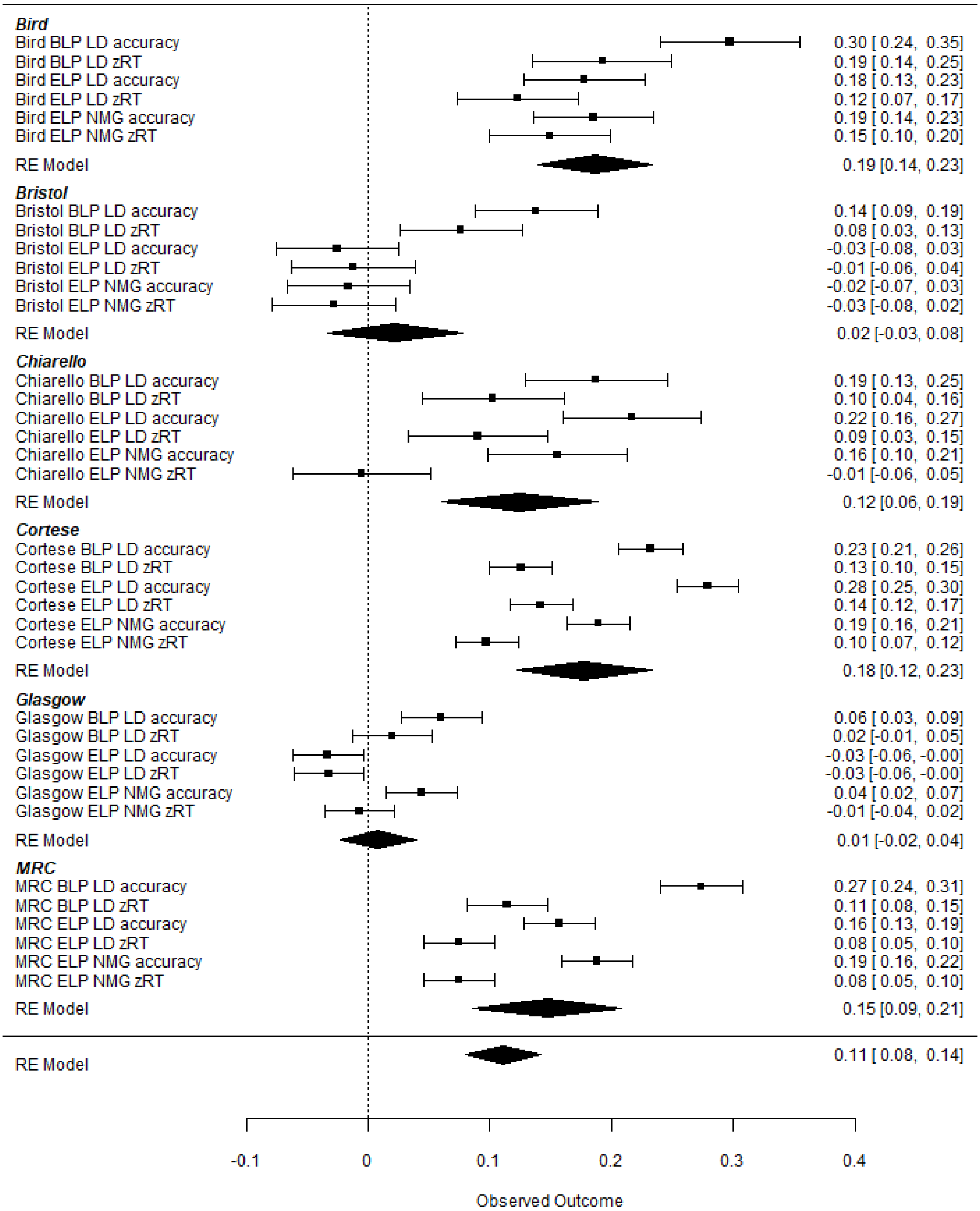


Figure 7: Forest plot of meta-analyses of each set of norms

4b) Meta-analysis of imageability effects with moderators

The observed differences between the results for different norms and tasks could be attributed to the fact that imageability has a different effect on different processes – for example, it may predict accuracy, but not speed of word recognition, or it may play a larger role in lexical decision than in word naming. Therefore, we entered a number of variables as candidate moderators in the analysis: Norms, Task type (LD/NMG), DV (accuracy/zRT), Source (BLP/ELP) and Dataset (BLP LD accuracy/zRT; ELP LD accuracy/zRT; ELP NMG accuracy/zRT). We then used the Bayesian Information Criteria (BIC), where lower BIC means better fit, to determine the best fitting model. The most parsimonious model included three moderators: norms (the 6 sets of ratings), DV (accuracy/zRT) and source (BLP/ELP). We used the same method as in Study 4a) – random effects model with Restricted Maximum Likelihood.

Results and Discussion

Compared to the meta-analysis in 4a, the heterogeneity between effect sizes decreased to 80% (I^2=79.90), indicating that inclusion of moderators accounted for some of the differences between the effects of imageability depending on norms and tasks.

We chose the MRC norms as the regression intercept, that is, the baseline imageability effect to which other norms are compared, as they are the oldest set of norms and have been used most commonly. The intercept represents a fixed level for all moderators, specifically the BLP and accuracy, to which the effects of the ELP and RT are compared. The results are presented in Table 9:

Table 9: Coefficients of meta-regression of moderating effects of imageability effect sizes

Moderator	Coefficient	SE	z	*p*
Intercept (MRC, BLP, acc)	0.225	0.020	11.065	< .001
Bird norms	0.040	0.025	1.615	.106
Bristol norms	-0.125	0.024	-5.127	< .001
Chiarello norms	-0.023	0.025	-0.907	.365
Cortese norms	0.030	0.023	1.341	.180
Glasgow norms	-0.138	0.023	-6.049	< .001
ELP data	-0.058	0.015	-3.859	< .001
RT data	-0.079	0.014	-5.610	< .001

Note: The intercept is MRC effect on BLP accuracy data

Table 9 shows that compared to the MRC effects on lexical processing, the Bird, Chiarello and Cortese norms performed the same ($p > .05$), while the Bristol and the Glasgow norms had a smaller effect than MRC. These differences emerged when the effects of the other two moderators had been taken into account. Further, effects of imageability were larger for accuracy than for speed of processing, and for the BLP than the ELP data.

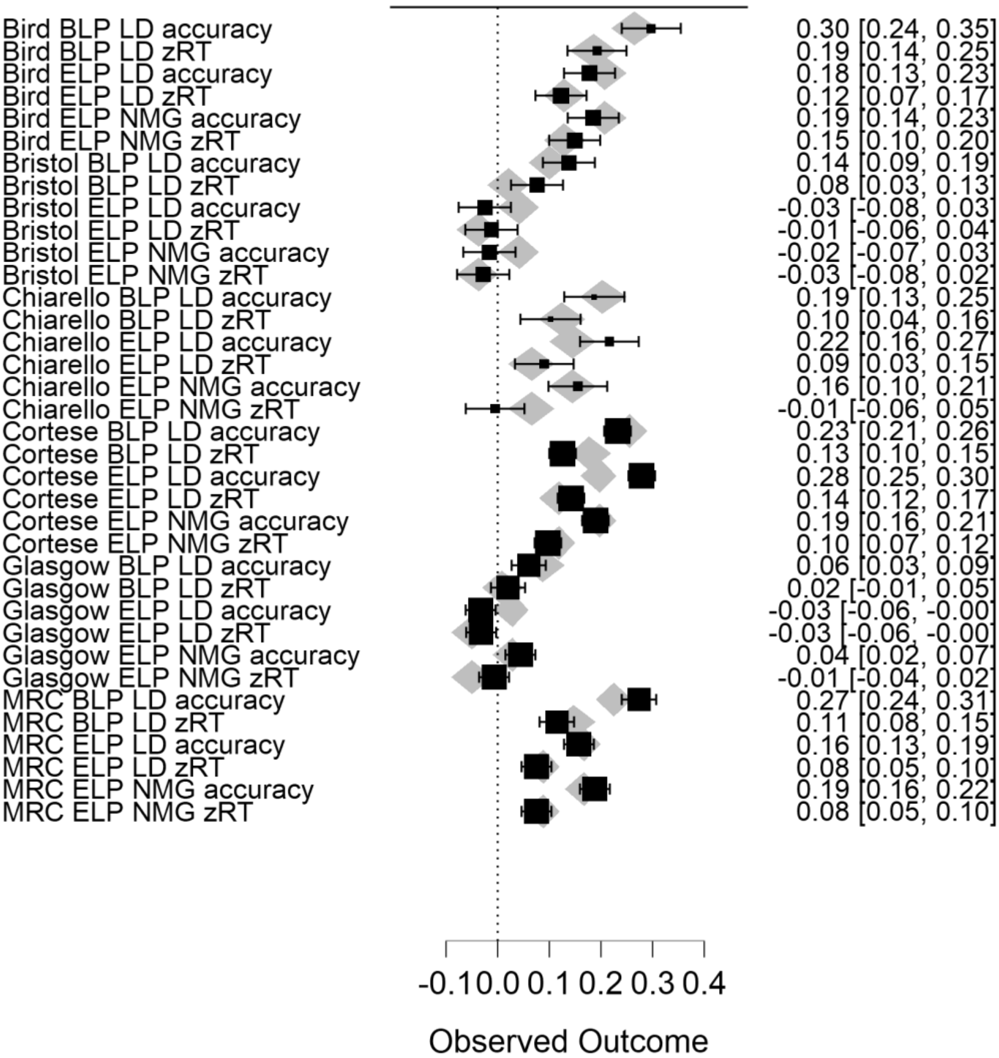


Figure 8: Forest plot of meta-analysis effect with moderators. Expected effect size is illustrated by shaded diamonds next to each observed effect size.

This time, the forest plot (Figure 8) shows different expected effect sizes, depending on the norms and DV. When the variance from norms, DV and data source was taken into account, most of the effect sizes showed regression to the mean. While most norms more or less adhered to their expected effect sizes, the Bristol and Glasgow norms consistently showed the lowest expected effect size, mostly equal to or just below 0, which means that, as in Studies 1 and 3, their effect size was in the wrong direction compared to what previous imageability literature had shown. Additionally, the expected effect sizes of the Cortese norms were closer to 0 than the observed positive effects, that is, based on the estimates from other studies, the Cortese norms were overperforming as a predictor. The variability was also evident between tasks: the Cortese norms had an unexpectedly large effect size on ELP lexical decision accuracy, while the effect of the Chiarello norms were most variable; higher than expected on the ELP lexical decision accuracy, but lower than expected on the ELP naming speed.

The results suggest that even when variance between different sources of ratings and different tasks is taken into account, imageability studies, with the same instructions, produce different ratings that sometimes capture a useful variable contributing to word recognition, but other times do not facilitate word recognition at all. Thus, at least some of the existing imageability norms do not reliably measure imageability.

Conclusion

We conducted an internal meta-analysis to assess the overall effects and reliability of imageability as a predictor of word recognition. The results corroborate our findings from Studies 1-3: different sources of imageability produce highly divergent ratings, which vary in their contribution to word recognition, with some norms performing well, but others performing much worse than in the previous literature. This suggests that imageability norms do not reliably measure the ease of creating a mental image, and that the construct of imageability cannot reliably predict word processing.

A full meta-analysis of the effects of imageability across sets of norms and tasks found that overall imageability explained only 1.2% of variance in word recognition, and that there was a high amount of variability between different sets of norms and tasks. When three moderating variables were included, heterogeneity decreased, but still some norms performed better than others, and some tasks were predicted better than others. Separate analyses of each set of imageability ratings further revealed that all norms had high heterogeneity between tasks, and that the contribution of each set to word recognition varied between 0.01% to 3.46% - which was a small effect at best, and a null effect at worst.

The overall pattern of results supported the findings from Studies 1 and 3: The Bird, Cortese and MRC norms performed a bit more consistently than others and had a small effect on word processing. The Bristol and Glasgow norms were quite variable and had no effect, and the Chiarello norms had variable effects depending on the task and measure. Finally, it is worth noting that while the Cortese norms elicited strong effects, they were also one of the most inconsistent, and therefore their reliability may have been overestimated in the original study, where the ELP lexical decision data was used to validate the ratings. It appears that even when imageability is normed using standard instructions and procedure, the participants who take part in the norming may do unpredictably different things when generating their imageability ratings. This finding, together with the inconsistencies found for other imageability norms, calls for a more rigorous, multi-faceted process of validating imageability ratings (or indeed any other lexical or semantic variable), if it is to be continued to be used reliably.

General discussion

Across four studies, we investigated the concept of imageability, what it measures and whether it reliably contributes to word recognition, as found in previous research (e.g., Bennett et al., 2011; Cortese & Schock, 2013, Sadoski & Paivio, 2004). Previous studies suggested that imageability facilitates word recognition by capturing semantic information associated with the word. However, the effect sizes, and even the direction of the effects, vary between different studies. We conducted a number of analyses to compare imageability norms from different sources in their contribution to word recognition, and to investigate the sources of variance between the norms. To our knowledge, this is the first study to conduct a comprehensive investigation into the reliability of imageability as a construct.

In Study 1, we compared 6 sets of imageability ratings using an empirically derived lexical baseline model, to eliminate any differences that were due to methodological variability between studies. We found that imageability performance in predicting word recognition still varied significantly between the datasets. This was especially the case for low frequency words, for which imageability effects were expected to be strongly facilitatory (Brysbaert et al., 2018; Raman & Baluch, 2001), but where the effects varied significantly, from facilitating to inhibiting processing, depending on the set of words or set of ratings used. In an attempt to determine the source of this variance, in Study 2 we conducted a principal components analysis to identify 6 principal components, which we then used to predict the imageability ratings; we found that 32%-57% of variance in imageability ratings was predicted by these components, with the 'Object' component, comprising haptic, visual, and hand action strength, combined sensorimotor strength, and part of speech (noun/non-noun), contributing most consistently. This was in line with previous literature suggesting that imageability ratings are particularly associated with nouns (Bedny & Thompson-Schill, 2006; Bird et al., 2001), since nouns often denote manipulable objects which are potentially the easiest to visualise, and suggests that when rating imageability participants generate sensorimotor simulations associated with objects and their use. However, this still varied between norms, and only accounted for up to half of the variance in imageability ratings.

The variability in imageability effects could be attributed to differences between the words used in the norms, or to the fact that participants in different norming studies may have considered different types of information when rating word imageability. Therefore, in Study

3, we examined whether imageability ratings capture the semantic content of words, measured as sensorimotor information associated with experiencing the concept. We found that when sensorimotor variables were partialled out, imageability effect sizes varied significantly between different norms, suggesting that the ratings did not capture the variance associated with sensorimotor content of the word, but perhaps something unrelated. Some ratings, namely the Bristol and Glasgow norms, consistently predicted very little variance in word processing over lexical and sensorimotor variables, while other norms actually predicted slower and less accurate word recognition. These effects of pure imageability, independent of word variation, were in contrast with the theoretical role of imageability in word recognition (Sadoski & Paivio, 2004).

The pattern of results suggests that imageability does not capture semantic content of the word, or at least it does not do so reliably. In other words, only the norms which performed well as predictors of word recognition appear to measure something which is useful in word processing. However, there are theoretical limits to what ease of generating imagery can contribute to word processing. Our analysis indicated that imageability refers to consciously generated, mostly visual information, which tends to involve manipulable objects. On the other hand, representing conceptual information through the sensorimotor system can be unconscious, involves multiple senses and modalities beyond vision, and can be applied to other types of concepts such as verbs and adjectives, or abstract concepts (Connell & Lynott, 2014; Pecher et al., 2009). Imageability should therefore not be treated as a placeholder for sensorimotor word content, or a catch-all variable responsible for word meaning. Additionally, when only some of the norms are able to capture something over and above sensorimotor word content, despite using the same method to obtain the ratings, then imageability as a concept cannot be trusted to be reliably tested.

There were some effects of imageability in Study 3 (the Cortese norms accounted for 7.4% of unique variance for ELP lexical decision accuracy, and the Bird norms explained 8.4% of BLP lexical decision accuracy), which was closer to the effects found in the literature (e.g. Bennett et al., 2011; Cortese & Schock, 2013). Nonetheless, these results were the exception, rather than the norm, and most imageability norms did not explain much of the variance in word recognition. Since the current study was able to control for variance associated with the use of different words and different baseline lexical and semantic variables, the variability in effect sizes can only be attributed to the differences in imageability ratings. Further, if this variability was due to participants in some norming studies doing better at generating mental imagery, or because the words in some norms were easier to rate, we would expect each set of norms to be internally consistent at predicting word recognition across different tasks, despite variability across imageability norms. This was not the case, as for example the Bristol norms had a small effect on the BLP lexical decision task, but no effects on the ELP tasks. Alternatively, the variability in the results could be due not to the role of imageability, but perhaps another variable which is more important and to which the effect can be attributed. We controlled for word frequency, eliminating the possibility that the differences were due to the participants' experience with different words. We also took into account semantic content of the words by using sensorimotor, including visual, strength as a semantic variable in Study 3, so the differences cannot be attributed to the visual nature of the concept of imageability. While we accounted for motor strength in our analyses, perhaps the body-object interaction variable, which measures how easy it is for a human body to interact with an object, could tap into an additional aspect of imageability contribution, due to its complementary nature to imageability (Bennett et al., 2011), and in particular could account for the consciously available information about objects. However, to our knowledge most BOI ratings focus on concrete words, which have high imageability ratings, and this makes the measure less useful for this type of comprehensive investigation. Alternatively, emotional valence may play a role in semantic processing: emotional experiences are processed faster (Kousta et al., 2009) and that can make words high in emotional valence easier to recognise (Citron et al., 2014, Westbury et al., 2013) and easier to generate an image of and thus confound with imageability. However, some of that emotional experience should be accounted for by the sensorimotor strength measure included in the present study (cf. Warriner et al., 2013), particularly interoceptive strength (Connell et al., 2018).

Overall, the results do not support the claim that ease of generating a mental image reliably predicts word recognition. A meta-analysis suggested that an overall imageability effect on word recognition was very small (1.2%) compared to previous effects reported in the literature. We found that the effects of imageability varied between different sets of norms and tasks, and heavily dependent on the combination of which imageability norms were used with which task (lexical decision or word naming from BLP or ELP). This variation cannot be explained by differences in word sampling between norming studies. Since all studies used the same instructions and procedure, the variance could be attributed to differences in participant strategies when rating ease of imageability. This means that imageability is not a reliably measured construct. Instead, other variables, such as sensorimotor strength, are more reliable predictors of word recognition, in part because they capture semantic information automatically, and thus should be used in future research.

References

Adelman, J. S., Brown, G. D. A., & Quesada, F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823.

Bakhtiar, M., & Weekes, B. (2014). Lexico-semantic effects on word naming in Persian: Does age of acquisition have an effect? *Memory and Cognition*, *43*(2), 298–313. https://doi.org/10.3758/s13421-014-0472-4

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283-316. https://doi.org/10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Nelson, D. L., Simpson, G. B. & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445-459. https://doi.org/10.3758/bf03193014

Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain and Language*, *98*(2), 127–139. https://doi.org/10.1016/j.bandl.2006.04.008

Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., & Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic nouns. *Behavior Research Methods*, *43*(4), 1100–1109. https://doi.org/10.3758/s13428-011-0117-5

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*(1), 73–79. https://doi.org/10.3758/BF03195349

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*(2), 215–226. https://doi.org/10.1016/S0001-6918(00)00021-4

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin and Review, 8*(3), 531–544. https://doi.org/10.3758/BF03196189

Chedid, G., Brambati, S. M., Bedetti, C., Rey, A. E., Wilson, M. A., & Vallet, G. T. (2019). Visual and auditory perceptual strength norms for 3,596 French nouns and their relationship with other psycholinguistic variables. *Behavior Research Methods*, *51*(5), 2094–2105. https://doi.org/10.3758/s13428-019-01254-w

Chiarello, C., Shears, C., & Lund, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 603–637. https://doi.org/10.3758/BF03200739

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2014). Arousal and emotional valence interact in written word recognition. *Language, Cognition and Neuroscience*, *29*(10), 1257–1267. https://doi.org/10.1080/23273798.2014.897734

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 371–383. https://doi.org/10.3758/BF03195584

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33*(4), 497-505. https://doi.org/10.1080%2F14640748108400805

Connell, L. (2018). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience, 34*(10), 1308-1318. https://doi.org/10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2012a). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*(3), 452–465. https://doi.org/10.1016/j.cognition.2012.07.010

Connell, L., & Lynott, D. (2012b). When does perception facilitate or interfere with conceptual processing? The effect of attentional modulation. *Frontiers in Psychology*, *3*, 1–4. https://doi.org/10.3389/fpsyg.2012.00474

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin and Review*, *20*(3), 542–550. https://doi.org/10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014). I see/hear what you mean: Semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General*, *143*(2), 527–533. https://doi.org/10.1037/a0034626

Connell, L., & Lynott, D. (2016a). Do we know what we're simulating? Information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1218–1232. https://doi.org/10.1037/xlm0000245

Connell, L., & Lynott, D. (2016b). Embodied semantic effects in visual word recognition. In M. H. Fischer & Y. Coello (Eds.), *Foundations of embodied cognition* (Vol. 2, pp. 71–92). Hove, UK: Psychology Press.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 384–387. https://doi.org/10.3758/BF03195585

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, *60*(8), 1072-1082.

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609. https://doi.org/10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*, *68*(8), 1489–1501. https://doi.org/10.1080/17470218.2014.945096

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, *66*(5), 946–972. https://doi.org/10.1080/17470218.2012.722660

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review*, *4*(2), 226-231.

Cumming, G. (2012). *Understanding the new statistics effect sizes, confidence intervals, and meta-analysis* (1st ed., Multivariate applications book series). New York: Routledge.

Davis, C. J., & Stadthagen-Gonzalez, H. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598–605.

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 824–845. https://doi.org/10.1037/0278-7393.15.5.824

Evans, G. A. L., Ralph, M. A. L., & Woollams, A. M. (2012). What’s in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin and Review*, *19*(2), 325–331. https://doi.org/10.3758/s13423-011-0213-7

Fernandino, L., Humphries, C. J., Seidenberg, M. S., Gross, W. L., Conant, L. L., & Binder, J. R. (2015). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia*, *76*, 17–26. https://doi.org/10.1016/j.neuropsychologia.2015.04.009

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*(4), 395-427. https://doi.org/10.3758/BF03201693

González-Nosti, M., Barbón, A., Rodríguez-Ferreiro, J., & Cuetos, F. (2014). Effects of the psycholinguistic variables on the lexical decision task in Spanish: A study with 2,765 words. *Behavior Research Methods*, *46*(2), 517–525. https://doi.org/10.3758/s13428-013-0383-5

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718-730. https://doi.org/10.3758/s13428-012-0278-x

JASP Team (2020). *JASP (Version 0.14.1)[Computer software]*.

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology*, *64*(9), 1683–1691. https://doi.org/10.1080/17470218.2011.605150

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287-304. https://doi.org/10.3758/s13428-011-0118-4

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978-990.

Lenci, A., Lebani, G. E., & Passaro, L. C. (2018). The emotions of abstract words: A distributional semantic analysis. *Topics in Cognitive Science*, *10*(3), 550–572. https://doi.org/10.1111/tops.12335

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2019). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods, 52,* 1-21.

Paivio, A. (1971). Imagery and language. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches.* (pp. 7-32). New York: Academic Press.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1), 1-12. https://psycnet.apa.org/doi/10.1037/h0025327

Pecher, D., van Dantzig, S., & Schifferstien, H. N. J. (2009). Concepts are not represented by conscious imagery. *Psychonomic Bulletin and Review*, *16*(5), 914–919. https://doi.org/10.3758/PBR.16.5.914

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin and Review*, *15*(1), 161–167. https://doi.org/10.3758/PBR.15.1.161

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Raman, I., & Baluch, B. (2001). Semantic effects as a function of reading skill in word naming of a transparent orthography. *Reading and Writing*, *14*(7-8), 599-614.

Reilly, M., & Desai, R. H. (2017). Effects of semantic neighborhood density in abstract and concrete words. *Cognition*, *169*, 46–53. https://doi.org/10.1016/j.cognition.2017.08.004

Sadoski, M. & Paivio, A. (2004). A dual coding theoretical model of reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1329-1362). Newark, DE: International Reading Association.

Sadoski, M., & Paivio, A. (2012). *Imagery and text : A dual coding theory of reading and writing.* Routledge.

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*(2), 374–379. https://doi.org/10.3758/s13428-011-0162-0

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2018). The Glasgow norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, *51*(3), 1258-1270. https://doi.org/10.3758/s13428-018-1099-3

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods*, *42*(2), 393–413. https://doi.org/10.3758/BRM.42.2.393

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used?. *Journal of Applied Psychology*, *72*(1), 146-148. https://psycnet.apa.org/doi/10.1037/0021-9010.72.1.146

Simonsen, H. G., Lind, M., Hansen, P., Holm, E., & Mevik, B. H. (2013). Imageability of Norwegian nouns, verbs and adjectives in a cross-linguistic perspective. *Clinical Linguistics and Phonetics*, *27*(6-7), 435–446. https://doi.org/10.3109/02699206.2012.752527

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vergallito, A., Petilli, M. A., & Marelli, M. (2020). Perceptual modality norms for 1,121 Italian words: A comparison with concreteness and imageability scores and an analysis of their impact in word processing tasks. *Behavior Research Methods*, *52*(4), 1599-1616. https://doi.org/10.3758/s13428-019-01337-8

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*(1), 55–79. https://doi.org/10.1002/jrsm.1164

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://www.jstatsoft.org/v36/i03/.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(02), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Westbury, C., & Moroschan, G. (2009). Imageability x phonology interactions during lexical access: Effects of modality, phonological neighbourhood, and phonological processing efficiency. *The Mental Lexicon*, *4*(1), 115–145. https://doi.org/10.1075/ml.4.1.05wes

Westbury, C. F., Shaoul, C., Hollis, G., Smithson, L., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2013). Now you see it, now you don’t: On emotion, context, and the algorithmic prediction of human imageability judgments. *Frontiers in Psychology*, *4*, 1–13. https://doi.org/10.3389/fpsyg.2013.00991

Wickham, H., Averick, M., Bryan, J., Chang, W., D’Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Milton Bache, S., Muller, K., Ooms, J., Robinson, D., Paige Seidel, D., Spinu, V. ... &

Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6-10.

Wilson, M. A., Cuetos, F., Davies, R., & Burani, C. (2013). Revisiting age-of-acquisition effects in Spanish visual word recognition: The role of item imageability. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(6), 1842–1859. https://doi.org/10.1037/a0033090

Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(5), 878–890. https://doi.org/10.1037/0278-7393.31.5.878

Yap, M. J., Lim, G. Y., & Pexman, P. M. (2015). Semantic richness effects in lexical decision: The role of feedback. *Memory and Cognition*, *43*(8), 1148–1167. https://doi.org/10.3758/s13421-015-0536-0

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*, 1–10. https://doi.org/10.3389/fnhum.2012.00072

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(6), 1385–1397. https://doi.org/10.1037/0278-7393.31.6.1385

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*(1), 1-29.

6 More is Not Necessarily Better – How Different Aspects of Sensorimotor Experience Affect Recognition Memory for Words

In Chapter 4 I examined working memory using the linguistic-simulation perspective, and in Chapter 5 I found that imageability is not a reliable predictor of word recognition over and above lexical and sensorimotor information. This allows me to turn to long-term semantic memory, and use sensorimotor strength ratings and imageability ratings to investigate whether word memory relies on sensorimotor grounding, as well as conscious imagery. More specifically, I will examine what kind of sensorimotor information associated with a concept might be activated when the concept is encountered in a word memory task, and hence supports a detailed memory trace which is easier to recognise. Additionally, I will investigate whether making a conscious effort to remember a concept will play a role in word recognition memory performance. Even though Chapter 5 demonstrated that imageability ratings are too variable to be reliably useful in word recognition, it is still possible that they are useful to memory. When generating mental imagery is used as a strategy to memorise words, imageability could facilitate an expected memory task. If, however, strength of sensorimotor information enhances memory trace, then in both surprise and expected memory tasks words should benefit from sensorimotor strength.

More is Not Necessarily Better – How Different Aspects of Sensorimotor Experience Affect Recognition Memory for Words

Agata Dymarska, Louise Connell, and Briony Banks

Department of Psychology, Lancaster University.

Author Note

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 682848) to LC.

Correspondence concerning this article should be addressed to Agata Dymarska, Department of Psychology, Fylde College, Lancaster University, Lancaster LA1 4YF, UK. Email: a.dymarska@lancaster.ac.uk

 Stimuli, data and analysis code can be found at: https://osf.io/rjdng/?view_only=7ce5801c9eeb4a08a7e678724fbc9b1a

Abstract

We investigated the contribution of semantic information to word memory using imageability and sensorimotor strength as predictors. Semantic richness theory predicts that more semantic information should facilitate performance on a memory task. However, sensorimotor strength represents a multi-dimensional experience with a concept, while imageability focuses on consciously available information biased towards visual experience, and therefore they could show diverging effects. Data from a mega-study of word recognition memory (Cortese et al., 2010; 2015), as well as from an online memory task, was analysed in a series of hierarchical linear regressions. Both sensorimotor strength and imageability had an effect on word memory performance, but not as strong as reported in previous literature. However, the effects were smaller when the memory task was unexpected, suggesting that the semantic effects are dependent on memory strategies (or context). Most importantly, we found that sensorimotor strength had varying effects on different memory measures, which was not in line with the prediction of the semantic richness theory. The results highlight the importance of a multi-dimensional approach to measuring and testing semantic experience, and its effect on cognitive processing. The findings have implications for the use of semantic variables in memory research.

Keywords: word memory; sensorimotor information; imageability; semantic richness

More is Not Necessarily Better – How Different Aspects of Sensorimotor Experience Affect Recognition Memory for Words

Memory for words, both immediate and long-term, relies on their lexical features and semantic content. The stronger a semantic variable (i.e., a variable measuring the representation of meaning which underlies a word's referent, such as number of features or strength of sensory experience) measuring the representation of meaning which underlies a word's referent, is for a particular word, the easier it is to recognise or to remember that word. As proposed by the semantic richness theory (Buchanan et al., 2001; Pexman et al., 2008), encountering a concept with a richer semantic representation activates a larger number of neuronal connections, which then facilitates processing and eliciting a response to the word. This is best demonstrated in the lexical decision literature: for example, the number of semantic neighbours, number of features, and body-object interaction have predicted performance in visual word recognition tasks in concrete concepts (Pexman et al., 2008; Recchia and Jones, 2012; Yap et al., 2011), while for abstract concepts, faster lexical decision task responses are predicted by contextual information (Zdrazilova & Pexman, 2013) or the number of semantic neighbours (Recchia and Jones, 2012). Words higher in sensory experience are also faster to process in a lexical decision task (Juhasz et al., 2011) and a semantic categorisation task (Zdrazilova & Pexman, 2013), and the same has been found for words rated high on emotional valence (Kousta et al., 2009; Zdrazilova & Pexman, 2013). The semantic information used in word processing is stored in semantic memory (e.g., Pexman et al., 2008; Kounios et al., 2009), and the facilitation effect is also present in memory research: word recall is facilitated by availability, which measures how often a word is given as an associate to another stimulus word (Rubin & Friendly, 1986), by emotional valence (Long et al., 2015), animacy (Madan, 2020; Nairne et al., 2013), number of senses and features, arousal, and body-object interaction (Lau et al., 2018), as well as imageability

(Rubin & Friendly, 1986; Lau et al., 2018). Similarly, word recognition memory performance is facilitated by emotional valence and arousal (Snefjella & Kuperman, 2016; Lau et al., 2018). These semantic effects are varied, but they all elicit facilitation effects in the same direction, in line with the semantic richness theory.

The most commonly investigated semantic predictor is imageability, which has received considerable attention due to its critical role in conceptual processing, according to the dual-coding theory (Paivio, 1971, 1990). The theory states that information about word meaning is stored using two codes: all concepts are encoded with language, but concrete concepts also benefit from an additional sensory image, which makes them easier to read, process and remember, simply because it provides more information. The ease of generating or activating the mental image associated with the concept can be defined and measured as imageability, which is sometimes used interchangeably with concreteness ratings (i.e., is it concrete/touchable/experienced with the senses). The role of imageability and the dual-code advantage in memory was evidenced by research where pairs composed of high-imageability words were remembered better than pairs composed of low-imageability words (Paivio, 1969; 1971). Other studies have also found that higher imageability words were better remembered in a recognition memory task (Fliessbach et al., 2006), as well as in verbal and written recall (Bourassa & Besner, 1994; Majerus & Van der Linden, 2003; Tse & Altarriba, 2007). Additionally, Cortese et al. (2010; 2015) found that imageability was the strongest predictor of word recognition memory, even stronger than lexical variables such as frequency or phonological and orthographic neighbourhood. The role that generating a conscious image plays in memory performance was also shown by Hamilton and Rajaram (2001) who found that when participants were asked to create an image of a word’s referent, they were later more likely to use that word in a conceptual knowledge task, compared to when they had just read the word passively. Similarly, Takashima et al. (2013) found that learning novel words through associated images, as well as their phonological form, resulted in better memory for the words 24 hours later than when learning only the phonological form. Additionally, using fMRI, they observed greater activity for the dual-coded items in the hippocampus, which is critically involved in memory formation. Other neuroimaging studies found greater hippocampal activity during study and test of high- compared with low-imageability word pairs (Caplan & Madan, 2016), and a larger positive wave in EEG (P600, which typically reflects syntactic and semantic anomaly processing: van Herten et al., 2005) while processing higher imageability words (Klaver et al., 2005). Both dual-coding theory and semantic richness theory propose that semantic information facilitates more accurate retrieval of information about a word and its referent. Evidence suggests that imageability contributes to word memory for this reason.

However, the concept of imageability is not without its problems. A comprehensive analysis of imageability norms for 9796 words from different sources (Dymarska et al., 2021; see Chapter 5) found that imageability ratings reflect a lot of noise and inter-participant and inter-norms variability. Critically, after an extensive range of other sources of variance is accounted for, ease of generating a conscious image is a weak and inconsistent predictor of word recognition, which suggests that previous evidence for apparently large imageability effects were an artefact of small word samples or fewer psycholinguistic control variables. Similar inconsistencies, though not comprehensively studied, can be found in the memory literature: for example, Cortese et al. (2010; 2015) found that imageability explained 14%-24% of variance in word recognition memory, but Lau et al. (2018) found no effect of imageability on recognition memory, only on free recall. It is not clear whether these inconsistencies may be due to variability in word samples, testing methods, or psycholinguistic controls. Nonetheless, unlike in word recognition tasks where semantic access is automatic and implicit, imageability may be a useful strategy in a memory task, because memory requires conscious retrieval of the particular instance when the word was encountered or studied. Generating conscious imagery for words encountered both at encoding and retrieval makes it easier to identify previously seen words, which also makes easy-to-image words better remembered. It is therefore important to examine the role of conscious imagery in memory in a more comprehensive manner in order to establish the robustness (or otherwise) of imageability effects.

In addition to variability in imageability effects, another problem with the concept of imageability, as well as the semantic richness theory, is the way they conceptualise word meaning. Much of the imageability and concreteness literature (Klaver et al., 2005; Paivio, 1971, 1990; Miller & Roodenrys, 2009; Walker & Hulme, 1999) define word meaning based on two main types of information: language-based and image-based. However, this distinction does not fully capture the complex nature of semantic representations, which rely on information from multiple sensory and motor modalities. For example, words such as “music” or “taste” are rated low in imageability, but high on experience via the senses, which makes them faster to recognise in a lexical decision task (Connell & Lynott, 2012a). In the semantic richness literature, meaning is referred to as any semantic variable which facilitates processing with an additive effect if multiple variables are at play. However, the effects of different variables seem to be task dependent, and the theory does not address their individual roles. For example, Hamilton and Rajaram (2001) found the semantic richness effect in a free recall task, but not in a number of other tasks, such as word fragment completion (where participants had been exposed to a word, e.g., “elephant”, and later had to fill the gaps in “_le_p__n_”). Similarly, Yap et al. (2011) found an effect of the number of senses (that is, how many meanings a word has, e.g., “board” can be interpreted as an object or as a committee) in lexical decision, but not in a semantic categorisation task. While the disparate effect of semantic variables in different tasks does not invalidate the concept of semantic richness, it reveals its incomplete definition of meaning, which refers to a number of disconnected characteristics that words happen to have, which are sometimes useful in conceptual processing and are linked to meaning in some ways, but not to each other (e.g., number of features, emotional valence, semantic neighbourhood).

A more comprehensive view of conceptual representations is offered by the idea that concepts are grounded in either physical, sensory and motor experience, or the experience of internal states, emotional states, or situational context (e.g., Connell et al., 2018; Wiemer-Hastings & Xu, 2005). This means that conceptual knowledge is stored and represented using similar neural pathways as when the concept was encountered, for example, those associated with sensory, motor or emotional experience which accompanies the concept. Based on this, conceptual representations can be operationalised as sensory experience (SER, Juhasz et al., 2011), body-object interaction strength (Yap et al., 2016) or sensorimotor strength (Lynott et al., 2019). We focus here on sensorimotor strength, as it is more nuanced and comprehensive than other measures because it includes separate ratings of 11 sensorimotor dimensions that are processed in discrete areas of the sensory and motor cortices. Sensorimotor information has also been shown to be a better predictor of lexical processing than sensory experience ratings (Connell & Lynott, 2016a) or imageability (Connell & Lynott, 2012), and it has been shown to guide behaviour without conscious awareness (Pecher et al., 2009), unlike imageability, which concerns the ease of consciously generating a mental image (Paivio et al., 1968). There is reason to believe that sensorimotor information should facilitate word memory, in line with the semantic richness effect: although support for the role of motor information in memory for words and objects is mixed (Zeelenberg & Pecher, 2016), Sidhu and Pexman (2016) found that words associated with a greater amount of body experience were recalled and recognised more accurately than words rated low in body experience. Sensorimotor information was also found to support memory for manipulable objects

(Dutriaux et al., 2018). When participants were unable to simulate actions performed with their hands, such as “take the cup”, later cued recall of “cup” was impaired. Further, sensorimotor information can be measured on a number of dimensions (senses and body parts – see Lynott et al., 2019), giving a more nuanced insight into what semantic information about the concept consists of. While these nuanced dimensions (e.g., sensations inside the body, Connell et al., 2018) affect word recognition (Connell & Lynott, 2012; Lynott et al., 2019), they have not been comprehensively examined for their effects on memory. For example, Sidhu and Pexman (2016) only tested the effects of body-object interaction on memory for verbs. Thus, using a nuanced measure of sensorimotor strength may provide new insight into semantic richness effects in memory.

Current Studies

In the current series of studies, our aim was to examine semantic effects on word memory, using imageability and sensorimotor strength as predictor variables in a number of exploratory analyses. More specifically, we wanted to test whether higher sensorimotor strength offers an advantage regardless of dimension, as would be predicted by the semantic richness effect. We investigated this issue in Study 1 using a megastudy approach (Balota et al., 2012) of hierarchical regression analysis of 5311 words from the combined word recognition memory datasets from Cortese et al. (2010) and Cortese et al. (2015), with sensorimotor strength measures derived from the Lancaster Sensorimotor Norms (Lynott et al., 2019) as predictors.

Additionally, given the unstable nature of imageability effects in visual word recognition (Dymarska et al., 2021), we wanted to investigate whether the effect of imageability on word memory is stable across different sources of imageability and whether it is due to sensorimotor grounding (i.e., the perceptual and action experience of the word’s referent) or imageability per se (i.e., the *ease of consciously generating a mental image* of the word's referent). If the effect is purely due to sensorimotor grounding captured in ratings of perceptual and action experience, then imageability should not have an effect on word memory over and above sensorimotor strength. However, if ease of generating a mental image offers an additional benefit to word memory, then imageability will have an effect on word memory even when the effects of sensorimotor grounding have been accounted for. We tested these questions in Study 2 using imageability ratings from 6 different sources (all based on the same norming instructions), on top of multidimensional measures of sensorimotor strength, as predictors of the same word memory recognition dataset used in Study 1.

Finally, we wanted to investigate whether participant strategies involved in different forms of word memory tasks influenced the effects of semantic variables on performance. In Study 3, we compared data from a surprise memory task (details in supplemental materials) and the same words from Cortese et al.'s (2010; 2015) expected memory task. We tested the effects of sensorimotor strength and imageability on performance in these two types of tasks. We expected that imageability would have a stronger effect on performance when participants expect to be tested on their memory for words, because consciously generating imagery can be used as a strategy to enhance the strength of the memory trace. On the other hand, sensorimotor grounding is automatic and its effects should be relatively consistent across the two different types of tasks.

Study 1a: Different Forms of Sensorimotor Experience

The aim of Study 1 was to test whether the strength of sensorimotor experience influences word recognition memory performance. We used data from existing megastudies of word recognition memory (Cortese et al., 2010; 2015) and sensorimotor strength ratings from the Lancaster Sensorimotor Norms (Lynott et al., 2019). Since the norms comprised information about 11 sensorimotor dimensions (six perceptual modalities and five action effectors), and we also wanted to control for many lexical, sublexical, and lexico-semantic variables, we used a principal components analysis (PCA) to collapse this large number of variables into the most important components of lexical and sensorimotor information. In addition, since the 11 sensorimotor dimensions are highly correlated with each other (Lynott et al., 2019) and with lexical variables such as frequency and length (Lynott & Connell, 2013; Dymarska et al., 2021, Chapter 5 of this thesis), PCA offered a useful way to segregate lexical and sensorimotor information into a set of uncorrelated component variables that could then be used as predictors in regression analyses without risk of multicollinearity.

Based on the semantic richness view (Buchanan et al., 2001; Pexman et al., 2008), we predicted that higher sensorimotor strength, regardless of dimension, would improve performance on a word recognition memory task. Specifically, we expected that if a word is more strongly associated with perceptual experience (vision, hearing, taste, smell, touch, and interoception; e.g., book) or motor action (using the leg/foot, hand/arm, torso, head, or mouth to experience a concept; e.g., football), encountering it in a memory task would benefit from a stronger memory trace than a word which is rated low in sensorimotor experience (e.g., enzyme), and therefore better performance, due to the rich semantic representation resulting from the consolidated sensorimotor simulation in which the word meaning is grounded. Based on the patterns of effects previously observed for imageability in similar tasks (Cortese et al., 2010, 2015), we expected higher sensorimotor strength of a target word to lead to higher hit rates, lower false alarm rates, better discrimination of old and new items, and a more liberal response bias.

Method

Materials

Items comprised a total of 5311 words[6]. Dependent measures came from two megastudies of word recognition memory, one focusing on monosyllabic words (Cortese et al., 2010) and one on disyllabic words (Cortese et al., 2015). In both studies, participants were asked to study lists of 50 words at a time for a later recognition task, which took the form of an old/new judgement on each target word presented (i.e., half the targets were old and half new). These data provided 5 measures of memory performance per word: Hit Rate (HR: how many items are correctly recognised as previously seen); False Alarm Rate (FA: how many items are incorrectly recognised as previously seen); Hit Rate minus False Alarm Rate (HR-FA: a common composite measure of word memory performance); *d'* (sensitivity: how well are old items distinguished from new); and *c* (criterion or response bias: how strong is the overall tendency to respond "old" or "new" to all items).

As predictor variables, we used 6 PCA components obtained in another study to consolidate lexical and sensorimotor information (Dymarska et al., 2021), and we summarise the method of extracting the components here for the benefit of the reader. The item set for the PCA was based on 9796 words used in the analysis of imageability in Dymarska et al. (2021). Variables used for the PCA are detailed in Table 1, and included a variety of sublexical (e.g., orthographic and phonological neighbourhoods), lexical (e.g., word length and frequency measures), and lexico-semantic (e.g., age of acquisition, linguistic distributional distance) properties that impact on word processing. In addition, the PCA incorporated 11 dimensions of sensorimotor strength from the Lancaster Sensorimotor Norms (Lynott et al., 2019), where each dimension contained a rating of the extent to which the word's referent was experienced with the specified perceptual modality or by performing an action with the specified action effector. We also included Lynott et al.'s composite measure

[6] 5307 words had *d'* and *c* measures used in our analyses

of all 11 dimensions of sensorimotor strength, which was weighted towards the dominant dimension(s).

Using JASP, we conducted the PCA via parallel analysis (95th percentile), with pairwise exclusions and varimax rotation, using the correlation matrix, which reduced the original 24 dimensions to an optimal 6 principal components that captured 75.1% of the original variance. We then used the Principal function[7] in R to calculate rotated component scores for each word. This produced six components that consolidated the most important information from the original variables: 2 lexical components (Frequency and Length) and 4 sensorimotor components (Body, Object, Food, Communication). Note that the components cleanly distinguished between lexical and semantic information, with the exception of the Object component, which included the noun (part of speech) variable in addition to sensorimotor variables; this contribution was not unexpected since object concepts are typically labelled with nouns (e.g., *apple*, *dog*) and tend to be strongly experienced with visual, haptic, and hand/arm action. Table 1 summarises how each component relates to the original variables in the PCA.

The overlap of the Cortese et al. (2010; 2015) datasets that contributed the dependent measures (5578 words) and the PCA components that acted as predictors (9796 words) resulted in a sample of 5311 words which were used in the current analysis.

[7] Which is also used to perform PCA in JASP

Table 1: Variables used in Principal Components Analysis in Study 1a and the rotated components to which they most strongly contributed with positive or negative weighting (*r* > .3 or < -.3).

Original variable	Source	Definition	PCA component
LgSUBTLWF	ELP	Log word frequency	+Frequency
LgSUBTLCD	ELP	Log contextual diversity (how many contexts a word appears in)	+Frequency
Zipf Frequency	Van Heuven et al. (2014)	Word frequency on Zipf scale	+Frequency
Prevalence	Brysbaert et al. (2018)	How many people know the word	+Frequency
Familiarity	Stadthagen-Gonzales & Davis (2006); Scott et al. (2018); Wilson (1988)	How subjectively familiar a word seems (ratings)	+Frequency
Age of Acquisition	Kuperman et al. (2012)[a]	Approximate age that the word was learned	–Frequency
Linguistic distributional distance (LDD20)	Dymarska et al., 2021	Distributional neighbourhood (mean cosine distance to closest 20 neighbours, based on vectors of log co-occurrence frequency)	–Frequency
Word length	ELP	Word length in letters	+Length
Number of syllables	ELP	Word length in syllables	+Length
Orthographic Levenshtein Distance (OLD20)	ELP	Orthographic neighbourhood (mean letter Levenshtein distance to closest 20 neighbours)	+Length
Phonological Levenshtein Distance (PLD20)	ELP	Phonological neighbourhood (mean phoneme Levenshtein distance to closest 20 neighbours)	+Length
Torso action strength	LSN	Motor strength in torso effector	+Body
Foot/leg action strength	LSN	Motor strength in foot/leg effector	+Body
Hand/arm action strength	LSN	Motor strength in hand/arm effector	+Body, +Object
Composite sensorimotor strength	LSN	Aggregated sensorimotor strength in all dimensions (Minkowski-3 distance of 11-dimension vector from the origin)	+Body, +Object, +Communication, +Food
Head action strength	LSN	Motor strength in head effector	+Communication
Auditory strength	LSN	Perceptual strength in hearing modality	+Communication
Mouth action strength	LSN	Motor strength in mouth effector	+Communication, +Food
Gustatory strength	LSN	Perceptual strength in taste modality	+Food
Olfactory strength	LSN	Perceptual strength in smell modality	+Food
Visual strength	LSN	Perceptual strength in sight modality	+Object
Noun (part of speech)	ELP	Whether or not word is a noun (binary coded: noun=1, non-noun=0)	+Object
Haptic strength	LSN	Perceptual strength in touch modality	+Object, +Body, –Communication
Interoceptive strength	LSN	Perceptual strength in interoceptive (sensations inside the body) modality	–Object, +Body, +Communication

[a] With extended norms from http://crr.ugent.be/archives/806

Note: ELP = English Lexicon project (Balota et al., 2007); LSN = Lancaster Sensorimotor Norms (Lynott et al., 2019).

Design and analysis

To investigate the extent to which sensorimotor information contributed to word recognition memory, we conducted item-level hierarchical linear regression analyses of the five dependent measures of memory performance: Hit Rate (HR), False Alarms (FA), HR-FA, *d'* and *c*. Step 1 entered the two lexical components (Frequency, Length) as baseline model predictors, then Step 2 entered the four sensorimotor components (Body, Communication, Food, Object) and their interactions with the Frequency component. We included these interactions because semantic effects in word recognition are typically larger for low-frequency words than high-frequency words (e.g., Connell & Lynott, 2016b; James, 1975). This has not been studied extensively in memory research, but there is some evidence that the effect of word concreteness on word recall is higher for low frequency words (Walker & Hulme, 1999), and word memory recognition is faster for low frequency words with high concreteness ratings (Taylor, 2017). We therefore expected similar patterns to appear in word recognition memory.

We ran Bayesian linear regressions in JASP (0.14.1: JASP Team, 2020) with default JZS priors (r = .354) and a Bernoulli distribution (p = 0.5), from which we report Bayes Factors for model comparisons between hierarchical steps and inclusion Bayes Factors of coefficients (i.e., relative likelihood of models including a particular predictor compared to models excluding it). In addition, to calculate part correlation coefficients for each predictor (i.e., the unique contribution each predictor makes to the dependent measure in question), we ran NHST linear regression analyses using the same structure as the Bayesian linear regression.

Results and Discussion

Overall, performance in the memory task was good, with high hit rates, low false alarms and low bias (see Table 2 for descriptive statistics). Lexical effects were largely consistent with Cortese et al. (2010; 2015): lower frequency words produced higher hit rates and HR-FA, but also higher false alarms, better *d'* sensitivity, and a more liberal response bias. Word length produced small effects in word memory performance, but the pattern of results indicated that shorter words elicited lower HR and FA, higher HR-FA, and a higher bias, with no effect on *d'*. Full statistics are available in supplemental materials. Contrary to predictions, the four kinds of sensorimotor experience affected memory in different ways, and therefore we will outline their effects separately.

Table 2: Average performance on each memory measure with its standard deviation.

DV	Mean	SD
Hit Rate	0.728	0.097
False Alarms Rate	0.202	0.096
Hit Rate-False Alarms	0.526	0.132
d'	1.524	0.472
c	0.126	0.243

Table 3: Percentage of variance in memory performance explained by each step of the regression model (change in R^2, with levels of Bayesian evidence) and uniquely explained by each sensorimotor component in the Step 2 model (squared part correlations).

Model / parameter	HR	FA	HR-FA	*d'*	*c*
Step 1: Lexical baseline	26.20***	0.30*	11.80***	8.20***	11.80***
Step 2: Sensorimotor	5.00***	3.00***	6.90***	6.00***	1.40***
Body	0.18	0.88	0.14	0.21	0.98
Body*frequency	0.20	0.00	0.05	0.05	0.14
Communication	0.08	0.01	0.08	0.06	0.01
Communication*frequency	0.04	0.01	0.05	0.05	0.00
Food	1.10	1.02	2.28	2.28	0.01
Food*Frequency	0.00	0.00	0.01	0.01	0.00
Objects	2.66	0.74	3.31	2.72	0.15
Objects*frequency	0.02	0.04	0.03	0.03	0.02
Total Step 1 + Step 2	31.20	3.30	18.70	14.20	13.20

* $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence

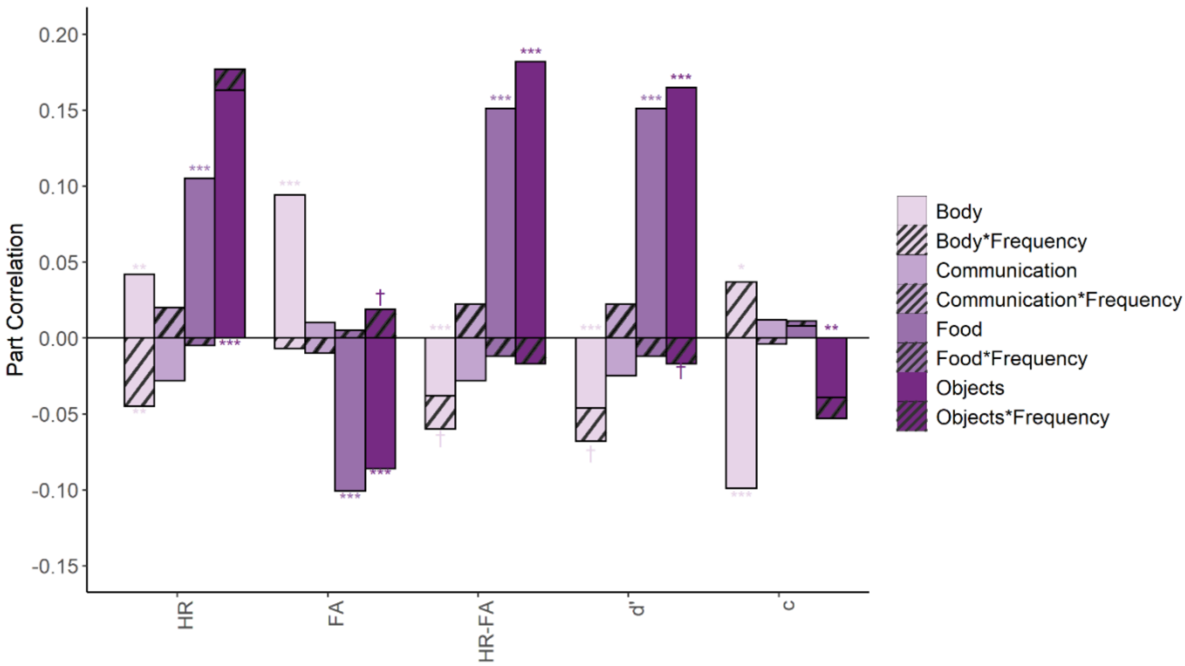


Figure 1: Part correlations of each sensorimotor component predictor in Study 1a (Step 2 model). Stacked bars represent the combined effect of each component (lighter shade) and the respective component*Frequency interaction (darker shade). The symbols per bar refer to the inclusion Bayes Factor (BF) of each predictor: *** $BF_{10} \geq 150$, constitutes very strong evidence; ** $BF_{10} \geq 20$, strong evidence; * $BF_{10} \geq 3$, positive evidence; † $BF_{10} \geq 0.33$ and <3.00, equivocal evidence; no symbol $BF_{10} < 0.33$, evidence against inclusion.

Body

Sensorimotor experience relating to the Body (i.e., involving motor action of the torso, feet/legs, and hand/arms, plus touch and interoceptive experience) had complex effects on word memory. Words scoring higher in the Body component had higher hit rates but also higher false alarms, meaning that they led to a large liberal response bias (negative *c*). In other words, participants were more likely to think that a word was “old” (i.e., previously seen) if its referent was strongly grounded in Body experience, even when they had not actually seen the word in the study list. This biasing effect was strong for low- and average-Frequency words, but (as shown by the Body*Frequency interactions for HR and c) it was attenuated for high-Frequency words. High Body scores overall had a negative effect on composite performance measures of word memory (negative HR-FA and *d′* sensitivity), and there was equivocal evidence that this effect was stronger for high Frequency words. That is, high Body strength hindered, rather than helped, performance on word recognition memory, and this was somewhat more likely if the word was also high in Frequency.

Communication

The Communication component did not elicit any effects on word memory performance, regardless of Frequency. That is, words relating to sound and overall sensorimotor experience, as well as mouth and head action, were not easier to remember or more likely to be recognised as previously seen.

Food

In contrast to Body effects, sensorimotor experience relating to Food (i.e., involving taste, smell, and mouth action) had more consistent effects on word memory, and generally followed our predictions. Words scoring higher in the Food component had higher hit rates and lower false alarms, but no effect on bias (*c*). Overall, high Food scores led to better HR-FA and *d′* sensitivity. However, this did not interact with word Frequency.

Object

The Object component (that is, words referring to manipulable objects which can be touched and seen, and denoting nouns) had the strongest effects on word memory. Words rated higher on the Object component had higher hit rates and lower false alarms, which led to a moderately liberal response bias – participants still had a tendency to judge words as previously seen when they were grounded in the experience of Objects, regardless of whether they had been seen in the study list. However, overall the Object component had a strong positive effect on composite performance measures of word memory (positive HR-FA and *d′* sensitivity), suggesting that high Object strength facilitated performance in a word recognition memory task. This was in line with our predictions, except for the finding that the effects of the Object component on FA and *d’* were somewhat attenuated for high Frequency words, as suggested by the equivocal evidence for the interaction.

Conclusion

In summary, sensorimotor variables contributed to performance in a word memory task. The sensorimotor components used in the analysis varied in the kind of information they captured, and we found that they predicted different patterns in memory performance. While some aspects of sensorimotor strength increased the likelihood of the word being correctly remembered, others inflated the likelihood of making an incorrect judgement. In particular, it appears that high Body strength led participants to perceive a word which was not studied as “old”, more than when the word was presented in the study phase, which was not the case for other components. In other words, experiencing a concept with the body gives a newly encountered concept an illusion of being a studied word with a strong memory trace, but this does not extend to recently studied concepts. We address this further in the general discussion.

The finding that different aspects of perceptual and motor experience affect memory in different ways is an important one, because it does not support the previous theoretical claims and empirical research on semantic richness effects, whereby stronger semantic information invariably facilitates word recognition and word memory (e.g., Madan, 2020; Nairne et al., 2013; Yap et al., 2011). In contrast, our findings suggest that sensorimotor strength could either facilitate or inhibit processing in recognition memory, depending on the type of experience it represents.

Study 1b: Composite Sensorimotor Experience

In Study 1a, we examined how word memory performance is affected by four semantic variables which capture different aspects of sensorimotor information, and we found that their effects on word memory diverge according to the form of information they represent. Nonetheless, most studies on semantic richness try to capture sensorimotor information in a single variable, such as imageability, number of features, body-object interaction (BOI; Sidhu et al. 2018), etc. These types of measures, obtained through asking participants to rate their overall experience with a concept, do not reliably capture the content of a representation (Connell & Lynott, 2012; Dymarska et al. 2021). Even ratings of the overall strength of sensorimotor experience in a single variable do not reflect the full sensorimotor grounding of a concept (Connell & Lynott, 2016a), given that, as demonstrated in Study 1a, different aspects of sensorimotor experience may act differently when processing different kinds of words. However, when perceptual strength, rated on multiple individual dimensions, is aggregated into a composite score calculated from the individual ratings, that score produces a more accurate measure of sensorimotor grounding of a concept (Connell & Lynott, 2016a).

The 11 sensorimotor dimensions of sensorimotor experience which were included in the sensorimotor components in Study 1a, were also used to create an aggregate measure of sensorimotor strength which best predicts lexical processing (Lynott et al., 2019). The most accurate variable was Minkowski 3 sensorimotor strength, which represents the sensorimotor strength in all dimensions, but with attenuated influence of weaker dimensions on the composite score. Rather than use 4 sensorimotor components as in Study 1a, we wanted to use a single composite measure of sensorimotor experience to test whether it can capture a large amount of variance in memory performance compared to the total variance explained by the 4 sensorimotor components in Study 1a. While the Minkowski 3 variable was included in the PCA that produced the Frequency and Length components, its contribution to those components was minimal as its variance was instead spread across the 4 sensorimotor components. We wanted to test whether on its own the composite sensorimotor strength variable could be used as a measure of semantic richness and a predictor of word memory.

Method

Materials

We used the same item set and lexical components (Length and Frequency) as in Study 1a. We also used the composite sensorimotor strength measure (Minkowski 3) from Lynott et al. (2019).

Design and Analysis

We ran hierarchical item-level linear regressions with the lexical components (Frequency and Length) as Step 1 predictors, and the composite sensorimotor strength (Minkowski 3) and its interaction with Frequency as Step 2 predictor, with the same five memory measures as in Study 1a (from Cortese et al., 2010; 2015) as dependent variables. We ran Bayesian linear regressions in JASP as in Study 1a, from which we again report Bayes Factors for model comparisons between hierarchical steps and posterior coefficient inclusion Bayes Factors (i.e., relative likelihood of models including a particular predictor compared to models excluding it). In addition, to calculate part correlation coefficients for each predictor (i.e., the unique contribution each predictor makes to the dependent measure in question), we ran NHST linear regression analyses using the same structure as the Bayesian linear regression.

Results and Discussion

The contribution of the lexical components was the same as in Study 1a, as this step in the model was identical.

Table 4: Percentage variance explained by each step of the regression model (change in R^2, with levels of Bayesian evidence) and uniquely explained by sensorimotor strength in the Step 2 model (squared part correlations).

Model / parameter	HR	FA	HR-FA	*d'*	*c*
Step 1: Lexical baseline	26.20***	0.30*	11.80***	8.20***	11.80***
Step 2: Sensorimotor	2.00***	0.20	1.60***	1.10***	0.60***
Minkowski3	1.85	0.03	1.25	0.92	0.55
Minkowski3*Frequency	0.07	0.06	0.00	0.00	0.18
Total Sensorimotor	1.92	0.09	1.25	0.92	0.73

* $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence

Figure 2: Part correlation of sensorimotor strength's contribution to memory performance from Study 1b (Step 2 models). Stacked bars represent the combined effect of sensorimotor strength (lighter shade) and the sensorimotor*Frequency component interaction (darker shade) in the Step 2 (final) regression model. The symbols per bar refer to the inclusion Bayes Factor (BF) of each predictor: *** $BF_{10} \geq 150$ constitutes very strong evidence; ** $BF_{10} \geq 20$, strong evidence; * $BF_{10} \geq 3$, positive evidence; † $BF_{10} \geq 0.33$ and <3.00, equivocal evidence; no symbol $BF_{10} < 0.33$, evidence against inclusion.

Overall, higher sensorimotor strength facilitated word recognition memory – it increased hit rates, with equivocal evidence that the effect was stronger for high Frequency words. However, composite sensorimotor strength had little effect on FA, and so it improved overall performance in terms of HR-FA and *d'* sensitivity, regardless of word frequency. It also induced a liberal response bias (negative *c*), which meant that higher sensorimotor strength in a presented word biased people towards thinking it was previously seen (i.e., "old" response), but this effect was attenuated for higher-frequency words.

While sensorimotor strength had no effect on FA in the regression, the inclusion BF was equivocal, suggesting a possibility that higher sensorimotor strength reduces false alarms, particularly for high Frequency words, but the effect was weak and should be treated with caution.

Compared to the effects of sensorimotor components in Study 1a, the combined measure explained less variance on all measures than the sum of the 4 components. More specifically, the model with sensorimotor components in Study 1a was $BF_{21} = 1.01^{41}$ times better at predicting HR than the model with Minkowski 3 in Study 1b, $BF_{21} = 7.48^{26}$ times better at predicting FA, and $BF_{21} = 4.85^{56}$ times better at predicting *d'*. Unique variance of sensorimotor strength in this analysis was higher than the unique variance of the Communication component, but lower than the Object and Food components, and similar to the unique variance of the Body component.

The sensorimotor strength measure followed the same direction as the Object component in Study 1a, which was the component with the strongest effect. This suggests that the composite sensorimotor measure was driven largely by the strength of the Object component, but the effects were attenuated by some of the other components pulling the effects in the opposite direction, leading in particular to an equivocal effect of sensorimotor strength on FA. This pattern of results indicates that the composite sensorimotor strength measure does not capture sensorimotor experience equally as well as the individual components.

Conclusion

We found that a composite measure of sensorimotor strength (Minkowski 3) facilitated word memory performance. Words rated higher in sensorimotor strength were more likely to be correctly judged as “old”. This was in line with the semantic richness effect. However, this variable predicted less variance in each of the measures of memory performance than sensorimotor strength derived from different aspects of sensorimotor experience and divided into 4 variables used in Study 1a. More specifically, in Study 1a sensorimotor strength predicted between 3.0% of variance in FA to 6.9% of variance in HR-FA, while the composite variable in Study 1b predicted only 0.2% of variance in FA and 1.6% of variance in HR-FA. We therefore concluded that the 4 sensorimotor components were better at capturing variance associated with different aspects of sensorimotor information, and that a single variable which combines their contribution to word representation cannot replace them as an equally useful predictor.

Study 2: Conscious Imagery

The sensorimotor effects of Study 1 partially resemble those of imageability, where a long history of research has shown that greater imageability enhances word memory (Bourassa & Besner, 1994; Caplan & Madan, 2016; Cortese et al., 2010; 2015; Fliessbach et al., 2006; Majerus & Van der Linden, 2003; Paivio, 1969; Tse & Altarriba, 2007; Xiao et al., 2012), in line with the semantic richness effect (Yap et al., 2012). However, sensorimotor strength and imageability are very different theoretical constructs (Connell & Lynott, 2012). While sensorimotor strength is concerned with the grounding of word meaning in perception and action experience which is automatically represented during word processing (Lynott et al., 2019), imageability is specifically concerned with the ability to consciously generate sensory imagery of word meaning (Paivio et al., 1968)[8]. Hence, imageability ratings conflate sensorimotor (largely sensory) grounding with the ease of generating imagery in that modality, which is problematic because it makes it difficult to know whether imageability effects on word memory are due to ease of imagery generation per se or the underlying sensorimotor grounding.

---

[8] Commonly used instructions from Paivio et al. (1968, p. 4) ask participants to "rate a list of words as to the ease or difficulty with which they arouse mental images", where mental images are defined as "a mental picture, or sound, or other sensory experience".

While "pure" imageability is not a good predictor of visual word recognition when the effects of imageability ratings from different sources (i.e., different sets of norms) are examined on top of sensorimotor grounding (Dymarska et al., 2021), it is possible that it may be a better predictor of memory. Imagery seems like a sensible strategy in any experimental paradigm that asks participants to memorise a list of words, as retrieving a memory may rely on consciously generating an image of the word's referent, which is what imageability represents. In Study 2, we therefore tested how imageability would perform as a predictor of word memory when all other sensorimotor variance is accounted for. We analysed imageability norms from multiple sources to ensure generalisable conclusions.

Method

Materials

Items, dependent measures (from Cortese et al., 2010; 2015), and lexical-sensorimotor components were the same as in Study 1a. As our predictor of interest in the present study, we collated imageability ratings from 6 different sets of imageability norms, each of which used the same instructions and scale to collect ratings from participants: the Bird norms (Bird et al., 2001); Chiarello norms (Chiarello et al., 1999); Cortese norms (Cortese & Fugett, 2004; Schock et al., 2012)[9]; Bristol norms (Davis & Stadthagen-Gonzalez, 2006); Glasgow norms (Scott et al., 2018); and the widely-used MRC norms (Coltheart, 1981; Wilson, 1988; featuring imageability ratings from Gilhooly and Logie, 1980; Paivio et al., 1968; Toglia & Battig, 1978). Because each set of norms covered a different sample of words with varying overlap with the dependent measures of the word memory dataset, and because previous work found large differences in predictive ability of different imageability norms (Dymarska

[9] Cortese & Fugett (2004) and Schock et al. (2013) norms were combined into a single variable since they came from the same laboratory and were used in Cortese et al. (2010) and Cortese et al. (2015) respectively, which we analyse as a single dataset.

et al., 2021), we analysed each separately. Descriptive statistics for each word subsample analysed in the present study are in Table 5.

Table 5: Descriptive statistics of the imageability norms used in Study 2

Norms	Number of words	Mean imageability rating	Imageability rating SD
Bird	810	4.38	1.02
Bristol	1223	4.01	1.38
Chiarello	989	4.92	1.33
Cortese	5311	4.50	1.40
Glasgow	2593	4.86	1.35
MRC	2563	4.80	1.00

Design and Analysis

In order to examine how the different sets of imageability norms may differ systematically in their lexical and sensorimotor characteristics, we first calculated Bayesian zero-order correlations between the imageability ratings from each set of norms with the lexical and sensorimotor components.

We then conducted hierarchical item-level linear regressions, similar to Study 1a, separately for each imageability dataset. In Step 1, we entered the lexical components of Frequency and Length. In Step 2, we entered the sensorimotor components (Body, Food, Object, Communication) and their interactions with the Frequency component. Finally, in Step 3, we entered imageability (centred) and its interaction with the Frequency component. As per Studies 1a-b, the DVs from the word memory task were HR, FA, HR-FA, *d'* sensitivity, and *c* response bias. We ran both Bayesian and NHST regressions with the same parameters. We report Bayes Factors of posterior coefficients and R^2 change from the Bayesian regressions, and part correlation coefficients from the NHST regressions.

Results and Discussion

We first report how different sets of imageability norms differ in the lexical and sensorimotor characteristics of their item sets, and then examine the ability of each set of norms to predict word recognition memory above and beyond sensorimotor information.

Correlation of Imageability with Sensorimotor Components

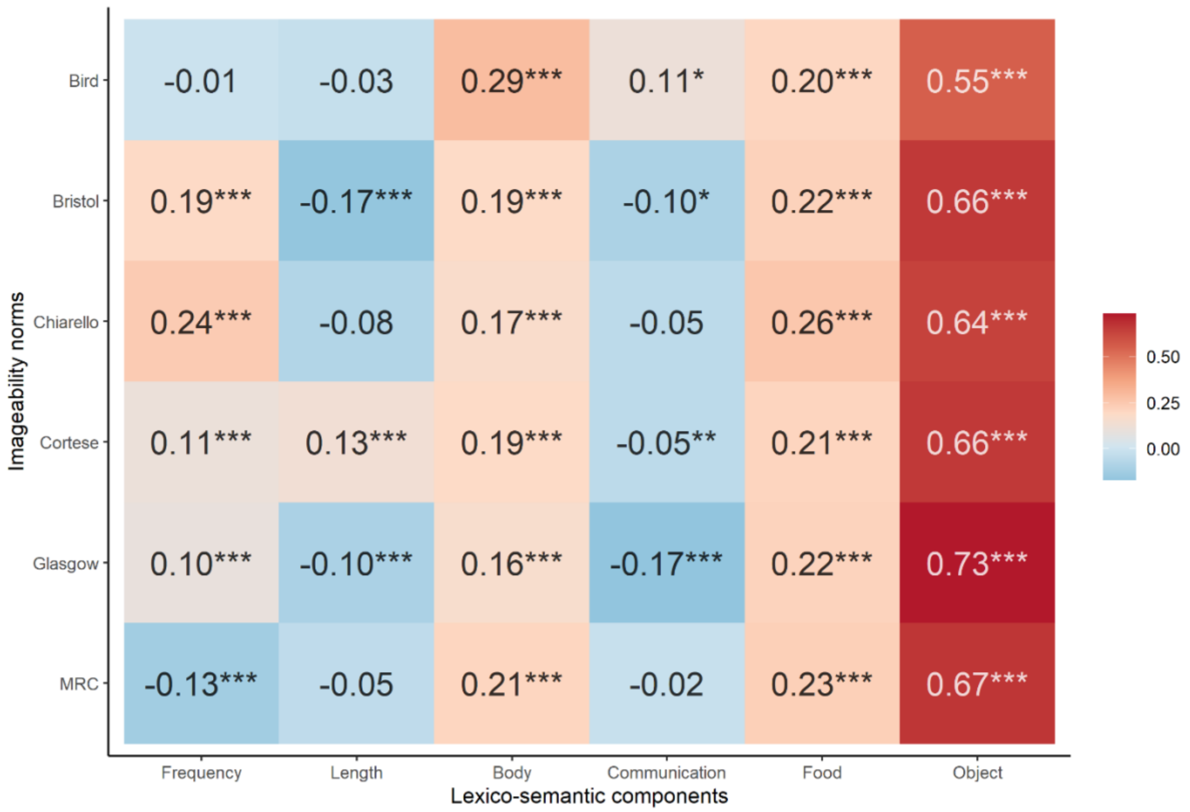


Figure 3: Heatmap of zero-order correlations between lexical and sensorimotor components and imageability ratings from different sources. The asterisks per cell refer to the Bayes Factor (BF) of the correlation coefficient, * $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq$, 150 very strong evidence.

Figure 3 shows that all sets of imageability norms were highly and positively correlated with the Object component and moderately positively with the Body and Food components. That is, words that were rated higher in imageability tended to be stronger in some aspects of sensorimotor experience. This pattern of correlations suggests that conceptual information captured by imageability ratings may instead be at least in part due to sensorimotor grounding, particularly experience associated with manipulable objects (which involves vision, touch, and hand/arm action), and also to some extent with whole-body experiences (which involves action of the torso, feet/legs, and hand/arms, plus touch and interoceptive sensations) and food interactions (smell, taste, and mouth action). While the magnitude of these effects varied somewhat across norms (e.g., the Object component has 53% shared variance with the Glasgow imageability ratings but only 31% with the Bird ratings), the overall consistency of these effects across sources indicates the relationship between imageability and Object, Body and Food experience is robust and consistent with similar analysis of a larger sample of words by Dymarska et al. (2021).

However, there were major inconsistencies in how each set of norms correlated with other components. For instance, while three sets of imageability norms were negatively correlated with the Communication component (Bristol, Cortese, Glasgow), the Chiarello and MRC norms showed no correlation at all, while the Bird norms showed a positive correlation. Even where negative correlations with Communication appeared, they varied in magnitude (e.g., 3% shared variance with the Glasgow imageability ratings but only 0.2% with the Cortese ratings). That is, words that were strongly grounded in Communication (which involves action of the mouth and head, plus hearing and interoception) were sometimes rated as easy to image and sometimes difficult. There were similar inconsistencies for the lexical components, which diverged from the previous reports in the literature. The correlation of the norms with the Frequency component varied from positive (Bristol, Chiarello, Cortese, Glasgow) to non-existent (Bird) to negative (MRC); highly common words (i.e., more frequent, familiar, prevalent, contextually diverse, with many neighbours, and learned earlier) were sometimes rated higher in imageability and sometimes lower. Finally, the Length component was not reliably related to the imageability ratings: words rated higher in imageability were sometimes shorter (fewer letters, syllables, and orthographic/phonological neighbours – in Bristol and Glasgow norms), sometimes longer (the Cortese norms), and sometimes there was no relationship between word length and imageability (Bird, Chiarello, MRC). It is unclear whether these inconsistencies result from sampling differences of words per set of norms, or different rating behaviour of the participants involved in each norming study (e.g., individual differences in their ability to generate imagery for auditory, interoceptive and/or mouth action sensations). Nonetheless, as found in previous analysis of a larger item set (Dymarska et al., 2021), the norms may vary in their ability to measure how much a word is associated with a mental image and may not reflect the same sensorimotor information about referent concepts. For example, the word “head” was rated 3.4 out of 7 in the Bird norms, but 6.8 in the Cortese norms, while the word “spire” was rated 5.5 in the Glasgow norms but only 2.7 in the Cortese norms. This could affect their reliability as a predictor of memory performance above and beyond sensorimotor grounding.

Effects of Imageability on word memory

Table 6: Percentage of variance explained by each step of the regression model (change in R^2, with levels of Bayesian evidence) and uniquely explained by imageability in the Step 3 model (squared part correlations).

Norms	Model / parameter	HR	FA	HR-FA	*d'*	*c*
Bird	Step 1: Lexical baseline	29.6***	4.2***	12.8***	8.6***	18.0***
	Step 2: Sensorimotor (ΔR^2)	6.1***	3.5***	8.8***	8.2***	1.6***
	Step 3: Imageability (ΔR^2)	3.4***	1.3*	5.0***	4.7***	0.2
	Imageability parameter (sr^2)	1.6	0.2	1.8	1.4	0.2
	Imageability*Frequency (sr^2)	0.1	0.2	0.3	0.4	0.1
	Total Step 1-3 R^2	39.1	9.0	26.6	21.5	19.8
Bristol	Step 1: Lexical baseline	15.0***	0.4	5.4***	4.3***	7.0***
	Step 2: Sensorimotor (ΔR^2)	5.3***	5.2***	9.3***	8.9***	1.8***
	Step 3: Imageability (ΔR^2)	6.0***	4.6***	11.0***	9.8***	0.2
	Imageability parameter (sr^2)	3.7	2.0	5.9	5.2	0.0
	Imageability*Frequency (sr^2)	0.1	0.5	0.5	0.5	0.2
	Total Step 1-3 R^2	26.3	10.2	25.7	23.0	9.0
Chiarello	Step 1: Lexical baseline	15.2***	4.2***	4.6***	3.3***	12.0***
	Step 2: Sensorimotor (ΔR^2)	8.0***	5.6***	13.0***	12.3***	1.8***
	Step 3: Imageability (ΔR^2)	9.5***	1.6***	9.9***	8.4***	1.4***
	Imageability parameter (sr^2)	7.6	0.4	6.4	5.1	1.4
	Imageability*Frequency (sr^2)	0.2	0.7	0.9	1.0	0.2
	Total Step 1-3 R^2	32.7	11.4	27.5	24.0	15.2
Cortese	Step 1: Lexical baseline	26.2***	0.3*	11.8***	8.2***	11.8***
	Step 2: Sensorimotor (ΔR^2)	5.0***	3.0***	6.9***	6.0***	1.4***
	Step 3: Imageability (ΔR^2)	8.0***	2.3***	9.7***	8.0***	0.8***
	Imageability parameter (sr^2)	7.6	1.5	8.5	6.8	0.7
	Imageability*Frequency (sr^2)	0.0	0.3	0.1	0.2	0.3
	Total Step 1-3 R^2	39.2	5.6	28.4	22.2	14.0
Glasgow	Step 1: Lexical baseline	24.7***	1.6***	7.5***	5.0***	15.0***
	Step 2: Sensorimotor (ΔR^2)	5.1***	6.4***	10.6***	9.9***	1.8***
	Step 3: Imageability (ΔR^2)	7.2***	5.1***	12.8***	11.7***	0.1
	Imageability parameter (sr^2)	4.5	3.0	7.8	7.1	0.0
	Imageability*Frequency (sr^2)	0.0	0.0	0.0	0.0	0.1
	Total Step 1-3 R^2	37.0	13.1	30.9	26.6	16.9
MRC	Step 1: Lexical baseline	33.6***	3.2***	12.7***	8.1***	21.2***
	Step 2: Sensorimotor (ΔR^2)	4.6***	3.4***	7.8***	6.7***	0.9***
	Step 3: Imageability (ΔR^2)	5.9***	1.4***	7.7***	6.2***	0.4*
	Imageability parameter (sr^2)	3.6	0.8	4.5	3.9	0.4
	Imageability*Frequency (sr^2)	3.6	0.8	4.5	3.9	0.4
	Total Step 1-3 R^2	44.1	8.0	28.2	21.0	22.5

Note: The imageability parameter and interaction do not add up to imageability ΔR^2, because the model also included non-unique variance and variance shared between imageability and interaction parameter, which had been partialled out when calculating part correlations. * BF_{10} >= 3 positive evidence; ** BF_{10} >= 20 strong evidence; *** BF_{10} >= 150 very strong evidence

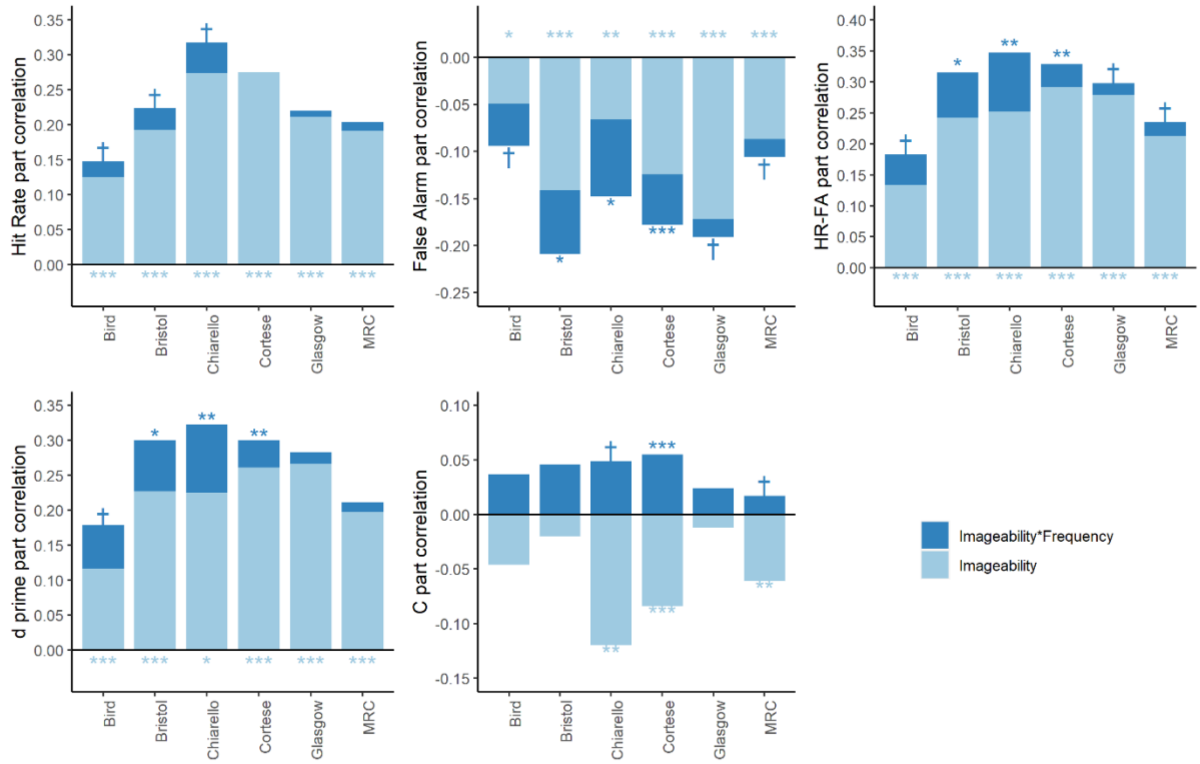


Figure 4: Part correlations of imageability effect on memory performance. Stacked bars represent the combined effect of imageability (green) and the imageability*frequency interaction (blue) in the Step 3 (final) regression model. The symbols per bar refer to the inclusion Bayes Factor (BF) of each predictor: *** $BF_{10} \geq 150$, constitutes very strong evidence; ** $BF_{10} \geq 20$, strong evidence; * $BF_{10} \geq 3$, positive evidence; † $BF_{10} \geq 0.33$, and <3.00, equivocal evidence; no symbol $BF_{10} < 0.33$, evidence against inclusion.

In general, memory performance was good across subsets of words, with high hit rates and overall memory performance measures (HR-FA and *d'*), and low false alarms. Higher frequency words produced higher hit rates and HR-FA, but also higher false alarms, lower *d'* sensitivity, and a more liberal response bias. Effects of word length were small and varied between the word subsets. Full statistics are available in supplemental materials.

Overall, part correlations showed that higher imageability predicted higher hit rate, regardless of Frequency, even when sensorimotor grounding had been taken into account. For half of the norms (Bird, Bristol, Chiarello) there was equivocal evidence that this effect was stronger for higher Frequency words. However, the unique effect of imageability (i.e., purely the ease of generating conscious imagery, and not the underlying sensorimotor grounding) varied in magnitude by nearly a factor of 5, from 1.6% in the Bird norms, to 7.6% in the Cortese norms (see sr^2 figures in Table 6).

Higher imageability also led to lower FA, that is, words rated high in imageability were less likely to be incorrectly judged as “old”. This effect was stronger for higher Frequency words, especially for the Bristol, Chiarello and Cortese norms, with equivocal support for this interaction for the other three sets. The magnitude of effects varied across the norms, from 0.4% of the unique variance being explained by the Bird norms, to 3.0% of unique variance explained by the Glasgow norms.

The overall performance measures, HR-FA and *d’*, were also influenced by higher imageability, which led to better word memory performance. Similar to the the FA results, the effects were stronger for higher Frequency words for the Bristol, Chiarello and Cortese norms on both measures. There was also equivocal evidence that the Bird norms elicited a stronger effect on higher Frequency words on both measures, and that the Glasgow and MRC norms elicited a stronger effect for higher Frequency words on HR-FA, but not *d’*. The magnitude of the results again varied across norms by a factor of four: the Bird norms explained the least unique variance in memory performance, 2.0% in HR-FA and 1.8% in *d’*, while the Cortese norms explained 8.5% of unique variance in HR-FA, and the Glasgow norms explained 7.1% of unique variance in *d’*.

Finally, the measure of bias was least affected by the imageability ratings. Higher imageability predicted lower bias for the Chiarello, Cortese and MRC norms, such that words were high imageability items were rated as old. The effect of the Cortese norms increased for low Frequency words, but there was only equivocal evidence for this for the Chiarello and MRC norms. The magnitude of the effects was lower than for the other memory measures, but still varied from 0.07% for the Glasgow norms to 1.7% for the Chiarello norms.

While all norms elicited some effects on memory performance, this varied in magnitude. In particular, the Bird norms had a small effect across all analyses, and the Glasgow norms had the lowest effect when predicting bias. The direction of results for average frequency words was consistent with the findings of Cortese et al. (2010; 2015) where higher imageability facilitated Hit Rate and led to lower False Alarms. However, the magnitude of the effects differed. In Cortese et al. imageability predicted 14-24% of HR when lexical variables were accounted for. Here, imageability explained 8% of total memory variance in the Cortese dataset HR, and less when other imageability norms were used (although on smaller sets of words).

Conclusion

We analysed imageability as a semantic predictor of recognition memory, over and above lexical and sensorimotor variables, using six different sources of imageability ratings. Overall, higher imageability words were more likely to be correctly recognised as old, or less likely to be incorrectly recognised as old (i.e., as a False Alarm). This suggests that imageability may help to remember words, at least when participants know that they are expected to remember them (as in the word recognition memory task analysed here). Higher imageability also appears to improve participants' sensitivity to whether words had been seen before. The results were consistent with the semantic richness effect, where a semantic variable facilitates conceptual processing, much like the results of Study 1b, but at odds with the findings of multi-component analysis of sensorimotor information in Study 1a. However, when other sources of semantic information were accounted for, the effects in the current study were smaller than previously found in the literature (Cortese et al., 2010; 2015), explaining up to 8.6% to word memory, rather than 14%-24% found by Cortese et al. Additionally, there were some differences in how well the effects were captured by different sets of imageability ratings. When using the imageability ratings from other sources, which covered different subsets of words than the Cortese norms, we found a much smaller effect size of imageability on memory than reported by Cortese et al. This was likely due to implementing a comprehensive lexical baseline model, which captured different types of word frequency, as well as their orthographic, phonological and contextual features, and the use of PCA components which had partialled out sensorimotor grounding. In other words, the findings of the current study suggest that a substantial part of the imageability effect size reported in word memory literature is due to sensorimotor grounding of word meaning enhancing the memory trace, rather than the ease of consciously generating sensorimotor imagery per se. Encountering a word and making a recognition judgment involves simulation of sensorimotor experience associated with the word, but that simulation is not always consciously generated.

Study 3: A surprise memory task

In the Cortese et al. (2010; 2015) megastudy, where the data in Study 1 and 2 was taken from, participants knew that they were supposed to memorise words for later recognition. It is possible that processing words with the intention of remembering might entail a more detailed semantic representation than, for example, a lexical decision task. We therefore investigated whether the same pattern of results would be found when participants do not know that they are going to be tested on their memory for the words they are presented with. To test this hypothesis, we performed an exploratory analysis of an existing dataset we collected from a surprise memory task, where participants were presented with a study list of words in the guise of a lexical decision task, and then were unexpectedly tested on their ability to remember the words in a recognition memory task. The dataset can be found in supplemental materials; we report the method below for clarity. In addition, to allow a fair comparison between expected and surprise memory tasks, we extracted data for the same set of items used in this study from the Study 1a dataset (Cortese et al.'s expected memory task) and analysed them with the same models.

Study 3a) Effects of Sensorimotor Strength

In Study 1 we found that sensorimotor grounding of word meaning facilitated word memory in its own right. Here we analyse whether the same effect can be found when participants do not expect to be performing a memory task. It is possible that the lack of an explicit strategy to remember will increase the role of sensorimotor grounding in memory trace, because participants will rely on the range of automatically available information rather than a subset of consciously available information. The study follows the analysis used in Study 1a.

Method

Participants. Participants were 154 native speakers of English (111 females; mean age=36.7 years, SD =13.2 years, 20 left-handed) recruited from Prolific.ac, for which they received a payment of £1.50. Forty-six participants were removed from the original sample of 200 due to their low scores on the memory task ($d' \leq 0$) suggesting that their discrimination between old and new items was below chance level.

Materials. 500 words[10] were selected from Cortese et al. (2010, 2015), sampled across the word frequency range by randomly choosing 125 words from each frequency quartile using the log subtitle word frequency measure (LgSUBTLWF; Balota et al., 2009). All words were one or two syllables long and comprised nouns, verbs, and adjectives. In addition, we generated 500 pseudowords using Wuggy (Keuleers & Brysbaert, 2010) that

[10] Two words used in the memory task: "yuppie" and "yummy" were excluded from the analysis as they did not have corresponding component scores. Additionally, the word "yodel" did not have a *d'*&*c* score in the Cortese et al. dataset, due to the FA rate being equal to 0. Thus, we excluded this word from the *d'*&*c* analysis of the surprise memory task. We refer to the sample as N=498 in the results section, but for *d'* and *c* the sample was N=497.

matched each original word in number of syllables and letters by changing some of the phonemes (e.g., “church” –> “chulks”). Where the generated item was not a clear pseudoword (i.e., could be considered a real English word; n=13), or where different items were assigned the same pseudoword (n=2), we replaced them with the second-best pseudoword option. Finally, we randomly divided the item set into 10 lists of 50 words each and corresponding 10 lists of 50 pseudowords each.

For the lexical decision task that constituted the study phase, each target word list was paired with a pseudoword list that was generated to match a different word list. That is, word list 1 was presented with pseudoword list 2, word list 2 was presented with pseudoword list 3, and so on.

For the word recognition memory task, the target word list (old words) was paired with a different word list, which had not been seen by the participant, to act as distractors (new words). For example, word list 1 was presented with word list 3 as distractors, word list 2 was presented with list 4 as distractors, and so on. In this way, each word list from the lexical decision task was subsequently tested in the memory task. Every participant saw 100 items in the lexical decision task (50 words and 50 pseudowords) and then 100 words in the memory task (50 old words from the lexical decision task, and 50 new words from a different, unseen list).

Procedure. The experiment was created and hosted through the online experiment builder Gorilla (http://www.gorilla.sc/; Anwyl-Irvine et al., 2019). We aimed to replicate the procedure of Cortese et al. (2010; 2015), with the exception of participants not knowing that they were performing a memory task during the study phase. After consenting to take part, participants were presented with instructions for the lexical decision task, where they were asked to decide whether a presented word was a real word in English (e.g., “young”) or not (e.g., “rilk”) by pressing the “Z” (not a real word) or “M” (real word) key on a keyboard.

Each trial began with a blank screen for 200 ms, followed by a fixation cross for 300 ms, and then the word (or non-word) presented individually in lowercase in the centre of the screen, 14 pixels in size, using black text on a white background, in Open Sans font (see Figure 5). The (non-)word stayed onscreen until the participant responded or until a timeout limit of 5000 ms was reached. A short practice task with 4 items was completed first, where participants received feedback for accuracy and speed; if they did not respond within 3000 ms, the message "Too slow" was displayed and the next trial commenced. There was no feedback in the main task.

When the lexical decision task was over, participants were presented with instructions for a distractor task that comprised 18 simple maths verification problems (e.g., "2+3=6?") as used by Cortese et al. (2010; 2015). Participants clicked on a button to begin the distractor maths task, and proceeded through questions at their own pace, without feedback. The entire distractor task took approximately 30 seconds.

Following the distractor task, the instructions for the surprise memory task appeared on the screen, where participants were asked to decide whether or not the displayed words had previously been seen in the lexical decision task, using the "Z" (new word) and "M" (old word) keys on the keyboard. Words were displayed as per the lexical decision task (see Figure 5) and stayed onscreen until the participant responded or until a timeout limit of 5000 ms was reached. There was no feedback during the recognition memory task. We measured accuracy of responses and response times from the onset of each word; accuracy per word was used to calculate the same dependent measures of memory performance that we analysed in Study 1a: HR, FA, HR-FA, *d'* sensitivity, and *c* response bias.

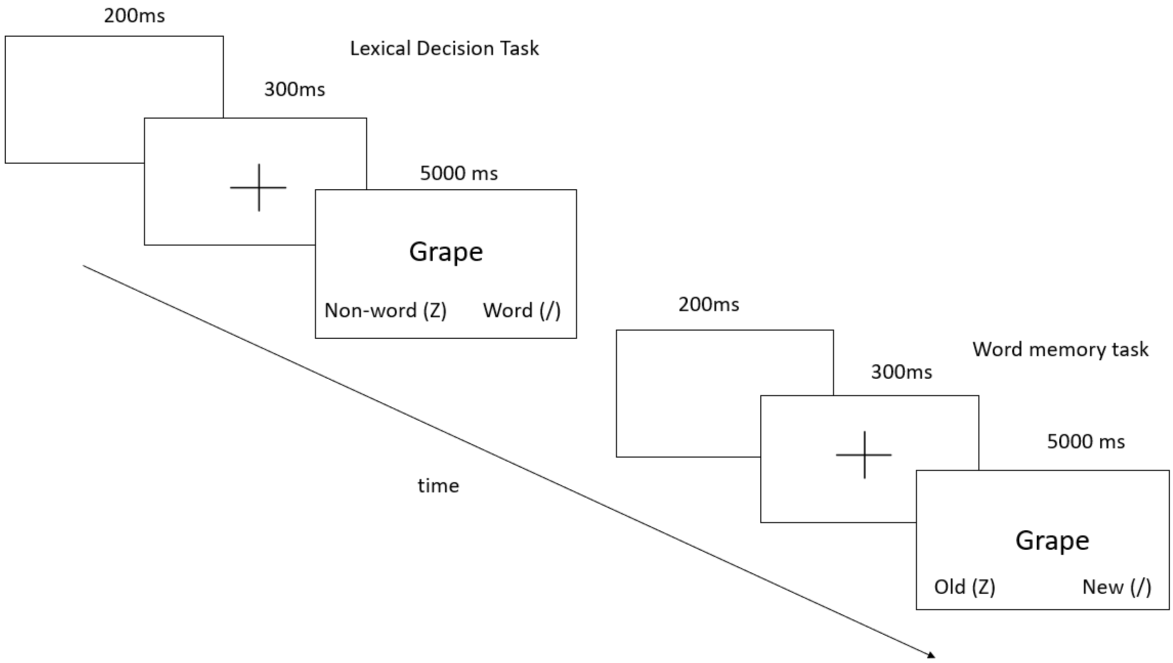


Figure 5: Trial diagram for surprise word memory task in Study 3a, comprising a lexical decision task for the study phase followed by a recognition memory task.

Design and Analysis. We ran hierarchical item-level linear regression analyses as in Study 1a. Step 1 entered the lexical components and Step 2 entered the sensorimotor components and their interactions with Frequency as predictors. The DVs were memory performance measures calculated from responses in the recognition memory task (HR, FA, HR-FA, *d'*, *c*) as well as response times calculated as item means for each word with a correct "old" response. The response time data was not available in the megastudies from Cortese et al. (2010; 2015) but it was measured in this study to gain a further insight into the time course of the word recognition memory process. RTs below 200ms were removed as motor errors. The *d'* and *c* measures were calculated using a log-linear approach (Stanislaw & Todorov, 1999) to compensate for ceiling performance (i.e., HR or FA at 0%). We ran Bayesian linear regressions in JASP from which we report Bayes Factors for model comparisons between hierarchical steps and posterior coefficients inclusion Bayes Factors (i.e., relative likelihood of models including a particular predictor compared to models excluding it). In addition, to calculate part correlation coefficients for each predictor (i.e., the unique contribution each predictor makes to the dependent measure in question), we ran NHST linear regression analyses using the same structure as the Bayesian linear regression.

Finally, in order to compare these results with an expected memory task on the same sample of words, we also conducted the above analyses on this subset of these 498 words from the dataset of Study 1a which featured the expected memory task data from Cortese et al. (2010; 2015) as dependent variables.

Results and Discussion

Overall, performance on the surprise memory task was good, with relatively high Hit Rates, low False Alarms, and good discrimination of old vs new items (see Table 7). When data on the same 498 words was analysed in an expected memory task, the Hit Rate and *d'* were slightly lower, but overall the tasks seemed to be of comparable difficulty.

As shown in Table 8, the lexical components explained a large amount of variance in most measures (more than in Study 1a), except for False Alarms and RT. Words higher in Frequency were less likely to be correctly recognised as old, and more likely to be incorrectly judged as old. Performance on higher Frequency words was also lower overall (negative effects on HR-FA and *d'*), and participants were more likely to judge them as “new” regardless of whether they had been studied (positive effect on bias). The Length component did not affect memory performance as strongly, but shorter words tended to be judged as old, regardless of whether or not they had been previously seen. Compared to the surprise memory task, the lexical components explained slightly more variance in the expected task in HR and *c*, but much less variance in *d'*. The direction of the effects of the lexical components was similar for both the surprise and the expected memory tasks.

Table 7: Average performance on each memory measure per task. Standard Deviation in brackets.

Task	HR	FA	HR-FA	*d'*	*c*	RT (ms)
Surprise Task	0.77 (0.14)	0.18 (0.14)	0.60 (0.21)	1.72 (0.74)	0.10 (0.33)	930 (82)
Expected Task	0.72 (0.10)	0.20 (0.10)	0.52 (0.14)	1.51 (0.49)	0.15 (0.25)	NA

Table 8: Percentage of variance in surprise and expected memory task performance explained by each step of the regression model (change in R^2, with levels of Bayesian evidence) and uniquely explained by each sensorimotor component in the Step 2 model (squared part correlations).

Model/Parameter	HR	FA	HR-FA	*d'*	*c*	zRT
Surprise memory task						
Step 1: Lexical baseline	32.70***	1.50	20.50***	17.00***	9.70***	2.80*
Step 2: Sensorimotor	0.90	3.30	2.30	2.60	1.80	1.40
Body	0.10	0.05	0.14	0.30	0.01	0.40
Body*Frequency	0.01	0.00	0.01	0.04	0.00	0.21
Communication	0.00	0.06	0.03	0.02	0.00	0.10
Communication*Freq	0.08	0.05	0.00	0.05	0.19	0.19
Food	0.13	0.92	0.15	0.19	1.06	0.14
Food*Frequency	0.01	0.00	0.00	0.02	0.10	0.22
Objects	0.36	1.90	1.74	1.99	0.34	0.02
Objects*Frequency	0.05	0.42	0.08	0.18	0.21	0.07
Total Step 1 + Step 2	33.6	4.80	22.8	19.60	11.50	4.20
Expected memory task						
Step 1: Lexical baseline	34.00***	0.80	19.90***	13.90***	12.50***	NA
Step 2: Sensorimotor	4.10	4.90	6.20**	5.90*	2.80	
Body	0.50	1.69	0.15	0.24	1.99	
Body*Frequency	0.52	0.72	0.00	0.02	0.98	
Communication	0.04	0.49	0.11	0.18	0.44	
Communication*Freq	0.28	0.03	0.08	0.07	0.09	
Food	0.71	0.31	1.10	1.23	0.03	
Food*Frequency	0.21	0.12	0.34	0.48	0.00	
Objects	1.10	2.22	3.35	2.96	0.10	
Objects*Frequency	0.32	0.14	0.03	0.00	0.19	
Total Step 1 + Step 2	38.10	5.70	26.10	19.80	15.30	

* $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence

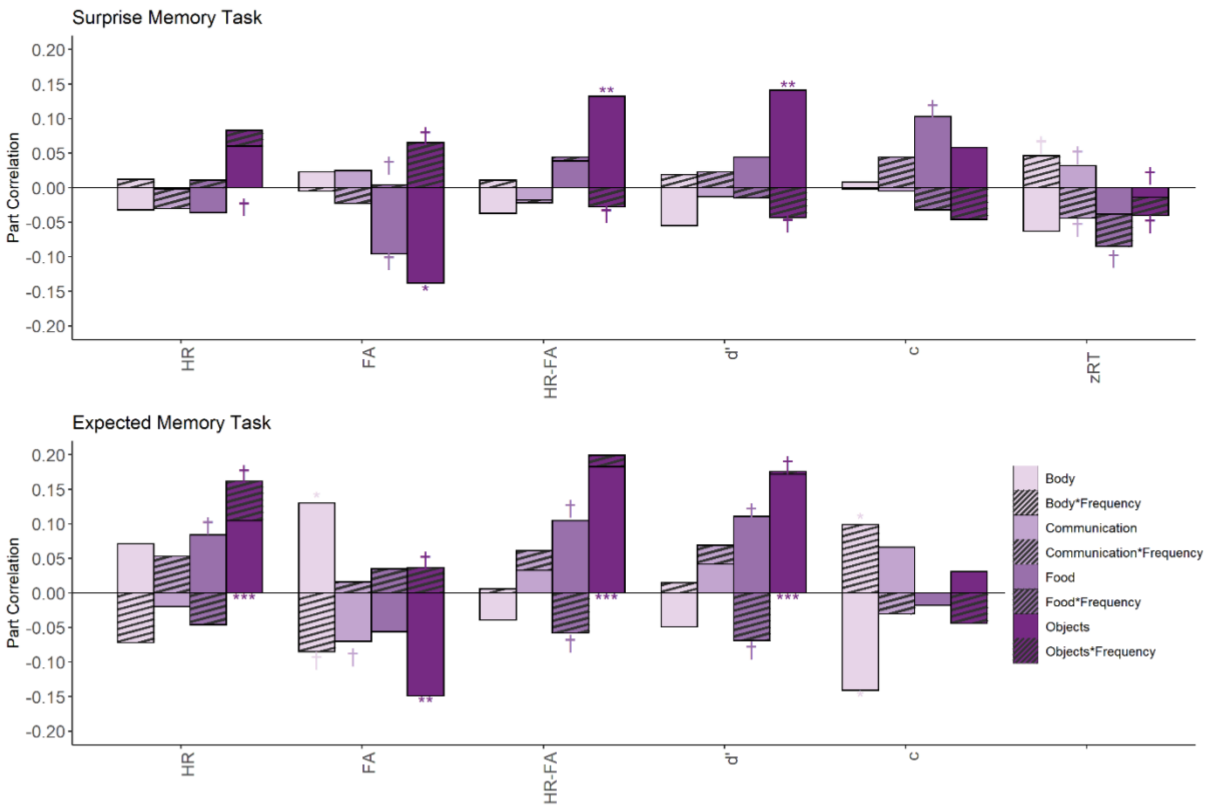


Figure 6: Part correlation of semantic components' contribution to memory performance from Step 2 (final) model, on a surprise memory task (left) and an expected memory task (right). Stacked bars represent the combined effect of sensorimotor component (lighter shade) and the sensorimotor*Frequency component interaction (darker shade) in the Step 2 (final) regression model. The symbols per bar refer to the inclusion Bayes Factor (BF) of each predictor: *** $BF_{10} \geq 150$ constitutes very strong evidence; ** $BF_{10} \geq 20$, strong evidence; * $BF_{10} \geq 3$ positive evidence; † $BF_{10} \geq 0.33$ and <3.00, equivocal evidence; no symbol $BF_{10} < 0.33$, evidence against inclusion.

Surprise memory task. In the analysis of the surprise memory task, the components varied in the size and direction of the effect they elicited, but the inclusion of most of them as reliable predictors was not supported by Bayes Factors. Overall, the model with all sensorimotor components explained from 0.9% of variance in HR to 3.3% of variance in FA.

Based on the part correlation analysis (see Figure 6), the Body and Communication components did not produce any effects on memory performance, apart from the equivocal evidence for their effects on RT. Words rated highly on the Body component were recognised faster, but this was only the case for low Frequency words, while words rated higher on the Communication component were responded to slower, and the effect was weaker for low Frequency words.

The Food and Object components elicited effects which were not consistent across different measures. The Food component had no effect on HR or the composite performance measures: HR-FA and *d′* sensitivity. There was equivocal evidence that higher Food ratings led to lower FA, and that this was somewhat attenuated for high Frequency words. There was equivocal evidence that high Food scores also predicted higher bias (*c*), that is, words strong in Food experience were less likely to be judged as "old", regardless of Frequency. There was equivocal evidence that words rated higher on the Food component were processed faster, but only when the words were of high Frequency.

The Object component elicited the strongest effects on memory performance. There was equivocal evidence that words rated higher on the Object component were easier to correctly judge as "old", regardless of their Frequency. They were also less likely to be incorrectly judged as "old" and there was equivocal evidence that this effect was stronger for low Frequency words. High Object scores predicted better performance on the overall performance measures, HR-FA and *d'*, although there was equivocal evidence that this effect was attenuated for high Frequency words. Object ratings did not affect response bias, but there was equivocal evidence that words higher in Object strength were faster to process, with the effect increasing for high Frequency words.

Expected memory task. Overall, the sensorimotor components explained between 2.8% of total variance in bias to 6.2% variance in HR-FA. Part correlations showed that the Body component did not elicit any effects on Hit Rates or the composite memory measures: HR-FA and *d'*. Words rated higher on the Body component were more likely to be incorrectly judged as "old", and there was equivocal evidence that this effect was smaller for high Frequency words. High Body words were also more biased towards being judged as "old" regardless of whether they had been seen previously, and this effect was smaller for high Frequency words.

The Communication component did not affect memory, apart from False Alarms, where there was equivocal evidence that high Communication scores predicted lower FA, regardless of word Frequency. The effects elicited by the Food component were also low, only supported by equivocal Bayesian evidence: high Food scores predicted higher HR, regardless of Frequency, and higher performance on overall performance measures (HR-FA, *d'*), where the effect was attenuated for higher Frequency words. The Food component had no effect on FA or bias.

Finally, the Object component predicted higher Hit Rates, and there was equivocal evidence that the effect was larger for high Frequency words. High Object scores predicted lower FA, with equivocal evidence for smaller effects when words were high in Frequency. Performance on the composite memory measures (HR-FA and *d'*) was also higher for words related more strongly to Objects, with equivocal evidence that the effect on *d'* was stronger for high Frequency words. The Object component did not elicit any effect on bias.

Conclusion

A surprise memory task was used to eliminate the element of prior knowledge about the task. To eliminate the differences in results and interpretation due to a different sample size, we also analysed a subset of data from Study 1a for comparison. Overall, the surprise memory task had smaller sensorimotor effects than the expected task on the same items. The pattern of differences shed some light on the processes behind word recognition memory performance. For example, while the Body component had no effects in the surprise memory task, it inflated FA and led to lower bias in the expected memory task, that is, when participants knew that they were supposed to memorise the words, higher Body strength made them judge the word as “old” regardless of whether it was previously seen. This suggests that Body strength contributed to the sense of having seen the word of the item. However, the Body effects on different measures were weaker in the current word subset than in Study 1a, suggesting that the sample of nearly 500 words was not enough to detect effects, or was not representative of the larger item set analysed in Study 1. Further research using the surprise memory task paradigm is needed.

The effects of the Food component revealed a tendency that participants were less likely to mistake words strong in Food ratings for “old” (equivocal evidence for lower FA rate and higher *c*) when they did not know that they were supposed to memorise the words, and that those words were processed faster. This was unexpected because these effects were not present in Study 1a, despite a much larger sample of words. The sensorimotor simulation of Food-related words must have been stronger during the test phase than the study phase – a lexical decision task which did not require deep semantic processing. This might have led participants to reject the possibility of having seen the word with greater confidence. Again, further research is needed to investigate the nature of the simulation in different types of task.

The effects of the Object component were also somewhat stronger in the expected memory task, and there was some inconsistency in the direction of the interaction with Frequency in the composite measures, although this was only supported by equivocal Bayesian evidence. The only exception was bias, which in Study 1a was affected negatively by the Object component, but in the current sample it was not. Since the Object component is closely linked to imageability (r=.66; see Figure 3), and imageability also had a positive effect on sensitivity (*d’*; see Figure 4), these effects were in line with the idea that imageability ratings may reflect experiences associated with interacting with manipulable objects (Dymarska et al., 2021).

The results of the expected memory task were broadly in line with the direction of results in Study 1a. The Communication component did not elicit any effects, which supported the conclusion that strength of experience with sound, mouth action, head action, or interoceptive experience does not influence whether a word is easier to remember. There was equivocal evidence for the role of the Food component in facilitating memory performance, likely due to a smaller sample size. The Body and Object components elicited effects in the same direction, suggesting that the different levels of support for the effects between Study 1a and Study 3a can be attributed to sample size differences. The general pattern of weaker effects of sensorimotor components on memory performance when participants performed a surprise memory task could be attributed to the fact that when participants were not actively trying to remember words, their processing relied more on lexical information and less detailed semantic representations in the study phase. We will discuss this in more detail in the general discussion. Overall, the results again did not support the semantic richness theory, as the effects of different aspects of sensorimotor experience varied in their tendency to facilitate or inhibit memory performance.

Study 3b): Effects of Imageability

In Study 2 we found that imageability had some effect on memory over sensorimotor information, although lower than reported in Cortese et al. (2010; 2015) where sensorimotor information was not included (2.3%-8.0% in study 2 vs 14%-24% in Cortese et al. of FA and HR, respectively). However, in Study 3a we found that the effects of sensorimotor strength were much lower when participants did not know they were going to perform a memory task. We therefore wanted to see how imageability performs as a predictor of a surprise memory task over and above sensorimotor information. Again, in order to compare these results with an expected memory task on the same sample of words, we also conducted the above analyses on the subset of 498 words from the dataset of Study 1a, which featured expected memory task data from Cortese et al. (2010; 2015) as dependent variables. We predicted that imageability would be useful in the expected memory task as shown in Study 2, because consciously generating imagery (e.g., visualising an apple upon reading the word “apple”) may be a useful strategy when participants know they will be tested on their memory of the words. Indeed, it has been found that imagery skills increase memory capacity (Keogh & Pearson, 2011; 2014). On the other hand, when participants are not expecting to be tested on memory for the words, they may not be consciously generating mental imagery during word processing (since word processing does not reliably rely on generating mental imagery, Dymarska et al., 2021) and therefore imageability will not be useful in a surprise memory task. This study follows the analysis used in Study 2.

Method

Materials. We used the same data as in Study 3a, as well as imageability ratings used in Study 2. We opted for the imageability ratings from Cortese et al. (2010; 2015) due to an overlap between our memory data and the Cortese imageability ratings, and because they also produced the strongest effects of imageability on word memory in Study 2. The same online memory task measures from Study 3a were used as DVs.

Design and Analysis. Using the hierarchical item-level linear regressions from Study 3a, we entered an additional Step 3, where we entered imageability and its interaction with the Frequency component. We report statistics from Bayesian regressions, as well as part correlation coefficients from the NHST regression, as in Study 3a. Again, the analysis was conducted on the surprise memory task data, and the subset of expected memory task data.

Results and Discussion

Results for the Step 1 and 2 of the regression, which reflect the contribution of lexical and sensorimotor components, are the same as in study 3a.

Table 9: Percentage variance explained by each predictor in the surprise memory task and an expected memory task. Unique variance is calculated with squared part correlations, and total variance is calculated with R^2 change between model.

	HR	FA	HR-FA	*d'*	*c*	zRT
			Surprise memory task			
Step 1: Lexical baseline	32.70***	1.50	20.50***	17.00***	9.70***	2.80*
Step 2: Sensorimotor (ΔR^2)	0.90	3.30	2.30	2.60	1.80	1.40
Step 3: Imageability (ΔR^2)	2.30***	4.00***	5.20***	5.70***	1.20	2.20*
Imageability parameter (sr^2)	2.34	3.24	5.02	5.48	0.40	1.46
Imageability*Frequency (sr^2)	0.38	0.05	0.08	0.10	0.45	0.16
Total Step 1-3 R^2	35.90	8.80	28.00	25.30	12.70	6.40
			Expected memory task			
Step 1: Lexical baseline	34.00***	0.80	19.90***	13.90***	12.50***	NA
Step 2: Sensorimotor (ΔR^2)	4.10	4.90	6.20**	5.90**	2.80	
Step 3: Imageability (ΔR^2)	7.80***	0.70	6.60***	5.20***	1.80*	
Imageability parameter (sr^2)	6.60	0.19	4.93	3.65	1.82	
Imageability*Frequency (sr^2)	0.03	0.26	0.27	0.34	0.29	
Total Step 1-3 R^2	45.90	6.40	32.70	25.00	17.10	

Note: Total variance expressed as R^2 change between the model with and without sensorimotor components. * $BF_{10} \geq 3$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence

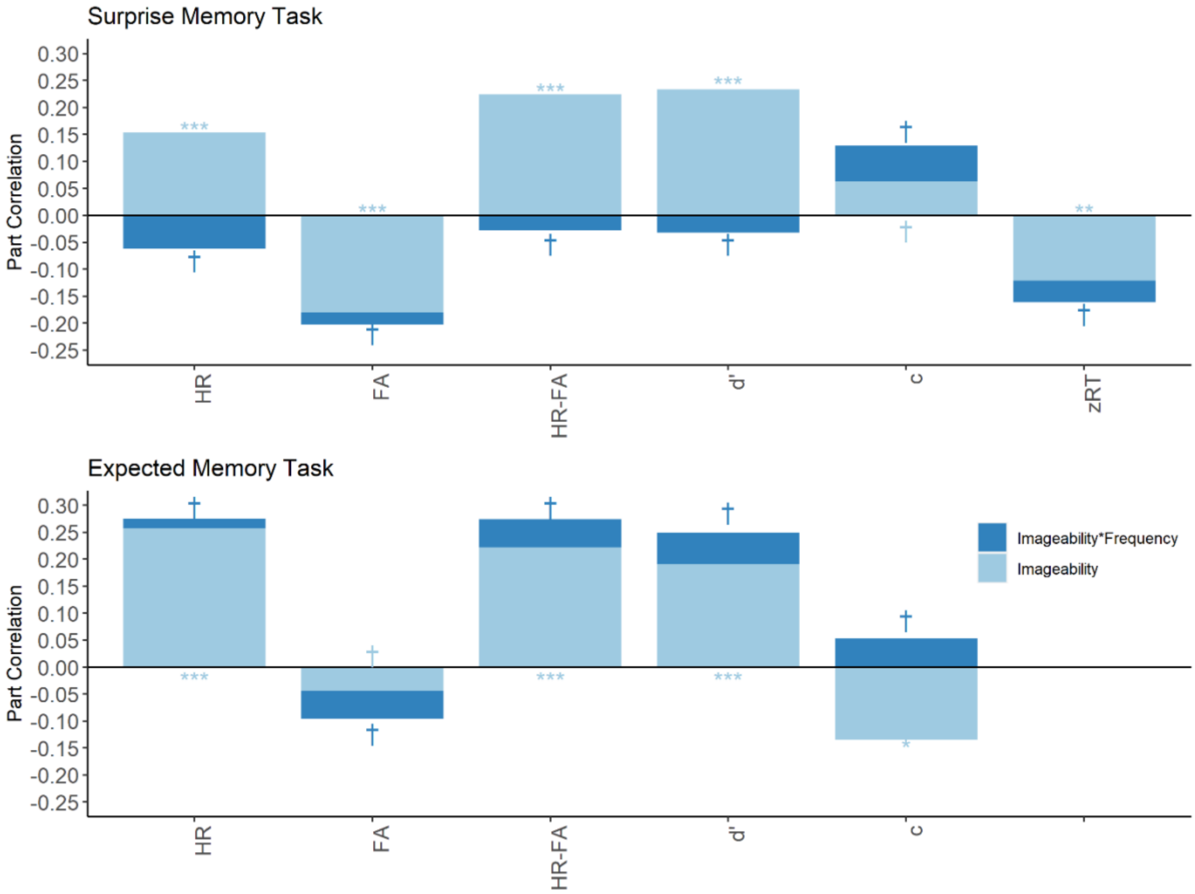


Figure 7: Part correlations of imageability's contribution to word memory in Step 3 (final) model, in a surprise memory task (top) vs an expected memory task (bottom). Stacked bars represent the combined effect of imageability component (light shade) and the imageability*Frequency component interaction (dark shade) in the Step 3 (final) regression model. The symbols per bar refer to the inclusion Bayes Factor (BF) of each predictor: *** $BF_{10} \geq 150$ constitutes very strong evidence; ** $BF_{10} \geq 20$, strong evidence; * $BF_{10} \geq 3$, positive evidence; † $BF_{10} >= 0.33$ and <3.00, equivocal evidence; no symbol $BF_{10} < 0.33$, evidence against inclusion.

Surprise memory task. Imageability facilitated memory performance on the surprise memory task, over and above the sensorimotor components. Part correlations showed that words higher in imageability were more likely to be correctly recognised as "old", with equivocal evidence that this was attenuated for high Frequency words, and less likely to be incorrectly judged as "old", with equivocal evidence that the effect was stronger for high Frequency words. Imageability also elicited positive effects on the composite memory performance measures, HR-FA and *d'*, with equivocal evidence that the effect was smaller for high Frequency words. There was equivocal evidence that imageability led to a more conservative response bias and evidence that it led to faster RT, that is, words with higher imagery scores were less likely to be judged as “old”, irrespective of whether they had been seen in the study phase, and were processed faster. These effects were stronger for high Frequency words, although this was also supported by equivocal evidence only.

Expected memory task. In the expected memory task imageability also facilitated word memory. High imageability predicted higher HR, with equivocal evidence that the effects were stronger for high Frequency words. The effect of imageability on FA was small – there was only equivocal evidence that words with high imageability scores were less likely to be incorrectly judged as “old”, and that this was stronger for high Frequency words. The overall performance measures, HR-FA and *d’*, were both predicted to be higher for high imageability words, with equivocal evidence for the effect being stronger for high Frequency words. Finally, higher imageability ratings led to lower bias, that is, words with high imageability ratings were more likely to be judged as “old”, regardless of whether they had been seen in the study phase. There was equivocal evidence that the effect was attenuated for high Frequency words. The effects of imageability in the expected memory task were smaller in magnitude than the effects of Cortese in Study 2, but followed the same direction.

Conclusion

We investigated whether imageability had an effect on memory over sensorimotor information when participants were not able to rely on conscious strategies in a surprise recognition memory task. We also used the same regression analyses and predictors to analyse the same words using a subset of data from Cortese et al. (2010; 2015) for comparison. We found that imageability had an effect on memory over and above sensorimotor strength in both tasks, consistent with the semantic richness theory, but the effects varied between tasks and measures. Overall, when participants knew that they were expected to remember the words, ease of generating mental imagery facilitated their performance more: imageability predicted 7.8% of total variance in HR, compared to 2.3% in the surprise memory task, and 6.6% of variance in overall performance measure (HR-FA,) compared to 5.2% in the surprise memory task. Similar to Study 2, we found that the effect of imageability, regardless of the type of task, was smaller than previously reported in Cortese et al. (2010; 2015). This supports the hypothesis that previously found effects of imageability could partially be attributed to perceptual and action information associated with the words.

While the effects of imageability on HR, HR-FA, and *d'* followed the same direction in both tasks, there were inconsistencies in the way that imageability interacted with Frequency, although the interactions were only supported by equivocal evidence and thus should be interpreted with caution. Additionally, the effect of higher imageability ratings leading to lower FA was stronger in the surprise memory task, suggesting that participants who used imageability as a strategy to memorise words (in the expected task) were then less able to reject distractors and judge them as "new" when they were rated higher in imageability. Finally, imageability had the opposite effects on bias in the two tasks. Consciously trying to remember words led participants to judge high imageability words as old more often. Participants in the expected memory task might have employed a strategy to imagine a word's referent in order to memorise it in the study phase, and therefore in the test phase high imageability ratings increased the feeling of having encountered the word previously across both old and new words. On the other hand, an unexpected encounter with high imageability words in the surprise memory task led participants to judge them as new more often, because they did not rely on consciously available information as a strategy, although evidence for this effect was equivocal.

Across Studies 3a-b, we also found that both sensorimotor strength and imageability produced smaller effects in a surprise memory task, suggesting that both sensorimotor grounding and an imagery strategy are independently useful when participants are actively trying to remember a word, compared to being exposed to words in a lexical decision task.

General discussion

In a series of exploratory analyses of word memory, we examined whether the contribution of sensorimotor information and imageability to word recognition memory performance was in line with the semantic richness theory (Buchanan et al., 2001). We used data from a mega-study of word recognition memory (Cortese et al., 2010; 2015) and an online surprise word recognition memory task. Sensorimotor strength contributed to word memory, but the effects varied depending on the type of experience, which was not in line with the semantic richness theory. Additionally, we found that imageability contributed to memory over and above sensorimotor information, but these effects varied between different rating datasets, as found in previous research (Dymarska et al., 2021). The effects of imageability were also smaller than those found in previous literature.

In this paper, we used a novel method of investigating how sensorimotor grounding affects word memory by examining how sensorimotor strength relating to various forms of experience (Body, Communication, Food, Objects) can predict word memory performance. The semantic richness theory predicted that greater sensorimotor strength would lead to richer representations and a stronger memory trace of sensorimotor simulation, which should make it easier to recognise previously-seen words. We found evidence against this prediction: it appears that different forms of sensorimotor information can sometimes help and sometimes hinder word memory. Simulating sensorimotor information about a concept associated with Food experience or manipulable Object experience did facilitate memory. This type of experience was useful at activating the memory trace of the studied word, which made words rated higher on sensorimotor strength more memorable. This was further supported by the findings that sensorimotor effects were weaker in the surprise memory task than the expected memory task, because the prior lexical decision task required a weaker or less detailed sensorimotor activation compared to intentionally memorising a list of words; therefore the trace activation after a lexical decision task is less helpful in the word recognition memory test. On the other hand, strength of Communication experience did not affect word memory at all. In Study 1a we also found that words rated high on the Body component were falsely judged as “old” when they were not previously studied, suggesting that Body strength elicited a bias on words, creating an illusion of having seen the word before. Further, we found that this could not be attributed to simply choosing “old” whenever a word high in Body strength was encountered, because the Body strength affected Hit Rate less than False Alarms, and led to negative discrimination – that is, signal and noise were differentiated. Hence, the biasing effect was not strong enough when the word had actually been studied, and there was a memory trace of sensorimotor representation associated with the word. It only elicited an effect when the word was new, and participants were misled into judging it as previously seen based on the strength of Body experience associated with the word. The results show that the semantic richness theory is not fully supported for memory. When sensorimotor information is analysed in detail, the way it influences word processing in a memory task is not as simple and straightforward as the semantic richness prediction: “more is better”.

This is a novel finding showing how sensorimotor information affects memory. A possible explanation for this pattern of results could be the degree of complexity of some concepts. When a concept is represented mainly with one modality (e.g., is strongly visual, such as “light” or “tree”), that modality can indeed facilitate memory recognition by strengthening the memory trace. For example, the Object component, which facilitated word memory, is associated with manipulable objects and visual and haptic experience, and these are rated high on modality exclusivity (Lynott & Connell, 2009; Speed & Majid, 2017; Vergallito et al., 2020), that is, words experienced with vision or touch are likely to be experienced through that one sense alone, perhaps making the experience and its simulation more distinct. On the other hand, some concepts are strongly multimodal and experienced through a number of senses, which may increase the impression that the word was encountered in study phase when it is presented in the test phase. This was the case for the Body component, which increased both Hit rate and False Alarm rate, and is associated with complex experiences, such as interoceptive and motor dimensions, which often come together to represent, for example, "running" (action to move legs, feeling of increased heartbeat and muscle fatigue). Future investigation of modality exclusivity effects could shed light on this pattern.

Alternatively, the unexpected effects of the Body component and its role in memory performance may be attributed to the ability of our body to serve as a contextual cue. Previous findings from memory research suggest that bodily movement and experience is implicated in enhancing memory traces. For example, the use of gestures, both by a participant or by an observed speaker/actor, has been found to facilitate memory for words or observed actions (Iani & Buciarelli; 2018). Similarly, body posture facilitated recall of situations associated with a similar posture (Dijkstra et al., 2007). Stooping posture also led to recall of more negatively valenced memories, due to its association with a negative mood (Veenstra et al., 2016), and performing a twisting or pressing action after encoding facilitated subsequent retrieval of verbs associated with the same action (van Dam et al., 2013). Thus, bodily experience seems to serve as a strong retrieval cue. Also relevant to the current finding, motor interference (finger tapping) disrupted recall and re-interpretation of novel tactile stimuli (Kamermans et al., 2019). That is, when the motor processing system was occupied, the manipulation and evaluation of the memory trace was not possible. This raises the possibility that in the current study, when participants encountered a strong Body word, they simulated it to make a memory judgment, but the process of simulation was confused with a retrieval of a strong memory trace, and gave rise to the belief that the word had been seen. Crucially, this interpretation is further supported by the findings that the illusion of a strong memory trace is dependent on the vividness of the memory (here, the strength of sensorimotor experience associated with the memory). Participants judge more vivid memories as true even when they have reason to believe that the memory is not real (i.e., it is not plausible to have happened, or they were told that it was not real by a family member or by an experimenter, in the case of autobiographical or experimentally induced memories, respectively; Otgaar et al., 2013; Scoboria et al., 2004). It is, however, important to distinguish that the Body effect was not a pure bias towards saying "old" when participants saw a word high in Body strength. Instead, seeing a new Body word likely elicited a simulation which was perceived as strong enough to judge it as previously encountered. Critically, when an "old" word with strong Body experience was encountered, the existing memory trace did its job of informing the response to some extent, which made it less susceptible to the biasing interference of activating a simulation (or whatever process was going on when a new Body word was presented). This is a novel and unexpected finding, which reveals complicated effects of different types of sensorimotor information on memory performance and will need to be investigated further in future research.

In addition to the role of sensorimotor simulation, we investigated whether we could replicate the effects of imageability on memory, as it has also been used as a proxy for semantic experience, and was found to be a strong predictor of memory in a number of studies (Caplan & Madan, 2016; Cortese et al., 2010; 2015; Fliessbach et al., 2006; Majerus & Van der Linden, 2003; Paivio, 1969). Our previous research found that imageability ratings do not reliably measure sensorimotor information and are not good predictors of word recognition (Dymarska et al., 2021). In the current studies, imageability facilitated word memory, in line with the predictions of the semantic richness effect. In Study 2 and the analysis of a subset of items in Study 3, when participants were explicitly asked to learn a list of words, consciously generating imagery was a good strategy and the ease of generating such imagery was therefore a good predictor of memory performance. Notably, imageability effects also emerged in the analysis of a surprise memory task, independent of sensorimotor grounding, which suggests that conscious generation of imagery played a role in the memory trace beyond a conscious strategy to memorise words. Since semantic activation in lexical decision is unconscious and automatic (i.e., sensorimotor information of word meaning is implicitly retrieved but not necessarily at a level of activation that is available to conscious awareness), it seems unlikely that our participants in Study 3 took a radically different approach of invoking unnecessary conscious imagery to make a word/non-word judgement. However, it is possible that they did so during the recognition memory part of the task. When asked if a word was old or new, participants generated conscious imagery of the word's referent, and words that are easier to image may have had stronger sensorimotor trace activation than words that are difficult to image, and hence benefited from both sensorimotor strength and imageability. Ease of generating mental imagery can thus be considered a useful predictor of word memory.

In the current study, we also found that some of the imageability effects in published literature can be attributed to sensorimotor grounding. Compared to Cortese et al. (2010; 2015) the effects were 50% smaller once sensorimotor information was included in the model, and it appears that the previous effects of imageability had been overestimated. Nonetheless, even when sensorimotor information was accounted for, the ease of generating a mental image had an effect on memory. These results indicate that some, but not all effects of imageability can be attributed to sensorimotor grounding. As demonstrated in Study 2, sensorimotor strength effects can be larger than those of some imageability norms in a conventional word memory task where participants study a list of words and expect to be tested on them later.

However, the pattern may have been reversed in a surprise memory task in study 3a, where sensorimotor information was activated less strongly when participants were exposed to words during a lexical decision task. That is, participants in the surprise memory task only attended to word meaning in a relatively shallow way during prior lexical decision, just enough to make the word/non-word judgement. The sensorimotor simulations activated during lexical decision are detailed enough to trigger semantic facilitation effects as measured by both perceptual strength (cf. Connell & Lynott, 2012; 2014; 2016; Connell et al., 2018) and action strength (Lynott et al. 2019), and as we found in this study, they can be detailed enough to facilitate word memory, at least for sensorimotor information relating to object experience. However, learning a list of words which needed to be remembered later (as in the expected memory task) most likely required deeper processing of word meaning than making a word/non-word judgement, that is, the resulting sensorimotor simulation was likely to be more highly activated and more detailed, which then facilitated word memory to a greater extent. Whether or not participants were engaging in a deliberate imagery strategy during the task (i.e., consciously generating a mental image of the word's referent), the differential task demands during the study phase of the expected versus surprise memory task were enough to produce differences in activation of sensorimotor information in word meaning, which then in turn led to differences in sensorimotor effects on the subsequent word recognition memory task. In future research, a different kind of task, focusing more on engaging detailed semantic representations, such as a semantic decision task, could be an alternative predecessor of a surprise memory task when examining the effects of sensorimotor strength on word memory.

In conclusion, the current study provides important theoretical and methodological insights into the study of word memory. First, it shows that the semantic richness effect, which predicts a facilitation of semantic memory by any semantic variable, does not fully explain word memory performance. While some aspects of sensorimotor strength do indeed facilitate word memory, others actually have the opposite effect, and the strong involvement of sensorimotor information in semantic processing may increase the sense of having seen the word, rather than the support the memory for whether the word is old or new. Second, imageability norms from different sources do not produce equally strong effects in word memory, but it nonetheless appears that generating conscious imagery contributes to word memory, when it is involved in a conscious effort to remember words. Finally, both sensorimotor information and imageability play a larger role and are more useful in an expected memory task than a surprise memory task. The decreased effects of sensorimotor information could be attributed to the nature of the lexical decision task presented to familiarise participants with the word, where participants did not have to rely on sensorimotor representation in as much detail, which led to retaining a weaker memory trace and smaller effects on recognition memory. On the other hand, the higher effects of imageability in some measures of the expected memory tasks could be attributed to participants employing a strategy to image words in the study phase. This could create the illusion of having seen the word before when it was rated highly on imageability, regardless of whether the word had been presented in the task or not. Together, these findings indicate that semantic effects in word memory are more complex than previous research suggested, as they encompass multisensory and motor information, as well as mental imagery, and are dependent on conscious and unconscious encoding strategies. In order to fully understand these complex effects, a multidimensional approach to semantic information is needed when investigating its effects in memory, rather than a single “all-in-one” variable which cannot capture the full semantic experience of the world.

References

Anwyl-Irvine, A.L., Massonié J., Flitton, A., Kirkham, N.Z., & Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods, 52,* 388–407. https://doi.org/10.3758/s13428-019-01237-x

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Nelson, D. L., Simpson, G. B. & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445-459. https://doi.org/10.3758/bf03193014

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing?. In: Adelman JS (Ed.), *Visual word recognition volume 1: Models and methods, orthography and phonology*. Hove, England: Psychology Press.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*(1), 73–79. https://doi.org/10.3758/BF03195349

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, *1*(1), 122–125. https://doi.org/10.3758/BF03200768

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin and Review*, *8*(3), 531–544. https://doi.org/10.3758/BF03196189

Caplan, J. B., & Madan, C. R. (2016). Word imageability enhances association-memory by increasing hippocampal engagement. *Journal of Cognitive Neuroscience*, *28*(10), 1522–

1538. https://doi.org/10.1162/jocn

Chiarello, C., Shears, C., & Lund, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 603–637. https://doi.org/10.3758/BF03200739

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 371–383. https://doi.org/10.3758/BF03195584

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*(3), 452–465. https://doi.org/10.1016/j.cognition.2012.07.010

Connell, L., & Lynott, D. (2014). I see/hear what you mean: Semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General*, *143*(2), 527–533. https://doi.org/10.1037/a0034626

Connell, L., & Lynott, D. (2016). Do we know what we're simulating? Information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1218–1232. https://doi.org/10.1037/xlm0000245

Connell, L., Lynott, D., & Banks, B. (2018). Interoception: the forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 20170143. https://doi.org/10.1098/rstb.2017.0143

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 384–387. https://doi.org/10.3758/BF03195585

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18*(6), 595–609. https://doi.org/10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*, *68*(8), 1489–1501. https://doi.org/10.1080/17470218.2014.945096

Davis, C. J., & Stadthagen-Gonzalez, H. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598–605

Dutriaux, L., Dahiez, X., & Gyselinck, V. (2018). How to change your memory of an object with a posture and a verb. *Quarterly Journal of Experimental Psychology*, 174702181878509. https://doi.org/10.1177/1747021818785096

Dymarska, A., Connell, L., & Banks, B. (2021). Imageability effects and variance in lexical decision and word naming Tasks. [Manuscript in preparation]. *Department of Psychology, Lancaster University.*

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage*, *32*(3), 1413–1421. https://doi.org/10.1016/j.neuroimage.2006.06.007

Hamilton, M., & Rajaram, S. (2001). The Concreteness effect in implicit and explicit memory tests. *Journal of Memory and Language*, *44*(1), 96–117. https://doi.org/10.1006/jmla.2000.2749

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(2), 130-136.

JASP Team (2020). *JASP (Version 0.14.1) [Computer software]*.

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology*, *64*(9), 1683–1691.

https://doi.org/10.1080/17470218.2011.605150

Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *PloS One*, *6*(12), e29221. https://doi.org/10.1371/journal.pone.0029221

Keogh, R., & Pearson, J. (2014). The sensory strength of voluntary visual imagery predicts visual working memory capacity. *Journal of Vision*, *14*(12), 1-13. https://doi.org/10.1167/14.12.7

Klaver, P., Fell, J., Dietl, T., Schür, S., Schaller, C., Elger, C. E., & Fernández, G. (2005). Word imageability affects the hippocampus in recognition memory. *Hippocampus*, *15*(6), 704-712. https://doi.org/10.1002/hipo.20081

Kounios, J., Green, D. L., Payne, L., Fleck, J. I., Grondin, R., & McRae, K. (2009). Semantic richness and the activation of concepts in semantic memory: Evidence from event-related potentials. *Brain Research*, *1282*, 95–102. https://doi.org/10.1016/j.brainres.2009.05.092

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *Quarterly Journal of Experimental Psychology*, *71*(10), 2207–2222. https://doi.org/10.1177/1747021817739834

Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin and Review*, *22*(5), 1328–1333. https://doi.org/10.3758/s13423-014-0791-2

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*(2), 558–564. https://doi.org/10.3758/BRM.41.2.558

Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*, *45*(2), 516-526. https://doi.org/10.3758/s13428-012-0267-0

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2019). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*, 1-21. https://doi.org/10.3758/s13428-019-01316-z

Madan, C. R. (2020). Exploring word memorability: How well do different word properties explain item free-recall probability?. *Psychonomic Bulletin & Review*, 1-13. https://doi.org/10.3758/s13423-020-01820-w

Majerus, S., & Van Der Linden, M. (2003). Long-term memory effects on verbal short-term memory: A replication study. *British Journal of Developmental Psychology*, *21*(2), 303–310. https://doi.org/10.1348/026151003765264101

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, *24*(10), 2099–2105. https://doi.org/10.1177/0956797613480803

Paivio, A. (1971). Imagery and language. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches.* (pp. 7-32). New York: Academic Press.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Paivio, A., Smythe, P. C., & Yuille, J. C. (1968). Imagery versus meaningfulness of nouns in paired-associate learning. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *22*(6), 427-441. https://psycnet.apa.org/doi/10.1037/h0082782

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, *76*(1), 1-12. https://psycnet.apa.org/doi/10.1037/h0025327

Pecher, D., van Dantzig, S., & Schifferstien, H. N. J. (2009). Concepts are not represented by conscious imagery. *Psychonomic Bulletin and Review*, *16*(5), 914–919. https://doi.org/10.3758/PBR.16.5.914

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin and Review*, *15*(1), 161–167. https://doi.org/10.3758/PBR.15.1.161

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Recchia, G., & Jones, M. N. (2012). The semantic richness of abstract concepts. *Frontiers in Human Neuroscience*, *6*, 1–16. https://doi.org/10.3389/fnhum.2012.00315

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition*, *14*(1), 79–94. https://doi.org/10.3758/BF03209231

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*(2), 374–379. https://doi.org/10.3758/s13428-011-0162-0

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322-339. https://doi.org/10.1037/met0000061

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2018). The Glasgow norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, *51*(3), 1258-1270. https://doi.org/10.3758/s13428-018-1099-3

Sidhu, D. M., & Pexman, P. M. (2016). Is moving more memorable than proving? Effects of embodiment and imagined enactment on verb memory. *Frontiers in Psychology, 7*, 1–13. https://doi.org/10.3389/fpsyg.2016.01010

Sidhu, D. M., Siakaluk, P. D., Pexman, P. M., Yap, M. J., & Muraki, E. (2018). Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior Research Methods, 51*(2), 453-466. https://doi.org/10.3758/s13428-018-1171-z

Snefjella, B., & Kuperman, V. (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition*, *156*, 135–146. https://doi.org/10.1016/j.cognition.2016.07.010

Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: Simulating perceptual modality in space. *Behavior Research Methods, 49*(6), 2204–2218. https://doi.org/10.3758/s13428-017-0852-3

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. https://doi.org/10.3758/BF03207704

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–278. https://doi.org/10.1016/j.neuroimage.2013.08.023

Taylor, R. (2017). *Concreteness, frequency, and bilingual language dominance: Implications for the impact of context availability in explicit memory* (Open Access Doctoral Thesis). University of Texas at El Paso, USA. https://scholarworks.utep.edu/open_etd/565

Tse, C. S., & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. *Memory*, *15*(6), 675–690. https://doi.org/10.1080/09658210701467186

Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*(2), 241-255. https://doi.org/10.1016/j.cogbrainres.2004.09.002

Vergallito, A., Petilli, M. A., & Marelli, M. (2020). Perceptual modality norms for 1,121 Italian words: A comparison with concreteness and imageability scores and an analysis of their impact in word processing tasks. *Behavior Research Methods, 52*, 1599–1616. https://doi.org/10.3758/s13428-019-01337-8

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*(5), 1256–1271. https://doi.org/10.1037/0278-7393.25.5.1256

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Milton Bache, S., Muller, K., Ooms, J., Robinson, D., Paige Seidel, D., Spinu, V. ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive science*, *29*(5), 719-736. https://doi.org/10.1207/s15516709cog0000_33

Xiao, X., Zhao, D., Zhang, Q., & Guo, C. yan. (2012). Retrieval of concrete words involves more contextual information than abstract words: Multiple components for the concreteness effect. *Brain and Language*, *120*(3), 251–258. https://doi.org/10.1016/j.bandl.2011.09.006

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience, 6*, 1–10. https://doi.org/10.3389/fnhum.2012.00072

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin and Review*, *18*(4), 742–750. https://doi.org/10.3758/s13423-011-0092-y

Zdrazilova, L., & Pexman, P. M. (2013). Grasping the invisible: Semantic processing of abstract words. *Psychonomic Bulletin and Review*, *20*(6), 1312–1318. https://doi.org/10.3758/s13423-013-0452-x

7 Thesis Discussion & Conclusions

7.1 Thesis Aims

This thesis examined the role of sensorimotor and linguistic information in conceptual processing in a number of tasks: working memory for objects, word memory, and word recognition. Both sensorimotor simulation and linguistic information contribute to conceptual representations (Barsalou et al., 2008; Connell & Lynott, 2013; 2014b), and their activation depends on task demands and available resources (Connell, 2018; Louwerse & Connell, 2011). Thus, when resources are limited, such as in a working memory task, linguistic labels may be employed as placeholders for complex sensorimotor representations, as predicted by the novel linguistic bootstrapping hypothesis (Connell & Lynott, 2014b). On the other hand, sensorimotor strength may be useful when access to semantic information is required to facilitate a lexical decision, as it provides more detailed information about the word's meaning (Barsalou et al., 2003). More detailed sensorimotor information may support a stronger memory trace in a word memory task, as predicted by the semantic richness theory (Buchanan et al., 2001; Pexman et al., 2008). Further, consciously available information about the word's meaning, which is captured by imageability, is thought to facilitate word processing in word recognition and word memory tasks (Cortese & Khanna, 2007; Cortese & Schock, 2013; González-Nosti et al., 2014). In the current thesis, I examined behavioural data from cognitive and psycholinguistic tasks to investigate how sensorimotor and linguistic information, accessed consciously or otherwise, come together to support conceptual representations used in those tasks. This was captured by three main aims of the thesis:

1. To examine whether the linguistic component of the conceptual system could bootstrap the capacity of working memory.

2. To study whether consciously available information about concepts supports word recognition.
3. To investigate whether sensorimotor information facilitates word memory.

Aim 1 was addressed in Chapter 4. I used an object recognition memory task to test working memory capacity for complex, contextually related, everyday-life objects when language is available and when it is not. Working memory capacity was indeed higher when linguistic placeholders could be employed, supporting the linguistic bootstrapping hypothesis. Around 10 objects were remembered when only sensorimotor information was available, and up to two more objects were remembered when language was also available. This was in line with the idea that language can serve as a bootstrapping mechanism for complex cognitive tasks, such as memory for real-life object sequences.

Turning to the second aim, I first examined a number of imageability ratings in their ability to predict word recognition (Chapter 5). Performance varied between different sets of norms and different tasks and measures, with some norms eliciting little effect, or inhibiting rather than facilitating word recognition. The imageability ratings themselves were also not consistently predicted by different aspects of sensorimotor experience, suggesting that they are not reliably measured in the first place, and reflect different types of experience with a concept, depending on the source of ratings. In Chapter 6, I also examined whether consciously generating mental imagery contributes to a stronger memory trace which then facilitates memory for words, as predicted by the semantic richness theory. A comparison of a surprise and an expected memory task revealed that imageability affects memory performance differently depending on task demands. When generating imagery can be explicitly used as a strategy to remember words, it facilitates word memory performance. On the other hand, when the study phase involves less detailed semantic representations, the strength of the memory trace is not influenced by imageability to the same extent, and participants are less likely to be driven by imageability when making a word judgement.

In Chapter 6 I then examined the contribution of sensorimotor information to word memory (as per aim 3). Data from an expected and a surprise memory task showed that not all types of sensorimotor information produced a facilitation effect on word memory, which was not in line with the semantic richness theory. Instead, while information related to object manipulation was useful in creating a stronger memory trace when encountering the word, information related to bodily experience elicited an unexpected effect. Participants judged words rated highly on body experience as previously seen when they were actually new words, even more so than when they were indeed previously seen. In other words when a new high body strength word was encountered, it gave participants a false sense of having seen the word before, in addition to facilitating memory for old high body strength words.

7.2 Contribution

7.2.1 Linguistic and sensorimotor contribution to memory

This thesis provides a number of novel contributions to the study of conceptual representations and their role in different conceptual tasks. First, the aspect of the concept that is represented during processing depends on task demands, as suggested by Connell and Lynott (2014b). Chapter 4 demonstrated that when resources are limited, relying on linguistic information allows us to remember a larger amount of information. Specifically, a linguistic label, which is smaller in size, can be used as a placeholder for a complex sensorimotor representation when holding information in working memory. The label itself can then be used to perform a task (e.g., asking for apples at the shop), or it can activate sensorimotor information needed for the task (e.g., looking for apples on the shelf).

This is a novel finding, as previous research indicated that working memory capacity for each specific type of information (e.g., objects) is fixed. It is also at odds with the prediction of dual-coding theory (Paivio, 1971), which proposes that encoding information with both sensorimotor and linguistic information (two codes) should always outperform encoding with either one of them individually, regardless of task demands. On the contrary, I found that the type of information and the availability of resources did matter. Initially, the type of information used at encoding made little difference to accuracy of performance. However, once participants had to hold more than 10 object concepts in mind, the working memory capacity for sensorimotor information was exceeded and participants had to employ the strategy of using a linguistic shortcut. This demonstrated the advantage of label availability to deputise for complex sensorimotor information.

On the other hand, in Chapter 6 I found that a memory task relies on sensorimotor information when a more detailed conceptual representation can facilitate processing. In a task which involved a more detailed semantic representation, such as retrieving word meaning from long-term memory, participants relied on sensorimotor information more than when superficial lexical decision about the word form had to be made. These findings put into question the role of semantic richness in accounting for word memory. Stronger sensorimotor experience facilitated word memory, at least when measured as experience with manipulable objects and nouns, reflected in haptic, visual or hand strength. However, other types of sensorimotor experience either did not elicit an effect on word recognition memory performance, or affected different measures of memory in different ways. For example, sensorimotor experience associated with the body was activated strongly during word processing, so much so that it led participants to make incorrect judgements about the words they had not studied. That is, it appears that participants were representing experience with the word associated with the body every time they were presented with a new word, and were likely using the strength of experience to inform their decision about whether the word was previously seen, which led to an inflated false sense of having studied the word. However, this was independent of the actual memory trace of the word. That is, if a word was indeed studied, then bodily strength supported a strong memory trace of that experience, and the judgment was not conflated with the general strength of representation of body related experiences. The discrepancy between the effects of different components could be attributed to participant strategies associated with representing concepts during the word recognition memory task, but also to the nature of the experience associated with different concepts. For example, visual experience, which tends to produce the strongest effects in visual tasks (Connell & Lynott, 2014a) was concentrated in the Object component. It also represents the type of information that is most easily available in conscious awareness, since it is strongly correlated with imagery ratings (see Chapter 5, Study 2). Indeed, the Object component elicited the strongest effects in word recognition, in particular the lexical decision task, in line with previous research. On the other hand, olfactory or gustatory information concentrated in the Food component, while important in many ways, is not necessarily activated in language-based task (Speed & Majid, 2018), which would explain its lower effects in word recognition performance.

Overall, it appears that sensorimotor experience plays an important role in conceptual processing, but not in a consistent manner. It is possible that information about auditory, olfactory or gustatory strength is not as easy to maintain as a memory trace as is the simulation of visual, haptic or interoceptive experience. The influence of sensorimotor information on word recognition memory depends on which aspect of the concept and which perceptual or motor dimension is involved.

7.2.2 Situating the findings in imageability research

The findings of Chapters 5 and 6 provided a novel insight into the concept of imageability and its role in conceptual processing. Defined as ease of generating a mental image, it represents information that comes to mind when encountering the name of a concept and can be used to support interpretation of its meaning. This extends to word recognition and word memory. However, as previously suggested by Connell & Lynott (2016) and Pecher et al. (2009), it appears that information about word meaning which is used in semantic tasks is not equal to consciously available semantic information, because people are unable to accurately report what kind of experience guides their conceptual processing. I extended this idea (in Chapter 6) to show that the exact nature of the contribution of conscious imagery is dependent on task demands, much like the contribution of sensorimotor and linguistic information. That is, when imagery is used as a strategy to reconstruct the experience of encountering a concept (e.g., in a study phase), it facilitates word memory performance. It also has a stronger effect when it is used actively as a strategy to enhance word memory through visualization/imagery when the memory task is expected. However, this is not the case when the task pertains to just lexical processing, where processing lexical information is sufficient to make a lexical decision or a word naming judgment. Ease of generating conscious imagery does not consistently make a word easier or faster to recognise or name.

Additionally, in Chapter 5 I evaluated a number of existing sets of imageability ratings in their contribution to conceptual processing, as well as what sort of sensorimotor information they reflect. The findings shed light on the problem of reliability of imageability as a construct. Contrary to previous research, imageability did not consistently facilitate word recognition. More surprisingly, imageability did not provide a boost in the form of semantic information to low frequency words, when frequent experience of encountering the word form was not available. This led to a further discovery that imageability ratings were not consistently predicted by the same types of sensorimotor experience. Depending on the word and the participant doing the rating, different forms of experience may come to mind when evaluating ease of imageability. Putting aside the problem of imageability being a strongly visual variable, sometimes a particular type of experience was captured well by the ratings, other times it did not predict them at all, overall indicating that the type of experience that is consciously available and represented may vary across words and across people, and therefore may not be captured well by a single semantic variable (cf. Connell & Lynott, 2016a).

In terms of the use of imageability ratings in psycholinguistic research, the findings suggest that participants are not able to reliably measure their experience with a concept when asked about the ease of generating mental imagery. Imageability is also not a reliable predictor of word recognition. However, in Chapter 6 imageability did have a fairly reliable effect on word memory. This suggests that despite the variability inherent to imageability ratings, they retain enough useful information to predict some variance in word memory, possibly because word memory benefits more from consciously available information, and this type of information is captured more reliably in the ratings. The decision to use imageability ratings in research on conceptual processing should therefore be made with caution. While it appears useful in certain types of tasks where relying on strategies to represent a more detailed semantic representation is required, using imageability in psycholinguistic research on lexical decision is likely to lead to unreliable results.

7.3 Limitations and future research

The experiments presented in Chapter 4 focused on the working memory capacity for object concepts, as well as the role that linguistic labels play in memory for concepts. The experiments revealed a differential effect of access to language on the encoding and retrieval processes in working memory. This warrants a further investigation into the way that linguistic and sensorimotor representations are activated during the encoding phase, how they are maintained in working memory while waiting to be retrieved, as well as how they interact with information from the environment to enable a response. As this was one of the first studies to combine working memory research with the linguistic-simulation approach to conceptual representations, a lot of questions remain unanswered. For example, in the current study all objects in a sequence were presented in the same order at both encoding and retrieval. However, it would be interesting to see how participants perform when stimuli are not in order. In that paradigm, any effect of stimuli acting as cues to each other, for example by pairing adjacent stimuli (chunking) or by remembering that “potato came after carrot” would be eliminated. While in many real-life contexts such as recipes or instructions objects are usually presented in a fixed order, other things, such as shopping lists, may have to be adapted and the order manipulated based on the order in which things are organized in a particular shop, for example. Thus, both a fixed order and a random order experimental paradigm would provide some insight into the capacity limits and capabilities of working memory, as well as separate the support offered by object-specific sensorimotor information from the support offered by the simulation of a broader situational context.

The studies in Chapters 5 and 6 demonstrated that sensorimotor information plays a role in retrieving conceptual information about words from semantic memory. However, only Chapter 6 investigated how different aspects of sensorimotor experience influenced the memory trace and the decision-making processes in a memory task. The same components representing sensorimotor information of different kinds could be used to investigate the role of sensorimotor information in word recognition in more detail. While there has been plenty of research into the role of semantic variables in word recognition (Balota et al., 2004; Buchanan et al., 2001; Cortese & Schock, 2013; Pexman et al., 2008; Yap et al., 2011), no other study, to my knowledge, has tested the semantic richness theory using multi-dimensional sensorimotor experience captured by multiple variables. Much like the results in Chapter 6, it is possible that only some aspects of sensorimotor experience facilitate word recognition. Whether some sensorimotor experience could also potentially hinder word recognition is less likely, since word recognition and word memory require a different type of decision, and it seems only logical that strength of sensorimotor experience would facilitate retrieval of information about a word from semantic memory. Nonetheless, research into this is required to rule out any unexpected effects that might take place, and it is possible that only selected dimensions actually play a role in this process. More broadly, using multi-variable sensorimotor strength measures appears to give a more nuanced insight into cognitive processing, and should be considered in other areas of research.

The analysis of the surprise memory task in Chapter 6 came from a relatively small pilot study. Although 500 words may be larger than other traditional memory research had used in the past, it is still much smaller than the dataset of over 5000 words provided by Cortese et al. (2010; 2015). Indeed, the comparison of the 500 word subset of the dataset suggested that it was not always enough to detect effect sizes present in the larger sample. Thus, in order to fully understand the difference in performance between the two types of tasks, a larger dataset of the surprise memory task should be collected. This will allow researchers to investigate the effects of using explicit strategies to remember stimuli, as well as to better understand the role of different semantic information in supporting a memory trace. Additionally, using a different type of task in the study phase could shed light on the implicit remembering processes – instead of a shallow lexical decision task, participants could be asked to imitate an action or to act out some information associated with the word, to boost the likelihood of engaging in a sensorimotor simulation. Other tasks that have been used to study semantic effects in conceptual processing have been for example, a semantic classification task – making a judgment on whether a word belongs to a particular semantic category, or a semantic decision task – a concrete/abstract word judgment (Pexman et al., 2008; Pexman et al., 2017). Alternatively, relying on paradigms commonly used in grounded cognition research, such as stimulating hands or feet (e.g., Connell et al., 2012), or presenting stimuli from a particular modality to keep it occupied (Nazir et al., 2008), could affect how information from different modalities is processed, and in turn allow us to examine the involvement of each perceptual modality and action effector in retrieving a memory.

Finally, the word memory research could benefit from studying linguistic distributional information as a factor which can affect how well words are remembered. While Chapter 6 has shed some light on how sensorimotor simulation supports memory (or when it does not), the linguistic part of the linguistic-simulation system is yet to be investigated in as much detail. Research into the role of frequently co-occurring pairs has shown long ago that frequency of co-occurrence, much like frequency of word occurrence, facilitates memory (Stuart & Hulme, 2000). However, with the currently available methods to use data from large corpora, as well as highly nuanced measures of distribution of words within a semantic space, this could be taken further. Establishing the exact level of co-occurrence needed to facilitate word chunking, or when distributional distance supports memory, would be an interesting next step. This could also include the linguistic distributional density of the word, which in the current research was merely input into the PCA as one of the many variables, but which could be investigated as an individual, novel predictor of word memory, among other things.

7.4 Concluding remarks

This thesis aimed to investigate the role of linguistic and sensorimotor information in conceptual representations, and how they interact to support retrieval of information from long-term memory, as well as maintaining information in working memory. It demonstrated that both linguistic and sensorimotor information support conceptual processing to varying extents and in different situations. Linguistic information supports working memory when cognitive resources are strained, while sensorimotor information affects the memory trace and the decision-making process when retrieval from long-term memory takes place, for example in word recognition and word memory tasks. This is mostly due to unconscious simulations, but consciously available information can support processing when the task allows for employing a strategy to improve performance.

The current work provides a starting point for future research on the intersection of memory and grounded cognition, which can be taken in many different directions by bringing the findings together. The thesis demonstrates that across different tasks, linguistic and sensorimotor information interacts in supporting conceptual representations in short-term memory, whether to remember an item, or to activate a representation in order to make a lexical judgment. However, it is clear that as discussed in section 2.1 of the thesis, further research is needed, in particular on the motor aspects of conceptual representations. The current findings about the role of information from different dimensions in conceptual processing could begin to inform these.

In Chapter 6, a clear subjective effect on judgment of body-related information emrged, which overrides the real old/new distinction of studied and non-studied words. One possible reason for that effect, albeit speculative, could be that motor information is reconstructed during processing. More specifically, when participants encounter a motor-related word during a study phase and a later test phase, the representation is not stable enough to be relied on over a period of time, and therefore the representation of, say, "running" is activated anew. This affects the vividness of the memory compared to a perceptual representation which is held in mind continuously. This idea is in line with the findings that motor interference (such as performing a finger tapping task) impairs processing of motor concepts in real time (Boulenger et al., 2006), while subsequent processing is independent (Nazir et al., 2008), and that strength of motor experience elicits a larger pupil response than the objective memory trace (Montefinese et al., 2013). Similarly, when participants hear a sentence "you gave a pencil to Tom", the speed of their sensibility judgment is not affected by the direction of response (towards or away from the body; Morey et al., 2021), because the representations of “gave a pencil to” and “press the response button” are constructed separately. This would then explain why the strength of the simulation during decision making overrides the strength of the memory trace of a motor-related concept, which is not neccessarily the case for other types of concepts.

The thesis also opens up new avenues for working memory research. In line with the patterns of interference when storing information within the visual, tacitle or motor dimensions, there seems to be a need to distinguish individual modalities through separate systems, in addition to the visuo-spatial sketchpad and phonological loop proposed by Baddeley and Hitch (1994). If that was the case, it would be expected that different types of sensorimotor information may be held in mind with varying levels of efficiency and varying capacity limits. For example, objects which afford interactions via vision or touch (as those captured by the Object component in the current studies), should be more easily maintained with an accurate and vivid memory trace, while for concepts which rely on motor experience (e.g., moving the arms or feet), the memory trace would be more easily disrupted by interference from similar concepts, or from ongoing processing. There is also scope to study the interaction between language and action in working memory more extensively (see e.g., Banks & Connell., 2021). The benefit of using labels to describe the complex experiences could be most apparent in motor processing, as labels can be stored elsewhere, outside of the motor simulation system, in a more efficient and stable manner, and thus boost working memory capacity for actions. Future research, in particular into processing of motor concepts, is needed to investigate these issues in more detail.

The behavioural studies in this thesis were pre-registered where possible. All materials, data, and analysis code are available online at:

https://osf.io/8sm3p/?view_only=853111f26db840b4a0558275496d77e9 (Chapter 4)

https://osf.io/hrj3c/?view_only=289c468c6ae4496f9f8771f9ea5a0ec3 (Chapter 5)

https://osf.io/rjdng/?view_only=7ce5801c9eeb4a08a7e678724fbc9b1a (Chapter 6)

List of Figures

List of Tables

Consolidated Bibliography

Adams, E. J., Nguyen, A. T., & Cowan, N. (2018). Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. *Language, Speech, and Hearing Services in Schools*, *49*(3), 340–355. https://doi.org/10.1044/2018_LSHSS-17-0114

Adelman, J. S., Brown, G. D. A., & Quesada, F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823.

Allen, R. J., Havelka, J., Falcon, T., Evans, S., & Darling, S. (2015). Modality specificity and integration in working memory: Insights from visuospatial bootstrapping. *Journal of Experimental Psychology: Learning Memory and Cognition, 41*(3), 820–830. https://doi.org/10.1037/xlm0000058

Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2009). Cross-modal binding and working memory. *Visual Cognition*, *17*(1-2), 83–102. https://doi.org/10.1080/13506280802281386

Allen, T. W., Marcell, M. M., & Anderson, P. (1978). Modality-specific interference with verbal and nonverbal stimulus information. *Memory & Cognition*, *6*(2), 184-188. https://doi.org/10.3758/BF03197444

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106–111. https://doi.org/10.1111%2Fj.0963-7214.2004.01502006.x

Ambrosi, S., Kalenine, S., Blaye, A., & Bonthoux, F. (2011). Modality switching cost during property verification by 7 years of age. *International Journal of Behavioral Development, 35*(1), 78–83. https://doi.org/10.1177/0165025410371603

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science, 6*(3), 359–370. https://doi.org/10.1111/tops.12096

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological review, 116*(3), 463.

Anwyl-Irvine, A.L., Massonié J., Flitton, A., Kirkham, N.Z., & Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods, 52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Atkinson, R. C., & Shiffrin, R. M. (1971). The Control of Short-Term Memory. *Scientific American*, *225*(2), 82–91.

Atkinson, R. C., Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In Spence, K. W., Spence, J. T. (Eds.), *The psychology of learning and motivation* (pp. 89–195). New York: Academic Press.

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology, 16*(18), 1818–1823. https://doi.org/10.1016/j.cub.2006.07.060

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

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 49*(1), 5–28. https://doi.org/10.1080/713755608

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2002). Fractionating the central executive. In D. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 246–260). New York: Oxford University Press

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In K. D. Federmeier and E. R. Schotter (Eds), *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485–493. https://doi.org/10.1037/0894-4105.8.4.485

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*(3), 438-456. https://doi.org/10.1016/j.jml.2009.05.004

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press.

Baddeley, A., Logie, R., Bressi, S., Sala, S. D., & Spinnler, H. (1986). Dementia and working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 603-618. https://doi.org/10.1080%2F14640748608401616

Bakhtiar, M., & Weekes, B. (2014). Lexico-semantic effects on word naming in Persian: Does age of acquisition have an effect? *Memory and Cognition*, *43*(2), 298–313. https://doi.org/10.3758/s13421-014-0472-4

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283-316. https://doi.org/10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing?. In: Adelman JS (Ed.), *Visual* *word recognition volume 1: Models and methods, orthography and phonology*. Hove, England: Psychology Press.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Nelson, D. L., Simpson, G. B. & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445-459. https://doi.org/10.3758/bf03193014

Banks, B. & Connell, L. (2021). Language enhances working memory for action events. [Manuscript in preparation]. *Department of Psychology, Lancaster University*.

Banks, B., Wingfield, C., & Connell, L. (2020). Linguistic distributional knowledge and sensorimotor grounding both contribute to semantic category production. Retrieved from PsyArXiv database. https://doi:10.31234/osf.io/7eucw

Bao, M., Li, Z.-H., & Zhang, D.-R. (2007). Binding facilitates attention switching within working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(5), 959–969. https://doi.org/10.1037/0278-7393.33.5.959

Barnard, P. J. (1999). Interacting Cognitive Subsystems: Modeling working memory phenomena within a multiprocessor architecture. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 298–339). Cambridge University Press.

Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J. C., Costumero, V., Parcet, M. A., & Ávila, C. (2012). Reading salt activates gustatory brain regions: FMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex, 22*(11), 2554–2563. https://doi.org/10.1093/cercor/bhr324

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

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

Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society*, *373*(1752). https://doi.org/10.1098/rstb.2017.0144

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, and A. C. Graesser (Eds), *Symbols, embodiment, and meaning* (pp. 246-283). Oxford: Oxford University Press.

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

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher and R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129-163). Cambridge University Press.

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*(4), 477–487. https://doi.org/10.3758/BF03197179

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain and Language, 98*(2), 127–139. https://doi.org/10.1016/j.bandl.2006.04.008

Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., & Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic nouns. *Behavior Research Methods, 43*(4), 1100–1109. https://doi.org/10.3758/s13428-011-0117-5

Bestgen, Y., & Vincze, N. (2012). Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavior Research Method, 48*, 998–1006. https://doi.org/10.3758/s13428-012-0195-z

Bidet-Ildei, C., Meugnot, A., Beauprez, S. A., Gimenes, M., & Toussaint, L. (2017). Short-term upper limb immobilization affects action-word understanding. *Journal of Experimental Psychology: Learning Memory and Cognition, 43*(7), 1129–1139. https://doi.org/10.1037/xlm0000373

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience, 17*(6), 905-917. https://doi.org/10.1162/0898929054021102

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33*(1), 73–79. https://doi.org/10.3758/BF03195349

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently?. *Cognition*, *118*(1), 123-129. https://doi.org/10.1016/j.cognition.2010.09.010

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2018). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews, 29*, 120–153. https://doi.org/10.1016/j.plrev.2018.12.001

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts*. New York, NY: Springer.

Borghi, A. M., Cimatti, F. (2009). Words as Tools and the Problem of Abstract Word Meanings. *Proceedings of the Annual Meeting of the Cognitive Science, 31*, 276–281.

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory and Cognition*, *32*(6), 863–873. https://doi.org/10.3758/BF03196865

Borreggine, K. L., & Kaschak, M. P. (2006). The Action-Sentence Compatibility Effect: It's all in the timing. *Cognitive Science, 30*(6), 1097–1112. https://doi.org/10.1207/s15516709cog0000_91

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia, 46*(2), 743–756. https://doi.org/10.1016/j.neuropsychologia.2007.10.007

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of cognitive neuroscience*, *18*(10), 1607-1615. https://doi.org/10.1162/jocn.2006.18.10.1607

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, *1*(1), 122–125. https://doi.org/10.3758/BF03200768

Bower, G. H. (1970). Analysis of a Mnemonic Device: Modern psychology uncovers the powerful components of an ancient system for improving memory. *American Scientist, 58*(5), 496–510.

Bozeat, S., Ralph, M. A. L., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*(9), 1207-1215. https://doi.org/10.1016/S0028-3932(00)00034-8

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*(4), 487–502. https://doi.org/10.1037/a0016797

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings* *of the National Academy of Sciences, 113*(27), 7459-7464. https://doi.org/10.1073/pnas.1520027113

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 157–165. https://doi.org/10.1037/0278-7393.18.1.157

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica, 104*(2), 215–226. https://doi.org/10.1016/S0001-6918(00)00021-4

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research, 24*(3), 355–363. https://doi.org/10.1016/j.cogbrainres.2005.02.020

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin and Review, 8*(3), 531–544. https://doi.org/10.3758/BF03196189

Buchsbaum, B. R., & D'Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex, 112*, 134–148. https://doi.org/10.1016/j.cortex.2018.11.010

Caplan, J. B., & Madan, C. R. (2016). Word Imageability enhances association-memory by Increasing hippocampal engagement. *Journal of Cognitive Neuroscience, 28*(10), 1522–1538. https://doi.org/10.1162/jocn

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus, 133*(1), 59-68.

Chedid, G., Brambati, S. M., Bedetti, C., Rey, A. E., Wilson, M. A., & Vallet, G. T. (2019). Visual and auditory perceptual strength norms for 3,596 French nouns and their relationship with other psycholinguistic variables. *Behavior Research Methods, 51*(5), 2094–2105. https://doi.org/10.3758/s13428-019-01254-w

Chen, I. H., Zhao, Q., Long, Y., Lu, Q., & Huang, C. R. (2019). Mandarin Chinese modality exclusivity norms. *PloS One, 14*(2), e0211336. https://doi.org/10.1371/journal.pone.0211336

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology, 62*(7), 1420–1429. https://doi.org/10.1080/17470210802453977

Chiarello, C., Shears, C., & Lund, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 603–637. https://doi.org/10.3758/BF03200739

Chiou, R., Humphreys, G. F., Jung, J., & Lambon Ralph, M. A. (2018). Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control' and hub-and-spoke 'semantic representation' systems. *Cortex*, *103*, 100-116. https://doi.org/10.1016/j.cortex.2018.02.018

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*(1-3), 83-121. https://doi.org/10.1016/0010-0277(91)90033-Z

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2014). Arousal and emotional valence interact in written word recognition. *Language, Cognition and Neuroscience*, *29*(10), 1257–1267. https://doi.org/10.1080/23273798.2014.897734

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 371–383. https://doi.org/10.3758/BF03195584

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240-247. https://doi.org/10.1016/S0022-5371(69)80069-1

Collins, J., Pecher, D., Zeelenberg, R., & Coulson, S. (2011). Modality switching in a property verification task: An ERP study of what happens when candles flicker after high heels click. *Frontiers in Psychology, 2*, 1–10. https://doi.org/10.3389/fpsyg.2011.00010

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33*(4), 497-505. https://doi.org/10.1080%2F14640748108400805

Connell, L. (2018). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience, 34*(10), 1308-1318. https://doi.org/10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2012a). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*(3), 452–465. https://doi.org/10.1016/j.cognition.2012.07.010

Connell, L., & Lynott, D. (2012b). When does perception facilitate or interfere with conceptual processing? The effect of attentional modulation. *Frontiers in Psychology, 3*, 1–4. https://doi.org/10.3389/fpsyg.2012.00474

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin and Review, 20*(3), 542–550. https://doi.org/10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014a). I see/hear what you mean: Semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General, 143*(2), 527–533. https://doi.org/10.1037/a0034626

Connell, L., & Lynott, D. (2014b). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science, 6*(3), 390–406. https://doi.org/10.1111/tops.12097

Connell, L., & Lynott, D. (2016a). Do we know what we're simulating? Information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning Memory and Cognition, 42*(8), 1218–1232. https://doi.org/10.1037/xlm0000245

Connell, L., & Lynott, D. (2016b). Embodied semantic effects in visual word recognition. In M. H. Fischer & Y. Coello (Eds.), *Foundations of embodied cognition* (Vol. 2, pp. 71–92). Hove, UK: Psychology Press.

Connell, L., Lynott, D., & Banks, B. (2018). Interoception: The forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170143. https://doi.org/10.1098/rstb.2017.0143

Connell, L., Lynott, D., & Dreyer, F. (2012). A functional role for modality-specific perceptual systems in conceptual representations. *PLoS One, 7*(3), e3321. https://doi:10.1371/journal.pone.0033321

Cornell Kärnekull, S. C., Jönsson, F. U., Willander, J., Sikström, S., & Larsson, M. (2015). Long-Term Memory for Odors : Influences of Familiarity and Identification Across 64 Days. *Chemical Senses, 40*(4), 259–267. https://doi.org/10.1093/chemse/bjv003

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*(3), 384–387. https://doi.org/10.3758/BF03195585

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60*(8), 1072-1082.

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609. https://doi.org/10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*, *68*(8), 1489–1501. https://doi.org/10.1080/17470218.2014.945096

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology, 66*(5), 946–972. https://doi.org/10.1080/17470218.2012.722660

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4*(2), 226-231.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin, 104*(2), 163-191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why. *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cumming, G. (2012). *Understanding the new statistics effect sizes, confidence intervals, and meta-analysis* (1st ed., Multivariate applications book series). New York: Routledge.

Damasio, A. R. (1989a). The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones. *Neural Computation, 1*(1), 123–132.

Damasio, A. R. (1989b). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33*(1-2), 25–62. https://doi.org/10.1016/0010-0277(89)90005-X

Damasio, H., Tranel, D., & Grabowski, T. (2004). Neural systems behind word and concept retrieval. *Cognition, 92*(1-2), 179–229. https://doi.org/10.1016/j.cognition.2002.07.001

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. https://doi.org/10.1016/S0022-5371(80)90312-6

Darling, S., Allen, R. J., Havelka, J., Campbell, A., & Rattray, E. (2012). Visuospatial bootstrapping: Long-term memory representations are necessary for implicit binding of verbal and visuospatial working memory. *Psychonomic Bulletin and Review, 19*(2), 258–263. https://doi.org/10.3758/s13423-011-0197-3

Darling, S., & Havelka, J. (2010). Visuospatial bootstrapping: Evidence for binding of verbal and spatial information in working memory. *Quarterly Journal of Experimental Psychology, 63*(2), 239–245. https://doi.org/10.1080/17470210903348605

Davis, C. J., & Stadthagen-Gonzalez, H. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*(4), 598–605

Davis, C. P., Joergensen, G. H., Boddy, P., Dowling, C., & Yee, E. (2020). Making it harder to "see" meaning: The more you see something, the more its conceptual representation is susceptible to visual interference. *Psychological Science, 31*(5), 505-517. https://doi.org/10.1177/0956797620910748

D'esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*(6554), 279-281. https://doi.org/10.1038/378279a0

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 824–845. https://doi.org/10.1037/0278-7393.15.5.824

de Hevia, M. D., Vallar, G., & Girelli, L. (2008). Visualizing numbers in the mind's eye: The role of visuo-spatial processes in numerical abilities. *Neuroscience and Biobehavioral Reviews*, *32*, 1361–1372. https://doi.org/10.1016/j.neubiorev.2008.05.015

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved*?. Memory & Cognition, 29*(2), 267-273. https://doi.org/10.3758/BF03194920

De Wijk, R. A., & Cain, W. S. (1994). Odor quality: discrimination versus free and cued identification. *Perception & Psychophysics, 56*(1), 12-18. https://doi.org/10.3758/BF03211686

Delaney, P. F., & Ericsson, K. A. (2016). Long-term working memory and transient storage in reading comprehension: What is the evidence? Comment on Foroughi, Werner, Barragán, and Boehm-Davis (2015). *Journal of Experimental Psychology: General, 145*(10), 1406–1409. https://doi.org/10.1037/xge0000181

Delvenne, J. F., & Bruyer, R. (2004). Does visual short-term memory store bound features? *Visual Cognition, 11*(1), 1–27. https://doi.org/10.1080/13506280344000167

Desai, R. H., Reilly, M., & Van Dam, W. (2018). The multifaceted abstract brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373(*1752). https://doi.org/10.1098/rstb.2017.0122

Di Luca, S., & Pesenti, M. (2011). Finger numeral representations: more than just another symbolic code. *Frontiers in Psychology, 2*, 1–3. https://doi.org/10.3389/fpsyg.2011.00272

Dijkstra, K., Kaschak, M. P., & Zwaan, R. A. (2007). Body posture facilitates retrieval of autobiographical memories. *Cognition, 102*(1), 139–149. https://doi.org/10.1016/j.cognition.2005.12.009

Dreyer, F. R., & Pulvermüller, F. (2018). Abstract semantics in the motor system? – An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex, 100*, 52–70. https://doi.org/10.1016/j.cortex.2017.10.021

Durda, K., Buchanan, L., & Caron, R. (2009). Grounding co-occurrence: Identifying features in a lexical co-occurrence model of semantic memory. *Behavior Research Methods, 41*(4), 1210–1223. https://doi.org/10.3758/BRM.41.4.1210

Dutriaux, L., Dahiez, X., & Gyselinck, V. (2018). How to change your memory of an object with a posture and a verb. *Quarterly Journal of Experimental Psychology*, *72*(5), 1112-1118. https://doi.org/10.1177/1747021818785096

Dymarska, A., Connell, L., & Banks, B. (2021). Imageability effects and variance in lexical decision and word naming tasks. [Manuscript in preparation]. *Department of Psychology, Lancaster University.*

Dymarska, A., Connell, L., & Banks, B. (2021). Linguistic bootstrapping in object memory [Data set]. Open Science Framework. http://osf.io/mwzfh

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245. https://doi.org/10.1037/0033-295X.102.2.211

Ericson, T. D. & Mattson M. E. (1981). From Words to Meaning: A Semantic Illusion. *Journal of Verbal Learning and Verbal Behaviour, 20,* 540-551.

Evans, G. A. L., Ralph, M. A. L., & Woollams, A. M. (2012). What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin and Review, 19*(2), 325–331. https://doi.org/10.3758/s13423-011-0213-7

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013). Where is the action? Action sentence processing in Parkinson's disease. *Neuropsychologia, 51*(8), 1510-1517. https://doi.org/10.1016/j.neuropsychologia.2013.04.008

Fernandino, L., Humphries, C. J., Seidenberg, M. S., Gross, W. L., Conant, L. L., & Binder, J. R. (2015). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia, 76*, 17–26. https://doi.org/10.1016/j.neuropsychologia.2015.04.009

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage, 32*(3), 1413–1421. https://doi.org/10.1016/j.neuroimage.2006.06.007

Fodor, J. A. (2008). The modularity of mind: An essay on faculty psychology. In J. E. Adler & L. J. Rips, (Eds), *Studies of Human Inference and Its Foundations* (pp. 878-914). Cambridge University Press. (Original work published 1983)

Gaillard, V., Destrebecqz, A., & Cleeremans, A. (2012). The influence of articulatory suppression on the control of implicit sequence knowledge. *Frontiers in Human Neuroscience, 6,* 1-9. https://doi.org/10.3389/fnhum.2012.00208

Gainotti, G. (2011). The organization and dissolution of semantic-conceptual knowledge: Is the "amodal hub" the only plausible model? *Brain and Cognition*, *75*(3), 299–309. https://doi.org/10.1016/j.bandc.2010.12.001

García, A. M., Moguilner, S., Torquati, K., García-Marco, E., Herrera, E., Muñoz, E., Castillo, E. M., Kleineschay, T., Sedeño, L., & Ibáñez, A. (2019). How meaning unfolds in neural time: Embodied reactivations can precede multimodal semantic effects during language processing. *NeuroImage, 197,* 439–449. https://doi.org/10.1016/j.neuroimage.2019.05.002

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395-427. https://doi.org/10.3758/BF03201693

Glenberg, A. M. (1997). Mental models, space, and embodied cognition. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (p. 495–522). American Psychological Association. https://doi.org/10.1037/10227-018

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

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *Journal of Neuroscience, 26*(18), 4917–4921. https://doi.org/10.1523/JNEUROSCI.5389-05.2006

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage*, *32*(2), 906-912.

González-Nosti, M., Barbón, A., Rodríguez-Ferreiro, J., & Cuetos, F. (2014). Effects of the psycholinguistic variables on the lexical decision task in Spanish: A study with 2,765 words. *Behavior Research Methods, 46*(2), 517–525. https://doi.org/10.3758/s13428-013-0383-5

Goodhew, S. C., McGaw, B., & Kidd, E. (2014). Why is the sunny side always up? Explaining the spatial mapping of concepts by language use. *Psychonomic Bulletin & Review*, *21*(5), 1287-1293. https://doi.org/10.3758/s13423-014-0593-6

Hamilton, M., & Rajaram, S. (2001). The concreteness effect in implicit and explicit memory tests. *Journal of Memory and Language, 44*(1), 96–117. https://doi.org/10.1006/jmla.2000.2749

Harnad, S. (1990) The Symbol Grounding Problem. *Physica D 42*, 335-346.

Harpaintner, M., Sim, E. J., Trumpp, N. M., Ulrich, M., & Kiefer, M. (2020). The grounding of abstract concepts in the motor and visual system: An fMRI study. *Cortex, 124*, 1–22. https://doi.org/10.1016/j.cortex.2019.10.014

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Hauk, O., Shtyrov, Y., & Pulvermüller, F. (2008). The time course of action and action-word comprehension in the human brain as revealed by neurophysiology. *Journal of Physiology-Paris*, *102*(1), 50–58. https://doi.org/10.1016/j.jphysparis.2008.03.013

Hayes, J. C., & Kraemer, D. J. M. (2017). Grounded understanding of abstract concepts: The case of STEM learning. *Cognitive Research: Principles and Implications*, *2*(1), 1-15. https://doi.org/10.1186/s41235-016-0046-z

He, K., Li, J., Wu, F., Wan, X., Gao, Z., & Shen, M. (2020). Object-based attention in retaining binding in working memory: Influence of activation states of working memory. *Memory and Cognition, 48*(6), 957–971. https://doi.org/10.3758/s13421-020-01038-0

Hendricks, R. K., & Boroditsky, L. (2017). New space–time metaphors foster new nonlinguistic representations. *Topics in cognitive science*, *9*(3), 800-818. https://doi.org/10.1111/tops.12279

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, *8*(5), 393-402. https://doi.org/10.1038/nrn2113

Hitch, G. J., Brandimonte, M. A., & Walker, P. (1995). Two types of representation in visual memory: Evidence from the effects of stimulus contrast on image combination. *Memory & Cognition*, *23*(2), 147-154. https://doi.org/10.3758/BF03197217

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory, 3*(3/4), 463-495. https://doi.org/10.1080/09658219508253161

Hoenig, K., Sim, E. J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience, 20*(10), 1799–1814. https://doi.org/10.1162/jocn.2008.20123

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718-730. https://doi.org/10.3758/s13428-012-0278-x

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-Frequency Effects on Short-Term Memory Tasks: Evidence for a Redintegration Process in Immediate Serial Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1217-1232.

Hume, D. (1978). *1739. A treatise of human nature.* London: John Noon. (Original work published 1739)

Hutter, R. R. C., Allen, R. J., & Wood, C. (2016). The formation of novel social category conjunctions in working memory: A possible role for the episodic buffer? *Memory, 24*(4), 496–512. https://doi.org/10.1080/09658211.2015.1020814

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 1*(2), 130-136.

Jaroslawska, A. J., Gathercole, S. E., & Holmes, J. (2018). Following instructions in a dual-task paradigm: Evidence for a temporary motor store in working memory. *Quarterly Journal of Experimental Psychology, 71*(11), 2439–2449. https://doi.org/10.1177/1747021817743492

JASP Team (2020). JASP *(Version 0.14.1)[Computer software]*.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*(1), 71-115. https://doi.org/10.1016/S0364-0213(81)80005-5

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology, 64*(9), 1683–1691. https://doi.org/10.1080/17470218.2011.605150

Kane, M.J., Bleckley, M.K., Conway, A.R.A., & Engle, R.W. (2001). A controlled-attention view of working-memory capacity. Journal of Experimental Psychology: General, 130, 169-183.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis?. *American Anthropologist*, *86*(1), 65-79. https://doi.org/10.1525/aa.1984.86.1.02a00050

Kellogg, R. T. (2001). Long-term working memory in text production. *Memory & Cognition*, *29*(1), 43-52. https://doi.org/10.3758/BF03195739

Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *PloS One, 6*(12), e29221. https://doi.org/10.1371/journal.pone.0029221

Keogh, R., & Pearson, J. (2014). The sensory strength of voluntary visual imagery predicts visual working memory capacity. *Journal of Vision*, *14*(12), 1-13. https://doi.org/10.1167/14.12.7

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*(1), 287-304. https://doi.org/10.3758/s13428-011-0118-4

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex, 48*(7), 805–825. https://doi.org/10.1016/j.cortex.2011.04.006

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

Kintsch, W., Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working memory in text comprehension. *Psychologia, 42*(4), 186–198.

Klaver, P., Fell, J., Dietl, T., Schür, S., Schaller, C., Elger, C. E., & Fernández, G. (2005). Word imageability affects the hippocampus in recognition memory. *Hippocampus, 15*(6), 704-712. https://doi.org/10.1002/hipo.20081

Koenigs, M., Acheson, D. J., Barbey, A. K., Solomon, J., Postle, B. R., & Grafman, J. (2011). Areas of left perisylvian cortex mediate auditory-verbal short-term memory. *Neuropsychologia, 49*(13), 3612–3619. https://doi.org/10.1016/j.neuropsychologia.2011.09.013

Kounios, J., Green, D. L., Payne, L., Fleck, J. I., Grondin, R., & McRae, K. (2009). Semantic richness and the activation of concepts in semantic memory: Evidence from event-related potentials. *Brain Research, 1282*, 95–102. https://doi.org/10.1016/j.brainres.2009.05.092

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The Representation of Abstract Words: Why Emotion Matters. *Journal of Experimental Psychology: General, 140*(1), 14–34. https://doi.org/10.1037/a0021446

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*(3), 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Kümmerer, D., Hartwigsen, G., Kellmeyer, P., Glauche, V., Mader, I., Klöppel, S., Suchan, J., Karnath, H-O., Weiller, C., & Saur, D. (2013). Damage to ventral and dorsal language pathways in acute aphasia. *Brain, 136*(2), 619-629. https://doi.org/10.1093/brain/aws354

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978-990. https://doi.org/10.3758/s13428-012-0210-4

Kuznetsova, A., Brockhoff, P. B., & Christensen, R.H.B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lakoff, G. (1993). Convergence zones and conceptual structure. Series: University of California, Berkeley, http://escholarship.org/uc/item/6n13989z

Latkowska, J. (2015). How Relevant is the Sapir-Whorf Hypothesis to Contemporary Psycholinguistic Research?. *Theory and Practice of Second Language Acquisition, 1*(1), 7-26.

Langerock, N., Vergauwe, E., & Barrouillet, P. (2014). The maintenance of cross-domain associations in the episodic buffer. *Journal of Experimental Psychology: Learning Memory and Cognition, 40*(4), 1096–1109. https://doi.org/10.1037/a0035783

Langerock, N., Vergauwe, E., Dirix, N., & Barrouillet, P. (2018). Is memory better for objects than for separate single features? The temporal hypothesis. *Journal of Experimental*

Psychology: Learning Memory and Cognition, 44(6), 898–917. https://doi.org/10.1037/xlm0000501

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source?. *The Quarterly Journal of Experimental Psychology, 56*(8), 1249-1268. https://doi.org/10.1080/02724980244000765

Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *Quarterly Journal of Experimental Psychology, 71*(10), 2207–2222. https://doi.org/10.1177/1747021817739834

Legge, E. L. G., Madan, C. R., Ng, E. T., & Caplan, J. B. (2012). Building a memory palace in minutes: Equivalent memory performance using virtual versus conventional environments with the Method of Loci. *Acta Psychologica*, *141*(3), 380–390. https://doi.org/10.1016/j.actpsy.2012.09.002

Lenci, A., Lebani, G. E., & Passaro, L. C. (2018). The emotions of abstract words: A distributional semantic analysis. *Topics in Cognitive Science*, *10*(3), 550–572. https://doi.org/10.1111/tops.12335

Lewis, J. W., Phinney, R. E., Brefczynski-Lewis, J. A., & DeYoe, E. A. (2006). Lefties get it “right” when hearing tool sounds. *Journal of Cognitive Neuroscience*, *18*(8), 1314-1330. https://doi.org/10.1162/jocn.2006.18.8.1314

Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin and Review, 22*(5), 1328–1333. https://doi.org/10.3758/s13423-014-0791-2

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin and Review, 15*(4), 838–844. https://doi.org/10.3758/PBR.15.4.838

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Louwerse, M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science, 35*(2), 381–398. https://doi.org/10.1111/j.1551-6709.2010.01157.x

Louwerse, M. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In M. deVega, A. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning: A debate* (pp. 309–326). Oxford, England: Oxford University Press

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition, 114*(1), 96–104. https://doi.org/10.1016/j.cognition.2009.09.002

Louwerse, M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in Psychology, 3,* 1–11. https://doi.org/10.3389/fpsyg.2012.00385

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281. https://doi.org/10.1038/36846

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology, 3*(54). https://doi.org/10.3389/fpsyg.2012.00054

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

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

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

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods, 41*(2), 558–564. https://doi.org/10.3758/BRM.41.2.558

Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods, 45*(2), 516-526. https://doi.org/10.3758/s13428-012-0267-0

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2019). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods, 52*, 1-21. https://doi.org/10.3758/s13428-019-01316-z

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin and Review, 23*(4), 1090–1095. https://doi.org/10.3758/s13423-015-0878-4

Madan, C. R. (2020). Exploring word memorability: How well do different word properties explain item free-recall probability?. *Psychonomic Bulletin & Review*, 1-13. https://doi.org/10.3758/s13423-020-01820-w

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris, 102*(1–3), 59–70. https://doi.org/10.1016/j.jphysparis.2008.03.004

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology, 60*, 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain ? *Trends in Cognitive Sciences*, *15*(3), 97–103. https://doi.org/10.1016/j.tics.2011.01.004

Majerus, S., & Van Der Linden, M. (2003). Long-term memory effects on verbal short-term memory: A replication study. *British Journal of Developmental Psychology*, *21*(2), 303–310. https://doi.org/10.1348/026151003765264101

Majid, A. (2020). Human olfaction at the intersection of language, culture, and biology. *Trends in Cognitive Sciences*, *25*(2), 1–13. https://doi.org/10.1016/j.tics.2020.11.005

Markiewicz, C. J., & Bohland, J. W. (2016). Mapping the cortical representation of speech sounds in a syllable repetition task. *NeuroImage*, *141*, 174–190. https://doi.org/10.1016/j.neuroimage.2016.07.023

Marschark, M., & Paivio, A. (1977). Integrative processing of concrete and abstract sentences. *Journal of Verbal Learning and Verbal Behavior*, *16*(2), 217-231. https://doi.org/10.1016/S0022-5371(77)80048-0

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25-45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Matheson, H. E. & Barsalou, L. (2018). Embodiment and grounding in cognitive neuroscience. In J. T. Wixted & S. Thompson-Schill (Eds), *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th Edition, Volume 3: Language and Thought). New York: Wiley. https://doi.org/10.1002/9781119170174.epcn310

May, J. (2001). Specifying the central executive may require complexity. In J. Andrade (Ed.), *Working Memory in Perspective* (pp. 261-278). Psychology Press.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310–322. https://doi.org/10.1038/nrn1076

McRae, K., Nedjadrasul, D., Pau, R., Lo, B. P. H., & King, L. (2018). Abstract concepts and pictures of real-world situations activate one another. *Topics in Cognitive Science, 10*(3), 518–532. https://doi.org/10.1111/tops.12328

Merck, C., Jonin, P., Laisney, M., & Vichard, H. (2014). When the zebra loses its stripes but is still in the savannah: Results from a semantic priming paradigm in semantic dementia. *Neuropsychologia, 53*, 221–232. https://doi.org/10.1016/j.neuropsychologia.2013.11.024

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. https://doi.org/10.1037/h0043158

Miller, T. M. C., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition, 171*, 172–179. https://doi.org/10.1016/j.cognition.2017.10.010

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The “subjective” pupil old/new effect: Is the truth plain to see? *International Journal of Psychophysiology, 89*(1), 48–56. https://doi.org/10.1016/j.ijpsycho.2013.05.001

Morey, R. D., Kaschak, M. P., Díez-Álamo, A. M., Glenberg, A. M., Zwaan, R. A., Lakens, D., Ibáñez, A., García, A., Gianelli, C., Jones, J. L., Madden, J., Alifano, F., Bergen, B., Bloxsom, N. G., Bub, D. N, Cai, Z. G., Chartier, C. R., Chatterjee, A. Conwell, E. ... & Ziv-Crispel, N. (2021). A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). *Psychonomic Bulletin & Review* (in press).

Morucci, P., Bottini, R., & Crepaldi, D. (2019). Augmented modality exclusivity norms for concrete and abstract Italian property words. *Journal of Cognition, 2*(1) 1-14. https://dx.doi.org/10.5334%2Fjoc.88

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81*(2), 111–121. https://doi.org/10.1111/j.2044-8295.1990.tb02349.x

Moseley, R., Carota, F., Hauk, O., Mohr, B., & Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cerebral Cortex*, *22*(7), 1634–1647. https://doi.org/10.1093/cercor/bhr238

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology, 21*, 263-267 https://doi.org/10.1037/h0082978

Myung, J., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano : manipulation knowledge of objects. *Cognition, 98*, 223–243. https://doi.org/10.1016/j.cognition.2004.11.010

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science, 24*(10), 2099–2105. https://doi.org/10.1177/0956797613480803

Nazir, T. A., Boulenger, V., Roy, A., Silber, B., Jeannerod, M., & Paulignan, Y. (2008). Language-induced motor perturbations during the execution of a reaching movement. *Quarterly Journal of Experimental Psychology*, *61*(6), 933-943. https://doi.org/10.1080%2F17470210701625667

Nisticò, R., Cerasa, A., Olivadese, G., Dalla Volta, R., Crasà, M., Vasta, R., Gramigna, V., Vescio, B., Barbagallo, G., Chiriaco, C., Quattrone, A., Salsone, M., Novellino, F., Arabia, G., Nicoletti, G., Morelli, M., & Quattrone, A. (2019). The embodiment of language in tremor-dominant Parkinson's disease patients. *Brain and Cognition*, *135*, 103586. https://doi.org/10.1016/j.bandc.2019.103586

Noorman, S., Neville, D. A., & Simanova, I. (2018). Words affect visual perception by activating object shape representations. *Nature: Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-32483-2

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

Oulasvirta, A., & Saariluoma, P. (2006). Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human Computer Studies*, *64*(10), 941–961. https://doi.org/10.1016/j.ijhcs.2006.04.006

Paivio, A. (1971). Imagery and language. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches*. (pp. 7-32). New York: Academic Press.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Paivio, A., Smythe, P. C., & Yuille, J. C. (1968). Imagery versus meaningfulness of nouns in paired-associate learning. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *22*(6), 427-441. https://psycnet.apa.org/doi/10.1037/h0082782

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1), 1-12. https://psycnet.apa.org/doi/10.1037/h0025327

Papesh, M. H. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. *Journal of Experimental Psychology: General*, *144*(6), 116-141. https://psycnet.apa.org/doi/10.1037/xge0000125

Parkin, A. J. (1998). The central executive does not exist. *Journal of the International Neuropsychological Society*, *4*(5), 518-522. https://doi.org/10.1017/S1355617798005128

Pecher, D., van Dantzig, S., & Schifferstien, H. N. J. (2009). Concepts are not represented by conscious imagery. *Psychonomic Bulletin and Review, 16*(5), 914–919. https://doi.org/10.3758/PBR.16.5.914

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin and Review, 11*(1), 164–167. https://doi.org/10.3758/BF03206477

Pederson, E. (2010). Cognitive linguistics and linguistic relativity. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics* (pp. 1012–1043). Oxford: Oxford University Press).

Peirce J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*(10), 1-8. https://doi.org/10.3389/neuro.11.010.2008

Peng, G., Deutsch, D., Henthorn, T., Su, D., & Wang, W. S. (2013). Language experience influences nonlinguistic pitch perception. *Journal of Chinese Linguistics*, *41*(2), 447-467.

Pexman, P. M. (2019). The role of embodiment in conceptual development. *Language, Cognition and Neuroscience, 34*(10), 1274–1283. https://doi.org/10.1080/23273798.2017.1303522

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). Neural correlates of concreteness in semantic categorization. *Journal of Cognitive Neuroscience*, *19*(8), 1407–1419. https://doi.org/10.1162/jocn.2007.19.8.1407

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin and Review, 15*(1), 161–167. https://doi.org/10.3758/PBR.15.1.161

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: concrete/abstract decision data for 10,000 English words. *Behavior research methods, 49*(2), 407-417. https://doi.org/10.3758/s13428-016-0720-6

Pezzulo, G., Barca, L., Bocconi, A. L., & Borghi, A. M. (2010). When affordances climb into your mind: Advantages of motor simulation in a memory task performed by novice and expert rock climbers. *Brain and Cognition, 73*(1), 68–73. https://doi.org/10.1016/j.bandc.2010.03.002

Phillips, L. H., Wynn, V. E., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory, 7*, 209–231. https://doi.org/10.1080/741944066

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia*, *48*(5), 1336–1342. https://doi.org/10.1016/j.neuropsychologia.2009.12.036

Postal, V. (2004). Expertise in cognitive psychology: Testing the hypothesis of long-term working memory in a study of soccer players. *Perceptual and Motor Skills, 99*(2), 403-420. https://doi.org/10.2466%2Fpms.99.2.403-420

Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience, 3*(1) 85–90. https://doi.org/10.1038/71156

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

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410-430. https://doi.org/10.1002/bs.3830120511

Quillian, M. R. (1969). The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM, 12*(8), 459–476. https://doi.org/10.1145/363196.363214

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rabin, M. D. (1988). Experience facilitates olfactory quality discrimination. *Perception & Psychophysics, 44*(6), 532–540.

Raman, I., & Baluch, B. (2001). Semantic effects as a function of reading skill in word naming of a transparent orthography. *Reading and Writing, 14*(7-8), 599-614.

Recchia, G., & Jones, M. N. (2012). The semantic richness of abstract concepts. *Frontiers in Human Neuroscience, 6*, 1–16. https://doi.org/10.3389/fnhum.2012.00315

Recchia, G., & Louwerse, M. M. (2015). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *The Quarterly Journal of Experimental Psychology, 68*(8), 1584–1598. https://doi.org/10.1080/17470218.2014.941296

Reilly, J., Peelle, J. E., Antonucci, S. M., & Grossman, M. (2011). Anomia as a marker of distinct semantic memory impairments in Alzheimer's disease and semantic dementia. *Neuropsychology*, *25*(4), 413-426. https://psycnet.apa.org/doi/10.1037/a0022738

Reilly, M., & Desai, R. H. (2017). Effects of semantic neighborhood density in abstract and concrete words. *Cognition*, *169*, 46–53. https://doi.org/10.1016/j.cognition.2017.08.004

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303-345. https://doi.org/10.1111/j.1756-8765.2010.01111.x

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

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rogers, T. T., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2004). Natural selection: The impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology*, *21*(2-4), 331-352. https://doi.org/10.1080/02643290342000366

Roome, H. E., Towse, J. N., & Crespo-Llado, M. M. (2019). Contextual support for children's recall within working memory. *Quarterly Journal of Experimental Psychology*, *72*(6), 1364-1378. https://doi.org/10.1177%2F1747021818804440

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronounceability for 925 nouns. *Memory & Cognition, 14*(1), 79–94. https://doi.org/10.3758/BF03209231

Rudner, M., Orfanidou, E., Cardin, V., Capek, C. M., Woll, B., & Rönnberg, J. (2016). Pre-existing semantic representation improves working memory performance in the visuospatial domain. *Memory & Cognition*, *44*(4), 608-620. https://doi.org/10.3758/s13421-016-0585-z

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* Vol. 1 (pp. 45-76). Cambridge, MA: MIT Press.

Sadoski, M. & Paivio, A. (2004). A dual coding theoretical model of reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1329-1362). Newark, DE: International Reading Association.

Sadoski, M., & Paivio, A. (2012). *Imagery and text : A dual coding theory of reading and writing.* Routledge.

Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition, 3*(1), 83-119. https://doi.org/10.1515/langcog.2011.004

Saults, J. S., & Cowan, N. (2009). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General, 136*(4), 663–684. https://doi.org/10.1037/0096-3445.136.4.663.A

Scerrati, E., Baroni, G., Borghi, A. M., Galatolo, R., Lugli, L., & Nicoletti, R. (2015). The modality-switch effect: Visually and aurally presented prime sentences activate our senses. *Frontiers in Psychology, 6*, 1–8. https://doi.org/10.3389/fpsyg.2015.01668

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*(2), 374–379. https://doi.org/10.3758/s13428-011-0162-0

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive psychology*, *22*(1), 36-71. https://doi.org/10.1016/0010-0285(90)90003-M

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods, 22*(2), 322-339. https://doi.org/10.1037/met0000061

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2018). The Glasgow norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods, 51*(3), 1258-1270. https://doi.org/10.3758/s13428-018-1099-3

Scribner, S. (1984). Studying working intelligence. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 9-40). Cambridge, MA: Harvard University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences 3*(3), 417-457.

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42*(2), 393–413. https://doi.org/10.3758/BRM.42.2.393

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex, 49*(1), 222–231. https://doi.org/10.1016/j.cortex.2011.10.005

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect. *Cognition, 106*(1), 433–443. https://doi.org/10.1016/j.cognition.2006.12.011

Sidhu, D. M., Kwan, R., Pexman, P. M., & Siakaluk, P. D. (2014). Effects of relative embodiment in lexical and semantic processing of verbs. *Acta Psychologica*, *149*, 32–39. https://doi.org/10.1016/j.actpsy.2014.02.009

Sidhu, D. M., & Pexman, P. M. (2016). Is moving more memorable than proving? Effects of embodiment and imagined enactment on verb memory. *Frontiers in Psychology, 7*, 1–13. https://doi.org/10.3389/fpsyg.2016.01010

Sidhu, D. M., Siakaluk, P. D., Pexman, P. M., Yap, M. J., & Muraki, E. (2018). Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior Research Methods, 51*(2), 453-466. https://doi.org/10.3758/s13428-018-1171-z

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used?. *Journal of Applied Psychology*, *72*(1), 146-148. https://psycnet.apa.org/doi/10.1037/0021-9010.72.1.146

Simon, H. A. (1979). Information processing models of cognition. *Annual Review of Psychology*, *30*(1), 363-396.

Simmons, W. K., & Barsalou, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology, 20*(36), 451–486. https://doi.org/10.1080/02643290342000032

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology*, *102*(13), 106-119. https://doi.org/10.1016/j.jphysparis.2008.03.014

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810. https://doi.org/10.1016/j.neuropsychologia.2007.05.002

Simonsen, H. G., Lind, M., Hansen, P., Holm, E., & Mevik, B. H. (2013). Imageability of Norwegian nouns, verbs and adjectives in a cross-linguistic perspective. *Clinical Linguistics and Phonetics, 27*(6-7), 435–446. https://doi.org/10.3109/02699206.2012.752527

Skipper, L. M., & Olson, I. R. (2014). Brain & Language Semantic memory: Distinct neural representations for abstractness and valence. *Brain and Language*, *130*, 1–10. https://doi.org/10.1016/j.bandl.2014.01.001

Snefjella, B., & Kuperman, V. (2016). It’s all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition, 156*, 135–146. https://doi.org/10.1016/j.cognition.2016.07.010

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*(2), 244-259. https://doi.org/10.3758/BF03196856

Song, X., Yi, F., Zhang, J., & Proctor, R. W. (2019). Left is “good”: Observed action affects the association between horizontal space and affective valence *Cognition, 193*, 104030. https://doi.org/10.1016/j.cognition.2019.104030.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition, 166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Speed, L. J., & Brysbaert, M. (2020). Dutch Sensory Modality Norms. Retrieved from PsyArXiv database. https://doi.org/10.31234/osf.io/zv6pn

Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: Simulating perceptual modality in space. *Behavior Research Methods, 49*(6), 2204–2218. https://doi.org/10.3758/s13428-017-0852-3

Speed, L. J., & Majid, A. (2018). An exception to mental simulation: No evidence for embodied odor language. *Cognitive Science*, *42*(4), 1146-1178. https://doi.org/10.1111/cogs.12593

Speed, L. J., van Dam, W. O., Hirath, P., Vigliocco, G., & Desai, R. H. (2017). Impaired comprehension of speed verbs in Parkinson’s disease. *Journal of the International Neuropsychological Society, 23*(5), 412–420. https://doi.org/10.1017/S1355617717000248

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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. https://doi.org/10.3758/BF03207704

Steyvers, M. (2010). Combining feature norms and text data with topic models. *Acta Psychologica*, *133*(3), 234-243. https://doi.org/10.1016/j.actpsy.2009.10.010

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 796–802. https://doi.org/10.1037/0278-7393.26.3.796

Suggate, S., Lehmann, J., Stoeger, H., & Jansen, P. (2019). Cognition embodied: mental rotation is faster for objects that imply a greater body–object interaction. *Journal of Cognitive Psychology, 31*(8), 876-890. https://doi.org/10.1080/20445911.2019.1678627

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage, 84*, 265–278. https://doi.org/10.1016/j.neuroimage.2013.08.023

Taylor, R. (2017). *Concreteness, frequency, and bilingual language dominance: Implications for the impact of context availability in explicit memory* (Open Access Doctoral Thesis). University of Texas at El Paso, USA. https://scholarworks.utep.edu/open_etd/565

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience, 17*(2), 273–281. https://doi.org/10.1162/0898929053124965

Towse, J. N. & Houston-Price C. M. T. (2001a). Reflections on the concept of the central executive. In J. Andrade (Ed.), *Working Memory in Perspective* (pp. 261-278). Psychology Press

Towse, J. N., & Houston-Price, C. M. (2001b). Combining representations in working memory: A brief report. *British Journal of Developmental Psychology*, *19*(3), 319-324.

Towse, J. N., & Saxton, M. (1997). Linguistic influences on children's number concepts: Methodological and theoretical considerations. *Journal of Experimental Child Psychology*, *66*(3), 362-375. https://doi.org/10.1006/jecp.1997.2389

Trumpp, N. M., Kliese, D., Hoenig, K., Haarmeier, T., & Kiefer, M. (2012). Losing the sound of concepts: Damage to auditory association cortex impairs the processing of sound-related concepts. *Cortex*, *49*(2), 474–486. https://doi.org/10.1016/j.cortex.2012.02.002

Tse, C. S., & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. *Memory*, *15*(6), 675–690. https://doi.org/10.1080/09658210701467186

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 830–846. https://doi.org/10.1037/0096-1523.24.3.830

Tulving, E., & Patkau, J. E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *16*(2), 83–95. https://doi.org/10.1037/h0083231

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological review*, *114*(1), 104. https://psycnet.apa.org/doi/10.1037/0033-295X.114.1.104

Unsworth N., & Robison M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin Review, 22*(3), 757-65. doi: 10.3758/s13423-014-0747-6. PMID: 25324180.

Van Dantzig, S., Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: Norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods, 43*(1), 145-154. https://doi.org/10.3758/s13428-010-0038-8

Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*(2), 241-255. https://doi.org/10.1016/j.cogbrainres.2004.09.002

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vannuscorps, G., & Caramazza, A. (2016). Impaired short-term memory for hand postures in individuals born without hands. *Cortex, 83*, 136–138. https://doi.org/10.1016/j.cortex.2016.07.019

Van't Wout, F., & Jarrold, C. (2020). The role of language in novel task learning. *Cognition, 194*(104036). https://doi.org/10.1016/j.cognition.2019.104036

Vergallito, A., Petilli, M. A., & Marelli, M. (2020). Perceptual modality norms for 1,121 Italian words: A comparison with concreteness and imageability scores and an analysis of their impact in word processing tasks. *Behavior Research Methods, 52*(4), 1599–1616. https://doi.org/10.3758/s13428-019-01337-8

Vermeulen, N., Chang, B., Mermillod, M., Pleyers, G., & Corneille, O. (2013). Memory for words representing modal concepts. *Experimental Psychology, 60*(4), 293-301. https://doi.org/10.1027/1618-3169/a000199

Vermeulen, N., Corneille, O., & Niedenthal, P. M. (2008). Sensory load incurs conceptual processing costs. *Cognition*, *109*(2), 287–294. https://doi.org/10.1016/j.cognition.2008.09.004

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods, 7*(1), 55–79. https://doi.org/10.1002/jrsm.1164

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://www.jstatsoft.org/v36/i03/.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*(2), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2013). The neural representation of abstract words: The role of emotion. Cerebral Cortex, 24(7), 1767–1777

Villani, C., Lugli, L., Liuzza, M. T., & Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition, 11*(3), 403–430. https://doi.org/10.1017/langcog.2019.23

Visser, M., & Lambon Ralph, M. A. (2011). Differential contributions of the ventral anterior temporal lobe and the anterior superior temporal gyrus to semantic processes, *Journal of Cognitive Neuroscience, 23*(10), 3121-3131. https://doi.org/10.1162/jocn_a_00007

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92–114. https://doi.org/10.1037/0096-1523.27.1.92

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: MIT Press. (Original work published 1934).

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review, 14*(5), 779–804. https://doi.org/10.3758/BF03194105

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning Memory and Cognition, 25*(5), 1256–1271. https://doi.org/10.1037/0278-7393.25.5.1256

Walker, P., Hitch, G. J., Dewhurst, S. A., Whiteley, H. E., & Brandimonte, M. A. (1997). The representation of nonstructural information in visual memory: Evidence from image combination. *Memory & cognition*, *25*(4), 484-491. https://doi.org/10.3758/BF03201124

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Westbury, C., & Moroschan, G. (2009). Imageability x phonology interactions during lexical access: Effects of modality, phonological neighbourhood, and phonological processing efficiency. *The Mental Lexicon, 4*(1), 115–145. https://doi.org/10.1075/ml.4.1.05wes

Westbury, C. F., Shaoul, C., Hollis, G., Smithson, L., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2013). Now you see it, now you don't: On emotion, context, and the algorithmic prediction of human imageability judgments. *Frontiers in Psychology, 4,* 1–13. https://doi.org/10.3389/fpsyg.2013.00991

Whorf, B. L. (1956). The relation of habitual thought and behavior to language. In J. B. Carroll (Ed.), *Language, thought, and reality*. *Selected writings of Benjamin Lee Whorf*. Cambridge, MA: The MIT Press.

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Milton Bache, S., Muller, K., Ooms, J., Robinson, D., Paige Seidel, D., Spinu, V. ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive science, 29*(5), 719-736. https://doi.org/10.1207/s15516709cog0000_33

Willems, R. M., Hagoort, P., & Casasanto, D. (2010). Body-specific representations of action verbs: Neural evidence from right-and left-handers. *Psychological Science*, *21*(1), 67-74. https://doi.org/10.1177%2F0956797609354072

Willits, J. A., Amato, M. S., & MacDonald, M. C. (2015). Language knowledge and event knowledge in language use. *Cognitive Psychology, 78*, 1–27. https://doi.org/10.1016/j.cogpsych.2015.02.002

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6-10.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin and Review, 8*(1), 44–57. https://doi.org/10.3758/BF03196138

Wilson, M. (2002). Six Views of Embodied Cognition. *Psychonomic Bulletin & Review 9,* 625–636. https://doi.org/10.3758/BF03196322

Wilson, M. A., Cuetos, F., Davies, R., & Burani, C. (2013). Revisiting age-of-acquisition effects in Spanish visual word recognition: The role of item imageability. *Journal of Experimental Psychology: Learning Memory and Cognition, 39*(6), 1842–1859. https://doi.org/10.1037/a0033090

Wilson, M., & Emmorey, K. (1998). A "word length effect" for sign language: Further evidence for the role of language in structuring working memory. *Memory & Cognition*, *26*(3), 584-590. https://doi.org/10.3758/BF03201164

Wingfield, C., & Connell, L. (2019). Understanding the role of linguistic distributional knowledge in cognition. Retrieved from PsyArXiv database. https://doi.org/10.31234/osf.io/hpm4z

Witt, J. K., Kemmerer, D., Linkenauger, S. A., & Culham, J. (2010). A functional role for motor simulation in identifying tools. *Psychological Science*, *21*(9), 1215–1219. https://doi.org/10.1177/0956797610378307

Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning Memory and Cognition, 31*(5), 878–890. https://doi.org/10.1037/0278-7393.31.5.878

Xiao, X., Zhao, D., Zhang, Q., & Guo, CY. (2012). Retrieval of concrete words involves more contextual information than abstract words: Multiple components for the concreteness effect. *Brain and Language, 120*(3), 251–258. https://doi.org/10.1016/j.bandl.2011.09.006

Yap, M. J., Lim, G. Y., & Pexman, P. M. (2015). Semantic richness effects in lexical decision: The role of feedback. *Memory and Cognition, 43*(8), 1148–1167. https://doi.org/10.3758/s13421-015-0536-0

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience, 6*, 1–10. https://doi.org/10.3389/fnhum.2012.00072

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin and Review, 18*(4), 742–750. https://doi.org/10.3758/s13423-011-0092-y

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning Memory and Cognition, 31*(6), 1385–1397. https://doi.org/10.1037/0278-7393.31.6.1385

Zdrazilova, L., & Pexman, P. M. (2013). Grasping the invisible: Semantic processing of abstract words. *Psychonomic Bulletin and Review, 20*(6), 1312–1318. https://doi.org/10.3758/s13423-013-0452-x

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*(1), 1-29.

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2018). The production effect and the generation effect improve memory in picture naming. *Memory, 27*(3), 1–13. https://doi.org/10.1080/09658211.2018.1510966

Zucco, G. M. (2003). Anomalies in cognition: Olfactory memory. *European Psychologist, 8*(2), 77–86. https://doi.org/10.1027//1016-9040.8.2.77

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, *10*(4), 954-958. https://doi.org/10.3758/BF03196557