The Role of Sensorimotor and Linguistic Distributional Information in Categorisation

Rens van Hoef

MSc

This thesis is submitted in partial fulfilment of the requirements for the degree

of Doctor of Philosophy

July 2021

Department of Psychology

Lancaster University

Table of contents

TABLE OF CONTENTS **3**

DECLARATION **6**

ACKNOWLEDGEMENTS **7**

THESIS ABSTRACT **8**

STATEMENT OF AUTHORSHIP **10**

CHAPTER 1: A REVIEW OF THE CATEGORISATION LITERATURE **12**

1.1. Key effects in object categorisation: effects of taxonomic level and typicality 15

1.2. Traditional accounts of categorisation 19

1.3. Alternative views of conceptual representation 53

1.4. A sensorimotor-linguistic account of taxonomic and typicality effects/processes 64

1.5. Thesis roadmap 67

1.6. References 69

CHAPTER 2: THE EFFECTS OF SENSORIMOTOR AND LINGUISTIC INFORMATION ON THE BASIC-LEVEL ADVANTAGE **85**

2.1. Chapter introduction 85

2.2. Abstract 87

2.3. Introduction 88

2.4. Experiment 1: The Basic-Level Advantage 94

2.5. Experiment 2: Typicality and the Basic-Level advantage 112

2.6. Experiment 3a: Replication of Experiment 1 127

2.7. Experiment 3b: Exploratory Replication of Experiment 2. 135

2.8. Experiment 4a: Combined Analysis for Experiments 1 and 3a. 142

2.9. Experiment 4b: Combined Analysis for Experiments 2 and 3b 148

2.10. General discussion 156

2.11 References 162

CHAPTER 3: HOW SENSORIMOTOR AND LINGUISTIC DISTRIBUTIONAL INFORMATION AFFECT PERFORMANCE IN THE CATEGORISATION OF CONCEPT LABELS **173**

3.1. Chapter introduction 173

3.2. Abstract 176

3.3. Introduction 177

3.4. Method 184

3.5. Results 192

3.6. General discussion 210

3.7. References 214

CHAPTER 4: TIMED PICTURE NAMING NORMS FOR 800 PHOTOGRAPHS OF 200 OBJECTS IN ENGLISH 220

4.1. Chapter introduction 220

4.2. Abstract 223

4.3. Introduction 224

4.4. Norming Method 228

4.5. Comparison with previous norms 238

4.6. Recognition latencies 240

4.7. Conclusions 243

4.8. References 245

CHAPTER 5: THE ROLE OF SENSORIMOTOR AND LINGUISTIC DISTRIBUTIONAL INFORMATION IN ULTRA-RAPID CATEGORISATION 248

5.1. Chapter introduction 248

5.2. Abstract 251

5.3. Introduction 252

5.4. Method 258

5.5. Results 269

5.6. General discussion 292

5.7. References 298

CHAPTER 6: GENERAL DISCUSSION 304

6.1. Thesis aims 304

6.2. Contributions 308

6.3. Limitations and future directions 316

6.4. Concluding remarks 323

CONSOLIDATED BIBLIOGRAPHY 323

LIST OF TABLES 353

Chapter 1 353

Chapter 2 353

Chapter 3 354

Chapter 4 355

Chapter 5 355

LIST OF FIGURES 358

Chapter 1 358

Chapter 2 358

Chapter 3 358

Chapter 4 359

Chapter 5 359

Declaration

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

This research has received funding from the European Research Council (ERC), under the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 682848.

Name: Rens van Hoef

Signature: _______________

Date: 26-07-2021

Acknowledgements

Many thanks to both my supervisors, Louise Connell and Dermot Lynott, for giving me the change to be part of a very interesting project, and of course for your guidance, patience and encouragement. Your attention to detail, level of knowledge and genuine passion for science are inspiring.

To my family and friends back home, dank jullie wel. It is good to know that you are always in my corner. To Rachael W. Cheung, thank you for the laughs, the talks, your endless support, and for making every day worthwhile.

Finally, to my friends in C8, all members of the Embodied Cognition Lab, and all the other lovely people in the Lancaster University psychology department I met along the way: thank you for making this journey so much more enjoyable.

Thesis abstract

How do people know what categories objects belong to? Traditional accounts of categorisation typically assume that concepts comprise perceptual or functional features. By contrast, recent accounts of conceptual structure emphasise the dual role of sensorimotor (i.e., perception-action experience of the world) and linguistic distributional information (i.e., statistical distribution of words in language).

This thesis contains a literature review and four empirical papers describing pre-registered experiments, which explore how the degree of sensorimotor-linguistic representational overlap between category- and member-concepts may drive object categorisation.

Chapter 1 presents a review of the literature, covering relevant theories of categorisation, as well as sensorimotor-linguistic theories of conceptual processing. Chapter 2 presents a study which explored the role of sensorimotor and linguistic distributional information in processing advantages. This study found that overlap in sensorimotor and linguistic distributional representations between category (e.g., *dog*) and member (e.g., *Labrador*) concepts reliably predicted performance (accuracy, RT) in a speeded picture category verification task. Chapter 3 reports a study that contrasted the traditional prediction of a basic-level advantage with the sensorimotor-linguistic prediction that representational overlap, not taxonomic level, is more important to categorisation. In a forced-choice categorisation task using labels only, participants decided between a basic- (e.g., *dog*) and superordinate-level label (e.g., *animal*) for a target object label (e.g., *Labrador*). While basic-level labels were overall chosen faster and more frequently, an exploratory analysis suggested that basic-level categorisation was slowed down when sensorimotor-linguistic overlap was greater between the target object label and the superordinate label. Chapter 4 describes the collection of a normed set of 800 photographs of 200 natural objects and artefacts, and their most frequent names. An exploratory analysis of the object recognition latencies associated with each photograph found that word frequency and length averaged over all names given to

an image predicted object recognition time better than the word frequency and length of the most frequent response. Chapter 5 reports a study that used the images and names collected in the study reported in Chapter 4, and examined the role of sensorimotor and linguistic distributional information in an ultra-rapid object categorisation paradigm with backwards masking. This study found evidence for the effect of linguistic distributional on sensorimotor information on categorisation accuracy, but not RT, nor was there a systematic relationship between perceptual processing time and sensorimotor-linguistic information.

In summary, the findings presented in this thesis provide support for a novel account of object categorisation based on sensorimotor and linguistic distributional representational overlap between category and member concepts.

Statement of authorship

The following contains a breakdown of the contribution made by Rens van Hoef (the student) and Louise Connell and Dermot Lynott (the supervisors) to each chapter. The order in which names are listed broadly indicates the proportion of the contribution (descending), unless joint contribution is specified.

Chapter 1 (literature review)

- Principle author: Rens van Hoef
- Comments: Louise Connell and Dermot Lynott

Chapter 2: The Role of Sensorimotor and Linguistic Distributional Information in the Basic-level Advantage **–** *empirical paper in preparation*

- **Conception and creation of the study:** Louise Connell, Rens van Hoef and Dermot Lynott.
- **Data collection:** Rens van Hoef, supervised by Louise Connell
- **Writing manuscript:** Rens van Hoef
- **Revision:** Rens van Hoef and Louise Connell
- **Comments:** Dermot Lynott

Chapter 3: How Sensorimotor and Linguistic Distributional Information Affect Performance in the Categorisation of Concept Labels **–** *empirical paper in preparation*

- **Conception and creation of the study:** Rens van Hoef, Louise Connell (jointly)
- **Data collection:** Rens van Hoef, supervised by Louise Connell
- **Writing manuscript:** Rens van Hoef
- **Revision:** Rens van Hoef, Louise Connell
- **Comments:** Dermot Lynott

Chapter 4: Timed Picture Naming Norms for 800 Photographs of 200 Objects in English **–** *empirical paper in preparation*

- **Conception and creation of the study:** Rens van Hoef, Louise Connell (jointly)
- **Data collection:** Rens van Hoef, supervised by Louise Connell
- **Writing manuscript:** Rens van Hoef
- **Revision:** Rens van Hoef, Louise Connell
- **Comments:** Dermot Lynott

Chapter 5: The Role of Sensorimotor and Linguistic Distributional Information in Ultra-Rapid Categorisation **-** *empirical paper in preparation*

- **Conception and creation of the study:** Rens van Hoef, Louise Connell (jointly)
- **Data collection:** Rens van Hoef, supervised by Louise Connell
- **Writing manuscript:** Rens van Hoef
- **Revision:** Rens van Hoef, Louise Connell
- **Comments:** Dermot Lynott

Chapter 6 (General Discussion)

- **Principle author:** Rens van Hoef
- **Comments:** Louise Connell and Dermot Lynott

Supervisor's name: Louise Connell Signature: ________________

Supervisor's name: Dermot Lynott Signature: ________________

Date: 16.07.2021

Chapter 1: A Review of the Categorisation Literature

Categorisation is fundamental to human perception, cognition, and language (Lakoff, 1987). Without categories, every object or event would be unique to us (E. E. Smith & Medin, 1981), leaving us overwhelmed by their sheer number and diversity. Instead, our ability to categorise greatly reduces the complexity of the world we perceive (Rosch, 1978), by grouping together and labelling those objects that we consider to be similar (e.g., both of those furry, four-legged objects are *dogs*) or distinguishing between objects that we think are not (e.g., the object on the left is a *dog*, but the object on the right is a *sheep*). Object categorisation allows us to infer information about previously unseen objects (the *dog* is a *pet*, the *sheep* is *livestock*), as well as allow us to evaluate possible interactions with them (e.g., if I throw a ball, the *dog* might fetch it; the *sheep* will not) and communicate about them to others (e.g., 'look, it's a *dog*').

Theories of object categorisation aim to answer a number of key questions. For example, by what mechanism do we group together some objects but not others? How do we determine that something (e.g., a *Labrador*) is like something else (e.g., a *spaniel*)? How do we judge that *ducks* and *Labradors* are sufficiently alike to both be *animals*, but not sufficiently alike to both be *dogs*? How is information about groups of objects (e.g., *dogs*) and their relationships with other groups of objects (e.g., *birds*) stored in semantic memory? Is there a relationship between the way we perceive objects and the way we represent them as concepts? If so, what are the building blocks that underpin both object-object and object-concept similarity?

These questions illustrate the close relationship between explanations of object categorisation and assumptions about conceptual representation. To explain one (categorisation) you need an idea about the other (conceptual representation); to explain that both *Labradors* and *spaniels* are *dogs*, one would need to account for the way *dogs*, *Labradors* and *spaniels* are represented in semantic memory. Therefore, to discuss theories of categorisation is to discuss theories of conceptual representation and the organisation of semantic memory. However, recent theories of conceptual representation have not explicitly addressed categorisation behaviour.

Traditional accounts of categorisation assume that instances and concepts may be deconstructed into perceptual and functional features[1] (e.g., *has wings*, *can fly*, *lays eggs*, etc.), and that categorisation of a given instance (e.g., a furry four-legged creature that barks) involves a comparison between features that are observed in the instance (e.g., *has fur*, *has four legs*, *barks*) to those stored in semantic memory with a given concept (e.g., *dog* or *bird*). However, traditional accounts differ in the way they believe categories are represented in semantic memory, and consequently, by what mechanism a feature-comparison results in a category decision. The definitional or classic view suggests that categories are represented in semantic memory as a collection of features that are necessary and jointly sufficient to describe it. By contrast, prototype theory assumes that categories are represented by the central tendency of their features; a summary of features common to a category (Posner & Keele, 1968; Rosch, 1973; Rosch & Mervis, 1975). Alternatively, exemplar theory suggests that categories are represented as a set of previously encountered instances of a category. According to traditional accounts then, categorisation is the process of verifying that an instance possesses either all of a category's necessary and sufficient features (definitional view), possesses a sufficient degree of features common to the category (prototype theory), or matches the features of previously observed category exemplars (exemplar theory).

Feature-based theories make assumptions about the manner in which concepts are represented in our minds, which are, by definition, impossible to test directly. Therefore, a key focus of traditional categorisation research has been to explain human categorisation behaviour. Two key effects have drawn a significant portion of attention: taxonomic effects and (proto-) typicality effects (hereafter, described as 'typicality'). Taxonomic effects concern the apparent variation in preference people have for categorisation at various levels of

[1] For the purpose of this work, the definition of feature is that of E.E. Smith and Medin (1981), and encompasses sets of binary and discrete features as well as dimensions, which may be perceptual (e.g., *has wings*) and functional (e.g., *can fly*) in nature.

abstraction (e.g., categorisation of a particular furry, four-legged animal as *Labrador*, *dog* and *animal*), referred to as the vertical dimension of category structure (Rosch, 1978).The most prominent of these effects is the basic-level advantage (Rosch, Mervis, et al., 1976), which refers to an increased ease of categorisation for categories of an intermediate level of abstraction (e.g., *dog*, *fish*, *car*). Typicality effects meanwhile pertain to a varying ease of categorisation for different members of the same category (e.g., *robin*, and *ostrich* as members of the category *bird*), depending on their degree of exemplariness (e.g., *robin* is judged as a better example of a *bird* than an *ostrich*, which affects categorisation speed and accuracy; Rips et al., 1973). Typicality effects have been argued to reflect the graded structure categories may have, and are referred to as the horizontal dimension of categories (Rosch, 1978).

However, in contrast to feature-based theories, more recent accounts on conceptual processing argue that concepts do not consist of discrete and binary features, but rather comprise a combination of sensorimotor information (i.e., perception-action and affective experience of the world) and linguistic distributional information (i.e., information about the statistical distribution of words in language; (Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse, 2011). Crucially, although some influential accounts (e.g., Rosch, 1978) argued against the notion that categorisation effects reflect conceptual structure, all theories of categorisation make assumptions about underlying conceptual structure in some shape or form. The inverse cannot be said for recent theories of conceptual processing. That is, while these theories make explicit assumptions about the organisation and nature of conceptual representations in semantic memory, they have largely not yet addressed what these assumptions mean for long-standing accounts of behavioural effects in object categorisation, such as taxonomic effects (e.g., basic-level advantage) and object typicality.

The aim of this thesis is therefore to critically evaluate the core tenets of traditional feature-based accounts of object categorisation, but also to propose and test a theory of object categorisation that is in line with modern sensorimotor-linguistic accounts of conceptual processing. In particular, the aim is to test whether categorisation behaviour – specifically, observable effects of taxonomic level and typicality – can be explained by measuring the degree of overlap in sensorimotor experience and linguistic distributional knowledge that is activated by a given instance in visual or verbal format (e.g., a photograph of a *Labrador* or the label *Labrador*) and a given category name (e.g., *dog*, *animal* or *Labrador*) across a range of categorisation tasks. The studies presented in this thesis are, to my knowledge, the first to explore this hypothesis, and make use of state-of-the-art measures of sensorimotor (e.g., Lynott et al., 2020) and linguistic distributional information (Wingfield & Connell, 2019).

1.1. Key effects in object categorisation: effects of taxonomic level and typicality

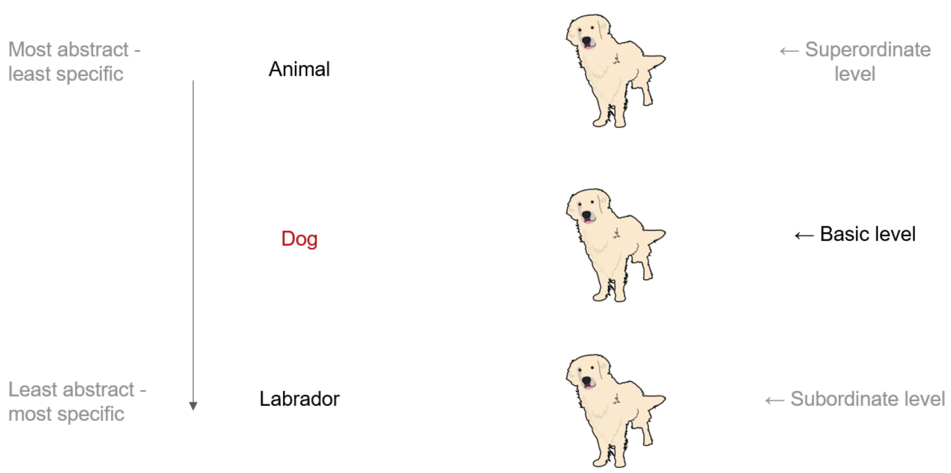
The mechanisms that are assumed to underly object categorisation and the representation of categories in semantic memory are hard or even impossible to observe. As a consequence, hypotheses about their nature must be tested by observing human categorisation *behaviour*. Of the various behaviours that research has associated with feature-based object categorisation, none have received more attention than taxonomic (e.g., Corter & Gluck, 1992; Jolicoeur et al., 1984; Markman & Wisniewski, 1997; Mervis & Crisafi, 1982; Murphy & Brownell, 1985; Murphy & Smith, 1982; Rogers & Patterson, 2007; Rosch et al., 1976; Tanaka & Taylor, 1991) and typicality effects (Armstrong et al., 1983; Hampton, 1998; Murphy & Ross, 2005; Rips et al., 1973; Rosch, Simpson, et al., 1976; Rosch & Mervis, 1975). Because of this, these effects form a robust basis to contrast feature-based explanations of categorisation with a sensorimotor-linguistic account.

The taxonomic effect follows from the notion that any given object may be categorised at multiple levels of abstraction (cf. Rosch, 1978). For example, a furry, four-legged object may simultaneously be a member of the categories *Labrador*, *dog*, *animal*, and *living thing.* These categories may be organised taxonomically: placed vertically along an axis of abstraction and inclusiveness (see figure 1), whereby members of each category are included in all categories that are more abstract (e.g., *Labradors* are *dogs* which are *animals*).

Crucially, while categorisation at any of these levels of abstraction is perfectly possible, researchers have observed that people generally prefer to categorise objects at an intermediate level of abstraction referred to as the basic level (e.g., *dog*) compared to more specific (i.e., subordinate) or abstract levels (i.e., superordinate). In an extensive series of experiments, Rosch et al. (1976) demonstrated that categorisation at the basic-level yields faster and more accurate responses in category-verification tasks, but also that basic-level categories are more frequently named in category production tasks, and that members of basic-level categories have the most similar shapes. Rosch et al. (1976) also argued that members of basic-level categories are learned and recognised earliest by children, although this particular finding has been contended (Mandler & Bauer, 1988; Mandler & McDonough, 2000).

Figure 1.

Example of the vertical dimension: categorisation of Labrador at three inclusive levels of abstraction.



Crucially, however, research has also uncovered a range of circumstances under which people seem to favour categorisation at taxonomic levels other than the basic level. For example, object typicality has been shown to interact with taxonomic level in predicting response times and accuracy in category-verification tasks (Jolicoeur et al., 1984; Murphy &

Brownell, 1985). That is, atypical items are categorised faster and more accurately at the specific subordinate level, whereas typical items display the basic-level advantage. Similarly, categoriser-expertise has been shown to reduce the advantage of the basic level over the subordinate level (K. E. Johnson & Mervis, 1997; Tanaka & Taylor, 1991). For example, Tanaka and Taylor (1991) found bird and dog experts are roughly equally as good at categorising objects from their area of expertise at the specific subordinate level (e.g., *robin*, *jay, cardinal*) as they are at categorising them at the basic level (e.g., *bird*). By contrast, research has found that a superordinate-level advantage occurs when people have limited time to perceive an object (Bacon-Macé et al., 2005; Mack et al., 2008; Mack & Palmeri, 2015). That is, people are faster and more accurate to decide that a rapidly flashed image contains an *animal* than they are to decide that it contains a *dog* or *bird*. Similarly, superordinate-level categorisation performance approaches basic-level performance when categorisation takes place in context (Murphy & Wisniewski, 1989). Finally, research into the effects of semantic impairment on categorisation show that superordinate-level categorisation is typically affected less than basic-level or subordinate-level categorisation (Rogers & Patterson, 2007). As a consequence, for the purpose of this thesis, the term 'taxonomic effects' is used to refer to a range of behavioural effects linked to classification at various levels of abstraction, of which the basic-level advantage is the most prominent example.

If taxonomic effects refer to the vertical dimension of categories, typicality effects pertain to a horizontal dimension of categories (Rosch, 1978). That is, where taxonomic effects concern performance differences for the categorisation of the same object at various levels of abstraction, prototypicality effects concern performance differences for the categorisation of different objects within the same category (Rips et al., 1973; Rosch, Simpson, et al., 1976) For example, while *robins*, *toucans*, *ostriches,* and *penguins* can all be categorised as *birds*, people may, when queried, judge some of these objects as more *typical* of that category (e.g., *robin*) than others (e.g., *ostrich*). Typicality is not equal to degree of membership. While emus are less typical birds, they are nonetheless considered to be birds (Hampton & Jönsson, 2012). As for the effects of taxonomic level, typicality effects are expressed in faster response times and higher accuracy in object categorisation tasks for objects with higher subjective typicality ratings (Armstrong et al., 1983; Rips et al., 1973; Rosch, 1973; Rosch, Simpson, et al., 1976; Rosch & Mervis, 1975; E. E. Smith et al., 1974) (Armstrong et al., 1983; Rips et al., 1973; Rosch, 1973; Rosch et al., 1976b; Rosch & Mervis, 1975; E.E. Smith et al., 1974). For example, Rips et al. (1973) found that the time people took to verify category membership depended on the rated semantic distance between a given object and the category.

Figure 2.

Object typicality in prototype theory, showing differences in exemplariness for typical and atypical members of the category bird. Atypical members share fewer features with the category as a whole, represented by the prototype.

Taxonomic and typicality effects are well-documented behavioural effects in object categorisation, that have proved to be a valuable testbed for broader theories of categorisation and conceptual representation. However, the most influential explanations of these effects are rooted in feature- and/or network-based accounts of conceptual representation and semantic structure, which will be outlined below. However, views on conceptual representation have changed considerably, and therefore existing explanations of categorisation and its most prominent behaviours (e.g., taxonomic and typicality effects) may require re-evaluation.

1.2. Traditional accounts of categorisation

Traditional explanations of key behavioural effects in object categorisation are closely tied to accounts of categorisation and conceptual representation. These accounts include the classical or definitional view, network accounts (e.g., Collins & Loftus, 1975; Collins & Quillian, 1969) and feature-similarity accounts such as prototype (Posner & Keele, 1968; Rosch & Mervis, 1975) exemplar theory (Brooks, 1978; Medin & Schaffer, 1978; Nosofsky, 1986; Nosofsky & Palmeri, 1997) and theory-theory (Medin, 1989; Murphy & Medin, 1985; Murphy & Spalding, 1995; Wisniewski & Medin, 1994) which are briefly reviewed here. These theories are united by the assumption that objects and representative concepts comprise features, which serve as the basis for computation and comparison of the similarity between objects and concepts. A short overview of each theory, followed by an examination of the extent to which it can explain taxonomic and typicality effects, is provided below.

1.2.1. The definitional view

Thinking about categorisation and concepts can be viewed as an iterative process, which starts from a straightforward idea and gradually grows more complex as attempts are made to account for a wider range of human behaviour. This iterative process begins with the classical or definitional view of categories and their representation in semantic memory (see Komatsu, 1992; Laurence & Margolis, 1999; Murphy, 2002; E.E. Smith & Medin, 1981; for reviews). The definitional view is not an explicit theory, but rather refers to the scientific and philosophical consensus at the time alternative views such as prototype and exemplar theory were developed.

Before addressing the definitional view’s perspectives on taxonomic and typicality effects (or lack thereof), it is important to outline its key assumptions. E.E. Smith and Medin (1981) describe these as follows:

1. Concepts are represented through an abstracted feature summary, which does not need to correspond to a specific instance. E.E. Smith and Medin argue that one of

the key functions of a concept is to summarise knowledge about the features that it is typically associated with (e.g., the concept *dog* is associated with the features *has four legs, has a tail, barks*), which is then used to categorise previously unseen entities (e.g., if has four legs, has a tail and barks, then it is a *dog*), or conversely to infer information about an entity which we cannot perceive (e.g., if it is a *dog* then it must *have four legs*, *have a tail* and *bark*).

2. The features that represent a concept are ‘singly necessary and jointly sufficient’ to define a concept.
3. In subset relations between concepts (e.g., *dog* may be regarded as a subset of *animal*) defining features of the more abstract concept are nested in those of the more specific concept (e.g., the features of *animal* are nested in those of *dog*).

Rey (1983) adds an implicit fourth assumption: having a concept means knowing its defining features, while Murphy (2002) also notes that the definitional view does not differentiate between members of the same category.

The definitional view held strong appeal as a theory of categorisation and concept acquisition and representation. For example, on this account, concepts are acquired by storing features which have been observed to correlate in the environment (Laurence & Margolis, 1999) - a notion that was in fact carried forward to later influential feature-based accounts (e.g., Rosch, 1978). Categorisation then is thought to be a matter of verifying that an object satisfies the features stored under the concept; category members are grouped together because they share a particular set of features with one another, but not with objects outside of the category. As such, in the definitional view, concept membership is clearly defined and discrete (Komatsu, 1992). This means that only if a concept possesses all a category’s necessary and sufficient features it is a member, otherwise it is not (e.g., if an object *has four legs, has a tail* but does not *bark,* then it cannot be a *dog*). These clear definitions for a concept, that fall apart into separate, discrete features, are extremely valuable for logical reasoning about concepts. For example, as Armstrong et al. (1983) illustrate, it explains compositional meaning, or how the meaning of a phrase and a word can be synonymous (e.g., “A barking, four-legged creature with a wagging tail” and *dog*), as the linguistic components of the sentence map directly onto the feature list stored in semantic memory under the concept *dog*. Furthermore, the definitional view offers a potential explanation for complex concepts, whereby the meaning of a complex concept (e.g., *guard dog*) is derived from the definitions of the simpler sub concepts (e.g., *guard* and *dog*). While the definitional view is compelling in its simplicity, it has several issues that make it difficult to accept as a plausible model of conceptual processing and categorisation.

1.2.1.1 Can the definitional view explain taxonomic effects?

The definitional view does not have an explicit explanation for taxonomic effects. This is partially for historical reasons; the basic-level advantage was first extensively documented in the mid-seventies (e.g., Rosch, Mervis, et al., 1976) by which point alternative theories of object categorisation were being developed (e.g., Posner & Keele, 1968; Rosch, 1973). When looking at the core principles of the definitional view as outlined in E.E. Smith and Medin (1981), one might conclude that the definitional view does not predict a basic-level advantage at all, although it might predict an advantage of near-superordinates over far superordinates (Medin & Smith, 1984) . That is, on the definitional view, a given subset category (e.g., *Labrador*) comprises the defining features of *Labrador* plus the defining features of its near superordinate (e.g., *dog*) as well as those of superordinates that are further removed (e.g., *mammal*, *animal*). As a result, *Labrador* shares more of its features with its nearer superordinates (e.g., *dog*) than it does with its far superordinate (e.g., *animal*; see table 1). From this follows the prediction that individuals should be better at judging whether a *Labrador* is a *dog* than whether a *Labrador* is an *animal*. In this example, that would result in some form of processing advantage of the basic *dog* over the superordinate *mammal* and *animal*. However, following this logic means it is not a particular taxonomic level that is at an advantage, i.e., that there is always an advantage of the nearer over the further category. In other words, people should also be faster to judging that a *dog* is a *mammal* than judging it is an *animal*. However, research has shown that people are not consistently better at judging membership of near compared to far superordinates. For example, Roth and Mervis (1983) showed that participant ratings of goodness of example for various animal exemplars were not consistently higher for their near superordinates than for their far superordinates (e.g., *chicken* was rated as a worse example of *bird* than of *animal*). In summary, no explicit definitional account of the basic-level advantage exists. Furthermore, the taxonomic effects that a definitional view might predict are not mirrored in the behaviour of subjects in goodness-of-example tasks (Roth & Mervis, 1983; E.E. Smith et al., 1974) .

Table 1. Illustration of defining features shared between near and far superordinates according to a definitional account.

Concept	Defining features	Features in common with near superordinate	Features in common with far superordinate
Animal	A, B	-	-
Mammal	A, B, C, D	2	-
Dog	A, B, C, D, E, F	4	2 (animal)
Labrador	A, B, C, D, E, F, G, H	6	2 (animal) 4 (mammal)

1.2.1.2 Can the definitional view explain effects of typicality?

The definitional view offers no explanation for the graded structure of categories (Hampton, 1993; Komatsu, 1992) as expressed by typicality effects. An implicit assumption of the definitional view is that because membership is discrete, identified members are all equal within their category. However, this assumption is contrasted by the fact that when queried, participants are readily prepared to rate category members as better (typical) or worse examples of their respective category. As described above, previous research regarding object typicality (Rips et al., 1973; Rosch, 1973; Rosch & Mervis, 1975) revealed distinct asymmetries in the representativeness of category members of their categories (e.g., *ostriches* and *penguins* are judged to be less representative of *bird* than *robins* or *sparrows*). Vitally, typicality ratings predict object categorisation speed and accuracy, with a processing advantage for typical over atypical items (Jolicoeur et al., 1984; Murphy & Brownell, 1985; Rosch, Simpson, et al., 1976). While Rosch (1978) stresses that such prototype effects do not necessarily reflect category structure, they have been taken to reflect a fundamental weakness in the definitional view, because it cannot readily explain performance differences within a category.

Moreover, the finding that categories have a graded structures has led some researchers to question what that would mean for its boundaries (McCloskey & Glucksberg, 1978). In the definitional view, a category's boundaries are discrete. Objects either do, or do not belong to the category. However, if the structure of a category can be expressed as a typicality spectrum, it is possible that category members on the lower end of that spectrum are less strongly linked to the category, for example because they share features with multiple categories (e.g., certain *amphibian cars* straddle the line between *car* and *boat*). Indeed, research has shown that judgments of category-membership differ not only between participants, but also within the same participant when asked to categorise an object at different points in time (McCloskey & Glucksberg, 1978), suggesting that the boundaries between categories may be *fuzzy* rather than clearly defined, as the definitional view would argue.

1.2.1.3 Other limitations of the definitional view

As Wittgenstein (1958) highlighted, for many instances of a concept, it can be hard or even impossible to produce a set of defining features for a given category. For instance, both *chess* and *Pacman* are members of the category *games*, but it is not clear what set of features sufficiently describes both these instances. This is supported by research which has shown that people do not consistently generate a set of defining and sufficient features for every concept (Hampton, 1979, 1981). Crucially, this not only conflicts with Rey's (1983) assumption that to have a concept means to know its defining features, but also begs the question why, if all concepts are defined by a set of necessary and sufficient features, do people fail to reliably produce such a set? By contrast, E.E. Smith and Medin (1981) note that the notion that a set of necessary and sufficient features cannot be determined for every concept does not mean such a set does not exist, simply that we have not yet determined it, or are looking for the wrong kind of features (e.g., perceptual rather than functional features). Furthermore, during tasks that require feature listing, participants may suppress features that could be defining, but are redundant (e.g., *living thing*, *physical object*, or *cross between two dogs* for *dog*), in favour of naming those they believe to be more characteristic (e.g., *barks*; Armstrong et al., 1983). As such, feature listing tasks may not reflect the full scope of features associated with a concept, falsely giving the impression that defining features do not exist.

In summary, the definitional view offers straightforward explanations for concept acquisition (i.e., concepts are constructed by observing features that co-occur naturally), representation (i.e., concepts are represented by a set of features that are necessary and jointly sufficient to define it), and categorisation (i.e., categorisation is the process of verifying that an object possesses all necessary and sufficient features defined within a concept). Nevertheless, the definitional view is unable to account for certain aspects of human behaviour – crucially, there is inconsistency in goodness of example at various taxonomic levels) and no explicit explanation for taxonomic processing advantages (i.e., the basic-level advantage) nor for typicality effects, or the lack of consistency in categorisation between people and at different time points. The discovery of these effects contradicts the assumption that clearly defined categories exist, which is further supported by the difficulty in determining necessary and sufficient features. As a result, the definitional view is too constrained to serve as an adequate theory of concepts and categories.

The failure of the definitional view to account for typicality (e.g., Rips, Shoben & Smith, 1973; Rosch & Mervis, 1975) and taxonomic effects ((Jolicoeur et al., 1984; Murphy & Brownell, 1985; Rosch, Mervis, et al., 1976) led researchers to develop alternatives that subsequently formed the main tenets of traditional accounts of object categorisation such as prototype, exemplar, and related theories such as theory-theory. However, much of the core principles of the definitional view have been carried forward to these subsequent theories.

1.2.2. Associative network models

The definitional view supposes that semantic memory is like a dictionary, where each concept is paired with a feature-definition. However, in addition to the issues outlined above, it is not immediately clear how such a system could support the enormous amount of variation in concepts in a conceptual system with limited resources. For example, while the definitional view allows for nested categories (i.e., in subset relations, features of the more abstract concept are nested in that of the more concrete concept), there are quite a few exceptions to this (e.g., the feature <u>can fly</u> could be associated with the concept *bird*, and should therefore generalise to all of its subset categories – but what about *penguins* and *ostriches*?). Moreover, what about concepts that are not simple noun concepts (e.g., *the birds in my garden*, *things I would take from my house in case of a fire*)? How are all these concepts stored?

One possible answer to this question comes in the form of spreading activation theory (Collins & Loftus, 1975; Collins & Quillian, 1969), which argues concepts and their properties may be represented in semantic memory as a network of nodes, with links (edges) of varying strength depending on the frequency with which they are accessed. Network models testing spreading activation theory aimed to account for differences in processing speed, such as the finding that people take longer to verify statements such as *robin* is an *animal* than statements such as *robin* is a *bird*. Crucially, the concept of cognitive economy is introduced, whereby properties of related concepts are generally (but not always; Collins & Loftus, 1975) stored at the highest possible level of abstraction, and may be generalised to subordinate categories. Importantly, subordinate concepts may modify inherited properties, if they are linked to a modifying property (e.g., *bird* may be linked to the properties <u>has wings</u> and <u>can fly</u>, which are inherited by *robin*, adding *red breast* and *sings*, but modified by *ostrich*, which is associated with a modifying property: *cannot fly*). That is, in learning category concepts with subset relations (e.g., learning that *robin* is a *bird*) properties of *bird* are generally not stored again with *robin*, except if they are highly indicative of a concept (e.g., according to Collins & Quillian, <u>*has leaves*</u> may be stored with *tree*, but some aspects of <u>*has leaves*</u> may also be stored at the subordinate *maple tree*, as its leaves are characteristic of this tree).

In network models, similarity between concepts is determined by the property links they share. Activation of a given concept (e.g., *vehicle*) spreads to other concepts via connected nodes. For example, activation of *vehicle* results in the activation of *taxi* and *fire engine* via property links (e.g., <u>*has wheels*</u>, <u>*has combustion engine*</u>, <u>*can be driven*</u>). The strength of their association depends on the number of links they have in common, as well as on what Collins and Loftus (1975) name the *criteriality* of the link (i.e., how essential a linked property is to a given concept). In categorisation, the decision whether a concept matches another concept involves a search of semantic memory, checking whether there exist connections either between the concepts themselves (i.e., superordinate ISA connections such as *robin-bird*) or between the features they possess, where shared features with higher criteriality for one of the concepts constitutes stronger evidence for their similarity (e.g., <u>*can fly*</u> could be considered highly essential for *birds*, if a concept possess the feature <u>*can fly*</u>, this constitutes strong evidence that it is a *bird*). Evidence may also be negative (i.e., *bat* is not a *bird*). Because of the way they link concepts to a specific set of properties, network models could be considered to be definitional. However, in also allowing direct links between concepts (i.e., superordinate ISA links, such as *Labrador* is a *dog*), network models allow categorisation to be possible even when defining properties are not known (e.g., *sponge* is linked to *animal* through a superordinate link, for example because we are taught that *sponges* are *animals*, even if it is hard to determine what properties of *sponge* make it an *animal*). As such, network-models may circumvent some of the issues with the definitional view.

1.2.2.1 Can network models explain taxonomic effects?

The best-known explicit interpretation of spreading activation theory and its role in taxonomic effects assumes that the time it takes to categorise an object (e.g., is a *Labrador* an *animal*) is proportional to the time it takes for activation to spread from the *Labrador* node to the *animal* node in semantic memory (Jolicoeur et al., 1984). On this account, basic-level concepts correspond to the most common entry-points into a taxonomically structured semantic memory. The delays in categorising objects at the superordinate and subordinate compared to the basic level stem from the time it takes activation to spread through the network from the initial basic-level node. The interaction between typicality and taxonomic level (i.e., subordinate-level advantage for atypical items) is subsequently explained as an entry-level shift, whereby for atypical items, the network is entered at the subordinate level (e.g., *Labrador*) as opposed to the standard basic-level (e.g., *dog*).

However, this account is seemingly contradicted by findings in patients with semantic dementia (Hodges & Patterson, 2007; Rogers & Patterson, 2007; Warrington, 1975). Semantic dementia patients exhibit gradual deterioration of semantic memory secondary to the destruction of anterior temporal lobes (Patterson et al., 2007). Rogers and Patterson (2007) showed that patients with semantic dementia show a reversal of the expected effect described previously. Although healthy individuals categorise fastest at the basic level as opposed to the superordinate, semantic dementia patients appear to have lost the ability to categorise objects at the specific subordinate and basic-levels, but have retained the ability to categorise at the abstract superordinate level. If the basic-level is the standard entry-point into a hierarchically organised network, and superordinate-level categorisation is impossible without the basic level, these findings cannot be explained. Moreover, the basic-level as entry-point only works if semantic memory is assumed to be organised hierarchically, and if semantic search only spreads in one direction (i.e., from the entry point upwards or downwards). However, even early network accounts have argued that in a typical categorisation paradigm (e.g., determining whether *Labrador* is a *dog*), activation runs in parallel from both the member and the category, with the only difference being that *Labrador* is activated slightly earlier than *dog* (Collins & Loftus, 1975), and that semantic memory is not rigidly hierarchical. As such, explanations based on network-models must ignore the way in which network-models themselves argued conceptual processing would occur.

1.2.2.2 Can network models explain typicality effects?

Object typicality effects (i.e., faster categorisation for objects that are rated as better examples of their respective categories) could be explained from network models by referring to the manner in which evidence accumulates in the model. That is, when determining whether a concept is similar to another concept, which is argued to be the basis for categorisation, the links between those concepts and their properties are investigated. For example, concepts may have superordinate links (e.g., *robin* is a *bird*) as well as property links (e.g., *robins* *have wings*). For concepts that have a superordinate link, distinguishing features may slow down the process of categorisation (Collins & Loftus, 1975). For example, when determining whether an ostrich is a bird, the evidence from the superordinate link between ostriches and birds is countered by the distinguishing feature *cannot fly*, which is atypical of birds, thus slowing down decision-making.

While network models have remained relevant to research into cognitive representations (see Kumar, 2021; Siew et al., 2019; for reviews), they have not been applied extensively to the categorisation behaviours discussed here, and their recent developments fall largely outside of the scope of this work. Crucially, network-models were a notable attempt at operationalising assumptions about categorisation that are broadly in line with the definitional view on categorisation, as well as a mechanistic account of some behavioural effects in categorisation. Moreover, the work by Collins and Quillian formed a springboard for later connectionist models of categorisation and semantic memory(e.g., McClelland & Rogers, 2003; Rogers & McClelland, 2004; Rumelhart & Norman, 1988), a number of which are relevant to the current work, and will be discussed in section 1.3.1.

1.2.2.3 Early responses to network and definitional accounts

A weakness of the definitional view is that it is hard to produce a set of necessary and defining features for every concept (cf. Wittgenstein, 1958). Network models circumvented this issue by allowing for superordinate links (e.g., learning that *Labradors* are *dogs*). However, where network models expressed the likeness of concepts as evidence summed over their various connections, a second influential group of theories chose a different measure: feature similarity. Like the definitional view, feature similarity accounts argue that concepts comprise values on a large set of binary features (e.g., <u>*has wings*</u>, <u>*can fly*</u>) and dimensions (e.g., *size*, *weight*). However, where feature-similarity accounts deviate is the idea that all of a concepts' features and dimensions must be jointly sufficient and necessary.

An early iteration of this view draws a clear distinction between *defining* and *necessary* features (E.E. Smith & Medin, 1974). On this account, categorisation is a multi-stage process, whereby first all features of a given instance (e.g., *robin*) and category (e.g., *bird*) are extracted and compared, to determine the number of features that are shared (e.g., <u>*has wings*</u>, <u>*can fly*)</u>, and the similarity of the values for each feature (e.g., comparable size). If the resulting similarity surpasses a predetermined threshold, an instance may be rejected or accepted as a member of the category (e.g., accepting the statement a *robin* is a *bird*). When similarity cannot be clearly determined in this first stage, a second comparison between the instance and category is required, this time focusing only on their defining features (e.g., the characteristic-defining model retains the core of the definitional view, with discrete category boundaries and a (partial) reliance on features that are jointly sufficient and necessary – defining – to determine category membership). However, as Hampton (1979) illustrates, it also inherits some of the definitional view's flaws, in that for some categories, defining features cannot be specified, and that the defining-characteristic distinction contrasts with McCloskey and Gluckberg's (1978) finding that category boundaries may not be discrete. Consequently, Hampton argues that features should not be treated as dichotomous (defining vs. characteristic), and that category membership might be defined as the degree to which an object shares features with the category of features an object, without making the possession of one or more features an absolute requirement for category membership. It is this very notion that also forms the basis of one of the most influential feature-similarity theories of object categorisation: prototype theory.

1.2.4. Prototype theory

Prototype theory was devised as an answer to the definitional view's apparent inability to explain categories with *graded structure*, whereby category membership was not an all or nothing process. Evidence for this graded structure came from behavioural studies showing typicality effects (Rips et al., 1973; Rosch, Simpson, et al., 1976) and mathematical work (Zadeh, 1996) suggesting categories may have *fuzzy* boundaries (i.e., have members that share features with more than one category in equal measure; McCloskey & Glucksberg, 1978). While different versions of prototype theory exist – making prototype theory a prototypical concept (Geeraerts, 2016) - a key aspect of any iteration of prototype theory is the notion that category members need not share all their characteristics. It is important to note that while prototype theory departs from the definitional view, it retains many of its key components, such as the idea that categorisation is the process of matching observed features to an internally stored feature summary, as well as the idea that categories may be related to one another taxonomically. In leaving definitions behind, however, prototype theory needs to account for what holds categories together. If objects within a category are allowed not to share *every* feature, then how *are* they related?

To allow for category members to be related but not defined by a single set of necessary and sufficient features, prototype theory adapts Wittgenstein's idea (1958) that concepts may be related to one another by means of family resemblance. This is the idea that feature characteristics are distributed unevenly among category members, and some members share more of the characteristics typical of the category than others. Contrasting the necessary and sufficient features of the definitional view, prototype theory requires only that a feature occurs more than once within a given category (Komatsu, 1992). Family resemblance, as a measure of category membership, may be defined either as an object's similarity to all other category members (e.g., how many features a *Labrador* shares with other *dogs*), or as an object's similarity to the central tendency or summary of the features that usually occur within the category (Hampton, 1979; Posner & Keele, 1968; Reed, 1972; Rosch & Mervis, 1975)

While these interpretations functionally yield the same outcome (Barsalou, 1985; Komatsu, 1992), the latter emphasises the relationship between a conceptual representation (e.g., the feature summary of the category *dog*) and observed object (e.g., *the Labrador*), and is the central assumption of prototype theory. As a result, according to prototype theory, categorisation is the process of determining an instance's similarity to the prototype, and of comparing the features observed in the instance to the prototype feature summary stored in semantic memory (e.g., *has wings*, *bipedal*, *has a beak*), until a sufficient degree of evidence has been gathered to either accept the instance as a member or to reject it as a non-member.

In prototype theory, similarity to the prototype is defined as the number of prototypical features (weighted by their importance; Tversky, 1977) an object possesses (Hampton, 1993; Rosch, Simpson, et al., 1976; Rosch & Mervis, 1975). The more prototypical features an object possesses, the more similar it is to the prototype and the more likely it is that it is a member of the category. Crucially, this begs the question of what determines feature importance. On one influential account (Rosch & Mervis, 1975), feature importance may be derived from the proportion of category members that possess a given feature (i.e., cue validity). If the combined weight of features an object possess exceeds a particular threshold (i.e., the membership threshold; Komatsu, 1992), an object may be considered to be a member of the category. However, later work suggested cue-validity was an imperfect measure of similarity-based membership (e.g., Murphy, 1982), in particular because it did not accurately predict a basic-level advantage.

Prototype theory has remained relevant as a theory in categorisation. Its alternatives (outlined below) are either too complex (in the case of exemplar theory) or vague (in the case of theory-theory) and are therefore not as widely used as an explanation of categorisation. As such, prototype theory is still the topic of research, either in its own right or as the baseline against which alternatives (e.g., exemplar models) are compared (e.g., Minda & Smith, 2001; Nosofsky & Stanton, 2005; J. D. Smith, 2014; J. D. Smith & Minda, 2000). It has been adapted in broader theories of conceptual processing, which focus on modelling reliance on prototype theory and its competitors in one conceptual system (Aerts et al., 2016; Voorspoels et al., 2008), and been referenced in various parallel fields such as cognitive linguistics (e.g., (Geeraerts, 2016), and healthcare psychology (Hofmann, 2017).

1.2.4.1 Can prototype similarity explain taxonomic effects?

Like the definitional view before it, prototype theory comes from a family of closely related ideas on the nature of conceptual representations, the process of determining similarity between objects and concepts, and various behavioural effects that confounded existing beliefs on the nature of categories. Prototype theory was closely tied to the discovery that categories varied along a horizontal dimension of typicality, and the effects this had on categorisation behaviour, although the relationship between explanations of taxonomic effects and prototype theory is less explicit. However, many influential explanations of varying categorisation as a result of taxonomic levels (Murphy & Brownell, 1985; Murphy & Smith, 1982; Rosch, 1978; Rosch, Mervis, et al., 1976) are also deeply rooted in prototype theory's underlying assumption that abstract feature summaries form the basis of conceptual representation.

On influential explanation of the basic-level advantage assumes that features that may be perceived in the world are not distributed at random (Rosch, 1978), but rather that they can be perceived to occur more frequently in certain combinations than in others (e.g., <u>*has wings*</u> is more likely to co-occur with <u>*lays eggs*</u> than with <u>*has gills*</u>), forming a complex structure of related and unrelated features. On this view, categorisation at different taxonomic levels refers to applying different ways of dividing up this ‘perceived world structure’ into representative and information-rich bundles, with the aim of being maximally informative and cognitively economical. Crucially, this account, also referred to as the differentiation account (Murphy & Brownell, 1985) assumes that categories differ in terms of the degree to which they maximise within-category feature similarity and minimise between-category feature overlap (e.g., within-category feature similarity is greater for *dog* than for *animal*, while between category-dissimilarity is greater between *dog* and *bird* than between *Labrador* and *Golden Retriever*). The basic-level, named for being the taxonomic level at which Rosch argued the most ‘basic’ cuts in the perceived world structure, is argued to be generally preferred not because it occupies a particular position in a hierarchically structured semantic memory, but because it is maximally differentiated from other categories at the same taxonomic level (Markman & Wisniewski, 1997; Murphy & Brownell, 1985; Murphy & Smith, 1982; Rosch, 1978; Rosch, Mervis, et al., 1976). That is, basic-level categories describe instances that are maximally similar within-category (e.g., instances of *dog* share many features with other *dogs*) while being maximally distinct between-category (e.g., *dogs* share few features with *fish*). By contrast, categories superordinate to the basic level are more abstract, and describe instances that are less similar within-category (e.g., *animal* describes *dogs* and *fish*), and between-category (e.g., instances of *animal* differ considerably from instances of *vehicle*). Meanwhile, categories subordinate to the basic level are more specific, and describe instances that are highly similar within-category (e.g., *Labrador* describes only a subset of four-legged, furry, medium-sized animals with wagging tails), but not very distinct between-category (e.g., *Labradors* share many features with *golden retrievers*). Another, related explanation of in particular basic-level advantages in category verification is Murphy and Smith’s (1982) preparation model. This is a process model of in particular label-image verification, which assumes that a given category label may activate a particular feature-representation. This representation is subsequently verified upon seeing an image of an object. If the object is a clear match or mismatch, a decision is made. If the object is not a clear match or mismatch, additional processing is required. Importantly, the preparation model assumes that superordinate-level categorisation is generally slower because there is no single perceptual representation of concepts at this level of abstraction (e.g., no single perceptual representation of *animal*), and requires more information to be active at the same time. The advantage over the subordinate model then is thought to stem from a higher membership threshold because of greater between-category similarity (e.g., the threshold for determining that something is a *bird* or a *dog* is lower than for determining whether something is a *sparrow* or a *finch*).

As such, influential explanations of the basic-level advantage build on the same assumptions that underly prototype theory: categories map onto feature-clusters, not feature-definitions, and similarity between objects and concepts may be determined by examining the features they share. As such, they inherit the same problem as all other feature-based accounts: they are increasingly contradicted by accounts of conceptual representation that argue concepts do not comprise discrete binary features and dimensions, but rather a combination of sensorimotor experience and linguistic distributional knowledge. However, mechanisms such as the ones proposed by preparation and differentiation accounts (Murphy & Smith, 1982; Murphy & Brownell, 1985) provide valuable insights and may still explain categorisation behaviour when feature-based representations are exchanged for sensorimotor-linguistic representations.

1.2.4.2 Can prototype theory explain typicality effects?

Prototype theory was developed with a graded structure of categories in mind. As such, it is uniquely positioned to explain object typicality effects. The crucial point prototype theory makes about categorisation is that what binds category members is not a set of defining features, but rather the degree to which they resemble one another. Allowing category members to possess a varying number of the features common to their category neatly fixes the issues with the definitional view. Moreover, in this way it could be argued that not only do categories map onto particular clusters of features, within those categories, members map onto further subdivisions within those clusters. That is, following on from Rosch's (1978) explanation of taxonomic effects, object typicality may be explained as an extension of the same process: prototypes are abstractions of or members of categories that are maximally similar to other category members (i.e., maximise family resemblance) while also being maximally distinct from contrasting category members (e.g., within the category *bird*, *robins* share more features with other *birds* than *ostriches*, and share fewer features with contrasting categories). Furthermore, the more similar a given instance is to the prototype, the more typical it is of its category, and the easier it is to categorise it as a member of said category. Evidence for this explanation comes from studies showing people are better at learning patterns that are closer to a target prototype, as well as better at classifying typical compared to atypical category members (Rosch et al., 1975).

However, several critical accounts have argued that prototype theory unjustly assumes that superficial typicality effects (e.g., Murphy & Brownell, 1985; Rips et al., 1973; Rips & Collins, 1993; Rosch, 1973; Rosch & Mervis, 1975) reflect a deeper category structure (Lakoff, 1987; Rosch, 1978). Prototype theory was conceived as a relaxation of the constraints from the definitional view, primarily based on the finding of typicality effects that affected categorisation performance and has taken similarity to the prototype to equal degree of membership. However, some researchers suggest that graded structure may not reflect a degree of membership (e.g., Lakoff, 1987). That is, while *ostriches* and *penguins* are rated as atypical *birds,* they are still considered to belong to the category *bird*, and therefore typicality cannot reflect the underlying categorical structure. While this example of a clearly defined category may be criticised[2], it is supported by other work (Armstrong et al., 1983), which showed that participants were perfectly willing and able to assign typicality ratings to

---

[2] Lakoff's example seems to imply that the biological category *bird* is in fact discrete, and membership of this category is constant. However, while this may currently be the case, the principles of evolution dictate that at some point, there may have been a *bipedal*, *winged* and *feathered* creature that straddled the boundaries between biological categories (e.g., *bird* and *reptile*, in the case of feathered dinosaurs).

instances of categories that have a clear definition (e.g., rating 7 as a more typical instance of the category *odd number* than 57). Armstrong et al. took this to mean that prototypicality, or 'exemplariness' is separate from category membership and indeed structure. Nevertheless, Hampton (1998) demonstrated a clear relationship between typicality ratings and the predicted probability of category membership, showing that participants in McCloskey and Glucksberg (1978) were more likely to decide an object belonged to a given category when it was rated as more typical.

By contrast, Rips and Collins (1993) found that, as expected, typicality ratings correlated with distance to the category's central tendency, but that membership rating correlated with frequency of instantiation (i.e., the number of times a given instance occurred in the distribution). Related to this, Barsalou (1983, 1985) found that the mechanisms for determining typicality in ad-hoc categories (i.e., goal-derived categories, constructed to achieve a goal, for example *things to take on a holiday*) differed from that in taxonomic (e.g., *bird*) categories. In taxonomic categories, similarity to the central tendency and frequency of instantiation proved to be the best measure of typicality, whereas in ad-hoc categories similarity to the *ideal* (i.e., the features an instance should have to serve a particular goal, such as *fits in a suitcase* for the ad-hoc category *things to bring on holiday*) proved to affect typicality judgments most. The finding that typicality and category ratings differ is a problem for prototype theory, which is built around the notion that category membership is determined by verifying similarity to the prototype (central tendency). Furthermore, it is not clear how this view can account for the typicality effects found in non-taxonomic categories. Finally, it is not immediately evident how the dimensional ideals (e.g., *how happy people are to receive it* as an ideal of *birthday present*; Barsalou, 1985) relate to the discrete features of prototype theory.

1.2.4.3 Other limitations of prototype theory

In addition to the questions regarding prototypicality as the basis of category membership, some accounts have criticised the way prototype theory has defined similarity as a measure of categorical structure. In a criticism that may be extended to all feature-based accounts of categorisation, Murphy and Medin (1985) argue that similarity (e.g., to a central tendency, to other category members) by itself is not enough to determine category membership, and that similarity requires constraints, or rules, to be meaningful and not circular (e.g., category members are similar, hence they belong to the same category).

However, in theory, feature-based accounts of categorisation *do* constrain similarity relationships. For example, the definitional view would argue that all members of a category are similar because they share the necessary features, and not similar when they do not. In prototype theory, similarity to the central tendency is defined as the sum of the features shared between the object (e.g., *Labrador*) and the category (e.g., *dog*; Hampton, 1979, 1995; Rosch & Mervis, 1975). When this number crosses a specific threshold, the object is a member of the category. Crucially, features are assumed to vary in their relative weight depending on the task at hand, their relative frequency in the category as well as familiarity (A. Tversky, 1977), so that possession of certain features may point more strongly towards category membership. For example, features such as <u>*has two legs*</u>, or <u>*has two eyes*</u> might be weighted as less important than <u>*has wings*</u> in determining membership of the category *bird*. This weighting might differ further depending on sentential context (Roth & Shoben, 1983; A. Tversky, 1977) for example, the sentence context: *The pet shop was full of animals*, constrains the relative important of features of *animal*, such that <u>*domesticated*</u> is weighted as more important than <u>*has wings*</u>. Taken together, this suggests that prototype theory does place some constraints on similarity: objects are similar to the central tendency if they share enough key features.

Alternatively, the problem with similarity in prototype theory may not be that it does not offer rules for its application, but rather, that the rules are vague. For example, as previously determined (e.g., McCloskey and Glucksberg, 1978), category boundaries may be fuzzy, and membership judgments may vary from person to person and even time to time. In terms of prototype theory, that would imply that the threshold for category membership (i.e., the number of shared features) shifts. What is enough feature-overlap in one context or person, is insufficient in the next. While some have embraced this variability as a strength of prototype theory (i.e., a shifting membership threshold accommodates fuzzy categories; (Hampton, 1995), the mechanism underlying this threshold shift is not readily explained within the framework of prototype theory. If feature-similarity to the prototype (or in fact other category members) determines category membership, how can this change within the same category (e.g., *animals*)?

One possibility is that the threshold for category membership depends on the selection of features for the similarity comparison. This leads to the criticism of prototype theory's specification of feature determination: the question of what counts as a feature in determining similarity between objects and prototypes to begin with, as objects may have a sheer infinite list of features that are technically true (Murphy & Medin, 1985). For example, it is technically a feature of *Labradors* that they are not blue, *have more than two hairs*, *weigh less than a grand piano* etc. It is unlikely that participants take all possible features an object may possess into account when making similarity judgments. And indeed, this is not what prototype theory proposes. Tversky (1977) suggests that similarity judgments, or in fact identification of an object, require the extraction of a limited list which allows for the completion of the task at hand. As Armstrong et al., (1983) point out, participants omit obvious features when listing features, focusing on only those features they deem relevant to the task at hand. Considering this, it is possible that, as Barsalou (1983, 1985) argues, the conceptual system is flexible. People may form multiple concept representations (e.g., prototypes) of the same category, allowing them to tailor representations to the context and the task at hand. The problem for prototype theory is that this does seem to relax its constraints to a degree that the theory fits any categorisation task.

In summary, prototype theory circumvents the issues with the definitional view by arguing concepts do not need a set of necessary and sufficient features but argues that concepts are represented by a summary of the features common to the category. Categorisation then becomes the process of summing the features observed in an instance and comparing them to the feature summary. Out of the three main theories of object categorisation discussed here, prototype theory has come closest to explaining taxonomic and typicality behavioural effects. It was designed to account for the typicality effects that reflect graded category structure, and influential accounts of the basic-level advantage share prototype theory's assumption of categorisation via feature-similarity and conceptual representation through feature clustering. Crucially however, features and prototypes are not sufficiently specified, that it is not clear what exactly determines what may count as a feature. Moreover, it has been suggested that prototype effects may not reflect conceptual structure. Consequently, much research has gone into developing alternative or complementary theories of conceptual representation and categorisation, most notably exemplar and theory-theory, but has also moved to develop (computational) models of conceptual processing, with a smaller focus on categorisation.

1.2.5. Exemplar theory

Prototype theory and other feature-similarity accounts assume that categories are represented through an abstracted summary of the features that are common to that category. Crucially, the prototype itself may form and change through exposure to category exemplars. For example, a given person's prototype representation for the category *dog* is likely dependent on their experience with various dog breeds. If they have only encountered *Labradors*, their prototype is likely to contain all features of *Labrador*. However, if they then encounter a mix of different dog breeds (e.g., *dachshund*, *spaniel, poodle*), their prototype will contain features that are common to all these dog breeds, and shed any idiosyncratic features. When this person is then faced with a categorisation task (e.g., is *x* a *dog*?) they presumably compare the features of *x* only to the *dog*-prototype. In this manner, prototype theory is an example of *early computation* (Estes, 1986) – only the result of category learning (i.e., the prototype) is called upon during categorisation.

A contrasting theory is exemplar theory (Brooks, 1978; Medin & Schaffer, 1978), which argues that concepts may be represented not through an abstracted feature summary, but rather through knowledge of specific instance or subset (e.g., the concept dog is represented either by a memory of a particular dog, or a subset of dogs such as *guard dogs*; Murphy, 2002; E. E. Smith & Medin, 1981). Exemplar representations may allow for more flexibility in categorisation – in contrast to prototype theory, idiosyncratic information about exemplars is retained, and are an example of *late computation* (Estes, 1986). On this notion, exemplars stored in semantic memory may be categorised differently depending on environmental and task contingencies. For example, a conceptual representation of *dog* that consists of a wide range of encountered dog-exemplars can be used to flexibly represent dogs in various scenarios, whereby task-dependent constraints determine which features are relevant (e.g., “watch out for the *dog*" invokes another set of exemplars than “that *dog* is cute”, depending on stored experience with various dogs).

Exemplar theory predicts that when deciding whether a given object or instance belongs to a particular category, people rely on experience with known exemplars (Brooks, 1978; Murphy, 2002). For example, when trying to identify a previously unseen animal, one might search their memory for exemplars that might match it (e.g., *dogs, horses, lions*). To determine which of the possible categories the unknown animal belongs to, you rank all possibilities by the degree of similarity to the animal you are observing. It may bear a few similarities to a horse, a few more similarities to a lion, and be most similar to a dog. According to Murphy (2002), this mechanism also allows exemplar theory to account for the prototypicality effects that led to prototype theory. Typical category members (e.g., *robins*) are similar to other bird exemplars, but not so similar to other animal exemplars (e.g., *fish*). Atypical members (e.g., *penguins*) are less similar to other bird exemplars, and more similar to other animals (e.g., *fish*), although this particular explanation is contested (see below; (J. D. Smith, 2002).

Different models of exemplar theory include Medin and Schaffer’s context model, (1978) Nosofsky’s (1986) generalised context model (GCM) and exemplar-based random walk model (EBRW; Nosofsky & Palmeri, 1997). The context-model involves a feature-comparison between all exemplars of competing categories. That is, the evidence for an instance's (e.g., *Labrador*) membership of a particular category (e.g., *dog*) is equal to the combination of the similarities of the instance to all *dog* exemplars divided by its similarity to all stored exemplars (Medin & Schaffer, 1978). A theoretical challenge for the context model is the lack of constraints on the features relevant to the task, as it compares *all* members for contrasting categories. While this predicts performance in small categorisation tasks with few competing categories, such as the one Medin and Schaffer tested, it is hard to see how this would extrapolate to a full-scale theory of conceptual processing. Some researchers have extended the basic exemplar model with additional parameters (e.g., to account for guessing; J. D. Smith & Minda, 2000)

As a possible way of constraining the number of retrieved exemplars, the GCM (Nosofsky, 1986) builds on the context model by transforming its simple similarity measure into a measure of distance in multidimensional space, which may vary in structure (i.e., minimising problem-solving space, for example by restricting attention to relevant dimensions, expressed as attention weights).

The EBRW model (Nosofsky & Palmeri, 1997) expands the generalised context model. It represents exemplars as points in a similarity space, which may be altered in structure by selective attention processes. Categorisation is the process of matching an object to the stored exemplars. For example, the categorisation of an instance (e.g., *Labrador*), in one of two categories (e.g., *dog* or *bird*) enters all exemplars stored in semantic memory in a 'race', where the probability of them completing the race within a given time is determined by their semantic distance to the instance. The random walk counter - which expresses evidence for either category - starts at zero, and changes positively when evidence in favour of *dog* is acquired, and negatively when evidence in favour of *bird* is acquired. Consequently, if the 'winning' exemplar belongs to the *dog* category, the counter increases, conversely, if the 'winning' exemplar belongs to the 'bird' category, the counter decreases. If the counter exceeds a given threshold, a category decision is made, otherwise a new race is started. Nosofsky and Palmeri found that this model fit response time data in a set of classification tasks well.

1.2.5.1 Can exemplar theory explain taxonomic effects?

There seems to be no explicit exemplar-theory account of taxonomic effects in object categorisation. It is unclear how once an individual has identified a group of objects as a category member (e.g., *dogs*) they can infer that *dog* exemplars are also *mammals.* Murphy (2016) argues that it is unlikely a person would encode every instance of *dog* as both a member of the categories *dog* and *mammal*, as this would take up too much of a person's cognitive resources. However, note that this is only true if exemplars are explicitly stored with each taxonomic category during learning. If, by contrast, the assignment of exemplars to various categories is considered to be a flexible process that may draw upon exemplar memory to accommodate various categorisation tasks (i.e., late computation; Estes, 1986) exemplar-based categorisation at various taxonomic levels may be possible. That is, computations for the categorisation of a given instance of *dog* (e.g., *Labrador*) as *dog*, *mammal* and *animal* involve similarity computations on the same set of exemplars. Moreover, it may be the case that not every exemplar is retrieved every time (Estes, 1986; Heit & Barsalou, 1996). In fact, some research on exemplar-based explanations of object typicality (see also below) suggests that correlations between exemplar-based predictors and various categorisation performance measures (e.g., category naming frequency, response time and typicality) for various categories (e.g., fruits, birds, vehicles etc.) increase in strength when these predictors are calculated based on up to 10 of the most frequent exemplars (Smits et al., 2002). As such, an exemplar-based explanation for taxonomic categorisation could be that it draws on a limited set of frequent exemplars which are flexibly selected from semantic memory depending on the categorisation task at hand.

However, this leaves two questions. Firstly, by what mechanism are relevant exemplars retrieved if they are not a priori stored in semantic memory at various taxonomic levels? Secondly, how could exemplar-based categorisation account for taxonomic effects (e.g., basic-level advantage)? In some ways, the former question is similar to that aimed at prototype theory: by what mechanism are relevant features extracted to form the prototype? Both questions concern a mechanism that is hard or even impossible to observe. In the case of prototype theory, the likelihood that a given feature is associated with the prototype representation of a given category is measured through feature generation tasks (Hampton, 1979; Rosch and Mervis, 1975), based on the assumption that features that are named more often are more salient because they are more common to a category. In the case of exemplar theory, the likelihood that a given exemplar is activated in response to a given category label may be measured through exemplar-generation tasks (Smits et al., 2002; Storms et al., 2000)based on the assumption that exemplars that are generated more frequently in response to a given category have a stronger association with said category, and are more likely to be activated in the exemplar representation. Both frequency-based measures of prototype-features and exemplars are highly related with various measures of categorisation performance (e.g., production frequency, response time; Smits et al., 2002). However, in the case of exemplar-models, it is still unclear how it is that exemplars become associated with a given category in the first place.

1.2.5.2 Can exemplar theory explain typicality effects?

Theoretically, exemplar theory could explain object typicality in a similar fashion to prototype theory. That is, prototype theory assumes that greater typicality means greater similarity to the prototype, which means to have more features common to the category. As a result, highly typical items are not only closer to the prototype, but also bear greater family resemblance to other members. It is therefore possible that degree of similarity to stored exemplars reflects object typicality just as well as similarity to the prototype might (Murphy, 2002).

Traditional exemplar models (e.g., Medin & Schaffer, 1978; Nosofsky, 1986) are not well positioned to test this hypothesis, because they typically use binary categories which contain a limited number of members and do not have a clearly defined graded structure

(Minda & Smith, 2001). The evidence that does exist is inconclusive. One the one hand, there is evidence from modelling work on the categorisation of dot-patterns (J. D. Smith, 2002) which suggests that similarity to the prototype (i.e., the prototypical dot-pattern) is a better way of determining object typicality than similarity to exemplars (i.e., individual dot-patterns), simply because a given instance will always be similar to some exemplars, and dissimilar to others. By contrast, similarity to the prototype, which collapses members in to a singular point, may be easily determined and reliable as a predictor of object typicality. However, Storms et al. (2000) found evidence that exemplar-based similarity ratings correlated more strongly with ratings of object typicality, as well as with response times, exemplar generation and category naming frequencies than prototype similarity calculated from Hampton (1979). Similarly, Smits et al. (2002) showed that exemplar-based GCM's predicted categorisation of both well-known and novel fruits and vegetables better than prototype-based multiplicative similarity models, and that categorisation of familiar but atypical fruits and vegetables was predicted worse by their feature-similarity to the prototype, and suggest that supplementing prototype models with exemplar knowledge as Minda and Smith (2001), might improve model performance. This would be in line with the earlier finding that atypical items are categorised faster and more accurately at the subordinate (i.e., more specific) level (e.g., (Jolicoeur et al., 1984; Murphy & Brownell, 1985). At the subordinate level, category exemplars (e.g., *Labradors*) are less likely to vary within-category than they are at the basic (e.g., *dog*) or superordinate (e.g., *animal*) level, thus stored experience with exemplars is a better predictor of subordinate category membership (also evidenced by the fact that experts are equally good at categorising at the subordinate and basic level, for categories that fall within their expertise). It may still be the case that exemplar experience is abstracted into a prototype-like representation (Hampton & Jönsson, 2008), however at more specific taxonomic levels exemplar knowledge and prototypes are harder to distinguish. That is, a prototype representation of *golden retriever* is necessarily more similar to *golden retriever* exemplars than the prototype of *dog* is to *dog* exemplars, which vary more within-category (e.g., *chihuahuas* and *Danish mastiffs*). As a consequence, similarity to the prototype may be more important for basic-level categories, whereas exemplar similarity may be more important for subordinate-level categories.

1.2.5.2 Other limitations of exemplar theory

The key criticism levelled at exemplar accounts is that their evidence comes primarily from computational models, where model predictions are compared to human classification behaviour. Moreover, the category sets that these models are typically trained and tested on have a number of limitations. Firstly, much of the most influential work on exemplar models centres around the classification of training and transfer instances into a binary set of contrasting categories (e.g., (Medin & Schaffer, 1978; Nosofsky, 1986). Secondly, these sets contain very few members, participants and models are often trained on 5 instances of one category and 4 of the other (i.e., 5-4 category structure; see J. D. Smith & Minda, 2000; for an extensive evaluation of this paradigm) and tested on a total of 16 instances (including 9 that were learned during training). Thirdly, the members of each category are only marginally more similar to one another than to members of the contrasting category (Minda & Smith, 2001; J. D. Smith & Minda, 2000). Note that all of this is a deliberate design choice in most exemplar studies; it allows control over prototypes and diagnostic information. Nevertheless, using a limited number of category members for each category, that are moreover relatively distinct could inadvertently cause participants to rely more heavily on the memorisation of particular exemplars (Murphy, 2016), and favour exemplar models over alternatives. Indeed, J.D. Smith and Minda found that when testing on data from a number of studies with a typical 5-4 category structure, exemplar models were generally favoured over additive and multiplicative prototype models in predicting categorisation outcomes, unless prototype models were allowed to retain exemplar knowledge. Nososfsky (2000) argues that enhancing prototype models in this manner is unlikely to work when training instances are not included in the test phase. While this may be correct, it is unclear whether exemplar models would work much better in a similar scenario.

On a more theoretical note, Murphy (2016) argues that exemplar theory is heavily focused on the classification aspect of categorisation, and not so much on inferring general knowledge from these categories (i.e., the knowledge that members of a given category possess a given feature). While prototype representations are a generalisation of features across all category members, exemplar theory does not store such knowledge directly. Rather, category-feature relationships are inferred from exemplars. For example, the *dog* exemplars I know all *have fur*, hence I might infer that dog *have fur*. A novel instance that *has fur* is then more likely to be categorised as a dog. Murphy argues that this mechanism is problematic because it does not allow for representing knowledge about a category may not be directly tied to experience with an exemplar (e.g., the knowledge that *birds* evolved from *dinosaurs*, or the fact that mammals have a diaphragm). In a strict exemplar view this may indeed be problematic, although it hinges on the kind of information one thinks may be stored in an exemplar, and whether exemplar knowledge can only be acquired through direct perceptual experience.

Furthermore, exemplars do not represent knowledge about causal relationships between features (such as *has wings, can fly*) that have been suggested to bias categorisation (Ahn, 1998). How do exemplars represent knowledge about the relationship between wings and flight? Whereas some (e.g., Rosch, 1978) suggest that features are not randomly distributed randomly, but correlate (exemplars which possess the feature *has wings* typically *can fly*, thus capturing the idea that wings are necessary for flight), others (Murphy, 2016) argue that not every correlating feature has the same causal relationship (e.g., *has a beak* frequently co-occurs with *feathers*, yet few people would argue having a beak is essential for having feathers). A potential counterargument is that while Murphy's (2016) examples work within the category bird, where *has a beak* and *has feathers* may co-occur about as much as *has wings* and *can fly*, the co-occurrence between wings and flight extends throughout the animal kingdom and beyond (e.g., various *mammals*, *insects*, *fish, airplanes* all use wings to fly) to the point that wingless flight may be rare, whereas featherless exemplars with beaks are not (e.g., *squid, fish*, *turtles*). Moreover, Murphy seems to suggest that prototype theory would be better suited for representing causal relationships, simply by allowing for generic features (e.g., <u>wings usually enable flying</u>). If prototypes may possess such generic features, why are they off-limits to exemplars?

In summary, exemplar theory argues that concepts are represented through their most salient exemplars. Evidence for exemplar theory comes from models that successfully predict category learning and response times in identification tasks. The strongest criticisms of exemplar theory are that its predictions are limited to models of classification, which often rely on small binary sets that favour memorisation of exemplars. If exemplar theory is to be a full-fledged theory of conceptual representation, it needs to account for a wider range of behaviour on more varied concepts. Furthermore, it is not clear how exemplar theory extends to knowledge inferring and displays the same lack of specification of the constraints on similarity that plagued prototype theory. In particular, the lack of constraints of what exemplars constitute a concept. At present, exemplar theory by itself has not been sufficiently extended to fit human behaviour, as pure-exemplar theory has been tested in models (Murphy, 2016). However, some researchers have argued that knowledge of exemplars *as well as* prototype knowledge might be relevant to categorisation (Hampton, 2007; Minda & Smith, 2001).

1.2.6. Summary of traditional theories of object categorisation

The definitional view, prototype theory, and exemplar theory share a number of assumptions. Firstly, they assume to some extent that the smallest representational unit is an abstract, binary perceptual or functional feature (e.g., <u>has wings, can fly</u>). That is, traditional accounts of categorisation assume concepts and objects are componential (Palmer, 1977). In other words, concepts and objects are the sum of their parts (e.g., the concept bird is represented as a combination of its features, such as <u>has wings</u>, <u>has feathers</u>, <u>bipedal</u>, <u>can fly</u> etc.). These accounts typically assume that features are static and transferrable, such that category membership may be dependent on the possession (or lack) of a particular feature, rather than variation within features themselves (perhaps with the exception of *dimensions* such as size and weight). A second crucial assumption of feature-based theories of categorisation is that features are the indivisible building blocks of concepts (Schyns et al., 1998), whose relationship to perceptual information is symbolic. That is, features *encode* perceptual and functional information, which can then be used in conceptual processing such as conceptual combination and categorisation. This notion is important. For practical reasons, studies on categorisation have frequently relied on participant-generated feature lists. A critical assumption was that the features listed by participants have been taken to be sufficiently grounded in the real world. That is, as Rosch (1978) illustrates when discussing her work on basic-level categories (see chapter 2), the idea was that the correlational structure of features in the perceived world (e.g., <u>*has wings*</u> frequently co-occurs with <u>*has feathers*</u>) would constrain participant feature listings to the extent that, while they might not know all features or even their internal relationships, they would not provide correlational structure where there was none (e.g., <u>*has wings*</u> and <u>*lives underground*</u>). As such, the features listed by participants were taken to be sufficiently reflective of perceptual information.[3]

However, as Rosch (1978) also notes, the features participants listed were not always uniformly interpretable away from their category. For example, features related to size (e.g., *large*) varied in meaning depending on the category they were assigned to. Moreover, some features had no clear meaning unless paired with their category (e.g., for example the feature <u>*has a seat*</u> requires functional knowledge of *chairs* to be meaningful). Rosch (1978) interpreted this to suggest that the analysis of objects into their features was a complicated process, that might only be possible after categories have already been established. At the very least, it suggests that some features are not useful for categorisation, as they only become informative once the category is known, although this leads us down back down the path we just came from (i.e., defining, and sufficient features).

---

[3] This refers to the view of features that was prevalent when the traditional accounts of categorisation discussed here were conceived. More recently, some have explicitly stressed that feature lists are not meant to be taken as a literal reflection of conceptual structure (McRae et al., 2005).

Another consequence of Rosch's (1978) notion is that features may not always be the static indivisible and independent building blocks that they are made out to be. Indeed, research has suggested that features may be interpreted differentially depending on context, as with the example of *seat*. For example, Medin and Shoben (1988) found that the relative importance of features in object categorisation changes depending on other features of the object that is categorised (e.g., *large* as a feature of *metal spoons* compared to *wooden spoons*). If features are the building blocks of concepts, how is it that they may change so freely? How can a feature such as *large* mean something else from one concept to the next? One answer is that the linguistic label *large* does not refer to something indivisible, but that it refers to a set of smaller properties that for reasons of cognitive economy cannot easily be listed but is grouped under a property that is flexible depending on context. Another explanation is that features are not the actual currency of semantic representation.

An additional question regards the lowest-level status of features. That is, feature-based theories typically do not address what makes up a feature. This begs the question what the lowest level of a feature is. What defines it? If *bird*, *chicken*, and *robin* all have the feature *has_wings*, then what features does *wing* have? If we do not allow *wing* to have features, why not? On the other hand, if we do allow *wing* to have features, what would these features be? Would they include *part of a bird*? If they do, that will result in a circular definition which is obviously problematic (e.g., Chinese Room argument, Searle, 1980)). If *wing* does not have features, then why not? Is it not a concept (e.g., like *robin*)? E.E. Smith and Medin (1981) seem to suggest that features are indeed concepts, but sidestep this question by arguing that the division between feature and concepts needs to be made on empirical grounds: in categorising an object, it is more useful to assign *wing* a feature-status, as it helps explain the difference between specific objects (e.g., *bird* and *mouse*), whereas the features of *wing* may be less suitable for this purpose.

Finally, the predictions from feature-similarity accounts of taxonomic effects are at odds with findings from research into the development and deterioration of conceptual knowledge. That is, feature-based accounts assume that basic-level categories are privileged, for example because they map onto feature-clusters that are highly informative and highly distinct (Murphy & Brownell, 1985) or because are the most likely entry point into a taxonomically structured semantic memory (Jolicoeur et al., 1984). However, research has shown that children may be earlier to differentiate between objects at the superordinate level than at more specific levels of abstraction (Mandler & Bauer, 1988; Mandler & McDonough, 2000). This broad-to-narrow pattern is mirrored by findings in research on the deterioration of semantic memory, which reports patients with semantic dementia (SD) generally lose the ability to categorise at specific levels of abstraction before the ability to categorise at broad abstract levels (Hodges et al., 1994; Hodges & Patterson, 2007; Warrington, 1975). However, the inverse effect has been observed in patients suffering from aphasic stroke (i.e., left hemispheric strokes), which showed a marked advantage for categorisation at the specific subordinate level compared to the more general basic level (Crutch & Warrington, 2008). Moreover, other work reports a case whereby a patient with damage across multiple areas of the brain (i.e., superior and middle frontal gyri, superior middle and temporal gyri, and lateral occipital gyri) lost the ability to categorise at the superordinate but not the basic level (Forde & Humphreys, 2002; Humphreys & Forde, 2005).

Research on effects of semantic impairments on taxonomic categorisation is a subset of a much broader area of work on category-specific impairments (see Clarke & Tyler, 2015; Mahon & Caramazza, 2009). However, the finding that SD patients are disproportionally restricted in their ability to categorise at specific taxonomic levels in particular has been explained as having profound implications for the structure of semantic memory, as well as traditional explanations of taxonomic and typicality effects in object categorisation (e.g., Ralph et al., 2017; Rogers & Patterson, 2007). With regards to taxonomic effects in categorisation, the finding that patients with aphasic stroke or semantic dementia retain only the ability to categorise either at the subordinate or superordinate level contrasts strong network accounts which argue that increased processing time for subordinate-level items can be explained by the time it takes to progress through a semantic network from a basic-level entry point.

In the following section, I will explore views that are alternatives to feature and network-based theories of conceptual representation. First, I outline theory-theory, which expands prototype theory with contextual knowledge. While it is a feature-based account, it bridges some of the gaps to more recent accounts. I then discuss linguistic distributional accounts, simulated accounts and finally integrated simulated-linguistic accounts. Finally, I outline a potential explanation of both taxonomic and typicality effects that builds on recent advances in our understanding of conceptual representation (i.e., sensorimotor-linguistic representations) that may be a viable alternative to the traditional feature-based accounts with which they are so strongly associated.

1.2.6.1 Adaptations of feature-based accounts: theory-theory

One theme that returns throughout the categorisation literature is the idea that a theory of categorisation must balance between allowing for exceptions while maintaining constraints. Categorical representations must simultaneously flexible enough to allow *ostriches* to be categorised as *birds*, but must also be constrained enough to allow people to use categories effectively in daily life. For prototype theory, this flexibility comes from allowing category instances to vary in the degree to which they share features common to said category. Constraints come in the form of feature importance. Not all features are assumed to be equally informative (e.g., when identifying an unknown animal, it may be more informative to verify that it *barks* than that it *has fur*). However, what features are more informative depends on the context in which a particular classification takes place (Minda & Smith, 2001), and prototype accounts generally do not elaborate on the mechanism that underlies the process of selecting features depending on context.

Theory-theory (Medin, 1989; Murphy & Medin, 1985; Murphy & Spalding, 1995; Wisniewski & Medin, 1994) builds on the criticisms of prototype theory, and is an attempt to address the crucial deficiency of all traditional accounts of concepts and categorisation; their reliance on feature-based similarity computations (see Goldstone, 1994; for a review on the role of similarity in categorisation). Theory-theory accounts argue that feature-based similarity judgments may not be based solely on perceptual input (Goldstone, 1994). Indeed, it has been suggested that similarity judgments are affected by information about the categoriser's expertise, intentions, and goals (Barsalou, 1982, 1983) as well as the context in which the judgment takes place (Roth & Shoben, 1983).Therefore, theory-theory accounts argue that for the feature-matching seen in prototype and exemplar theory to work, the notion of a feature may need to be expanded to include a *theory* about relations to other features and concepts and the processes that operate on it (Murphy & Medin, 1985). Consequently, theory-theory is an attempt to include knowledge about the world around us into the conceptual representation. In theory, this neatly solves the problem of unconstrained feature selection; we may select features for a concept based on our prior world knowledge. For example, features such as *has four legs*, *flat surface* are more relevant to the concept *table* than *can float*, even though upon inspection we may determine that a table indeed floats. However, because of the knowledge that furniture is rarely used in a context where *floating* is required, a feature summary for *table* is unlikely to include *can float*.

Evidence supporting theory-theory comes from category-learning experiments, which show that participants rely on prior knowledge to constrain which features are relevant for a given concept within a given context (Bridges et al., 2020; Murphy & Allopenna, 1994; Wisniewski & Medin, 1994). However, modelling work has shown that such prior knowledge may naturally arise from the available data. For example, Rogers and McClelland (2008) describe a neural network that was with a set of instances that either did or did not possess a given set of properties (e.g., has roots, can move). Crucially, they artificially assigned the properties (*is big*, *is small*, *is bright* and *is dull*) so that they were only indicative of one concept (e.g., all trees were *big*, all flowers were *small*, but both could be either *bright* or *dull*. By contrast, all birds were *bright* and all fish were *dull*, but both could be *big* or *small*).

Crucially, Rogers and McClelland found that the model selectively weighted size as more important in the categorisation of plants as trees or flowers, and brightness as more important in the categorisation of animals as birds or fish. This suggests that when learning to categorise objects based on their properties, if the model has learned that an object that *has roots* is a plant, *size* is more informative than *brightness* to determine whether it is a tree or a flower. This is not because the model has an a priori notion of the importance of size, but rather because it has *learned* that size distinguishes trees from flowers.

1.3. Alternative views of conceptual representation

1.3.1. Distributional models

The model employed by Rogers and McClelland (2008) is an example of a connectionist model of semantic memory (Rumelhart, McClelland, et al., 1986). Such models are part of a larger family of distributional semantic models (DSMs; see Günther et al., 2019; Kumar, 2021; Wingfield & Connell, 2019; for recent reviews). Distributional models assume that the meaning of a concept is not represented in semantic memory as a definition, summary or through exemplars, but rather as a distributed pattern across many dimensions (Kumar, 2021), most notably the manner in which concept-labels co-occur with other words in language. Connectionist models are part of a subclass of DSMs referred to as error-driven models. Models in this class are characterised by an architecture consisting of nodes in input, hidden and output layers. Activation from a cue spreads from the input to hidden layer(s), where nodes are activated depending on predetermined bias or weight, and propagates from these nodes to the output layer where an answer is generated. Crucially, most connectionist models are *trained* (e.g., on large bodies of text or feature sets) and receive feedback about their performance (e.g., through error backpropagation; D. E. Rumelhart, Hinton, et al., 1986) that is used to adjust the weights in their hidden layers, making it more or less likely that a node will activate or be suppressed. Recent incarnations, referred to as predict models, which are typically trained on very large text corpora (Winfield & Connell, 2019) include the highly influential Word2Vec (Mikolov et al., 2013) which is trained to predict a given linguistic cue following a given context (CBOW) or vice versa (skipgram), and improves its predictions gradually over many training examples using stochastic gradient descent to minimise computational load.

Connectionist models have been applied successfully to approach and even surpass human behaviour on a range of semantic tasks (e.g., see Baroni et al., 2014). In object categorisation, connectionist models have been used to successfully model the effects of semantic impairments on performance object categorisation, showing narrow-to-broad deterioration similar to that observed in patients with SD (Rogers & McClelland, 2004; Rogers & Patterson, 2007). The performance of these models, as well as their similar response to 'lesioning' has lent further support to claims of a semantic 'hub' that represents concepts as patterns of activation, much like a neural network might. The anterior temporal lobes, which are disproportionally affected in SD patients, have subsequently been earmarked as the possible location of such a semantic hub in the brain, although other brain regions have been suggested, as well as a system of multiple, specialised hubs (Jefferies et al., 2020).

While connectionist models have offered valuable insights into semantic representation, there are a number of aspects to their application that been argued to limit their cognitive plausibility. Some researchers have argued that major developments in neural network architectures were driven by a mathematical search for optimisation rather than a search for biologically plausible systems. An example is their reliance on methods of backpropagation and to inform the model where to adjust weights in the hidden layers, which seems to have no neurobiological correlate (Crick, 1989), although a number of biologically plausible alternatives have been suggested (see Lillicrap et al., 2020; Marblestone et al., 2016). Moreover, it is not consistently clear what connectionist models argue conceptual knowledge comprises. For example, Rogers and Patterson (2007) argue that conceptual representation is not feature-based, but rather consists of patterns of activation across multiple sensory, motor and linguistic modalities. However, models are typically trained on hand-picked sets of features (e.g., Hoffman et al., 2018) or network-like propositions (e.g., Rogers & McClelland, 2008), and/or argued to preferentially activate output nodes that correspond to properties (e.g., <u>*has wings*</u>) that may be used to differentiate between concepts (e.g., <u>*has wings*</u> allows the model to differentiate between *birds* and *fish*). Of course, it is likely that connectionist models use features as a convenient methodological and narrative shorthand (e.g., it is clearer to describe the differences between concepts in terms of differing features than it is to describe patterns of activation), but it necessarily restricts model predictions to those concepts for which comprehensive feature sets exist.

Alternatives to connectionist models include error-free or count-vector DSMs. These models may compute the representation of a given concept (e.g., *Labrador*) as a vector of word co-occurrences, which can be used to determine similarity to other words. That is, in a simplified example, the words *Labrador* and *spaniel* may co-occur with similar words (e.g., *leash, bowl*, *walk*, *bark*), and are therefore more similar than *Labrador* and *soup*, which co-occur with relatively fewer similar words (e.g., *bowl*). Examples of such models include Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996) Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) and Bound Encoding of the Aggregate Language Environment (BEAGLE; Jones & Mewhort, 2007) These models vary in their complexity and applicability. For example, HAL relies on relatively straightforward co-occurrence vectors, whereas LSA involves weighted log-transformation as well as dimensionality reduction steps, and BEAGLE accumulates information through repeated exposure.

Crucially, count-vector and predict DSMs have been highly successful in a range of semantic tasks. Relevant to this work is the finding that LSA predicted typicality ratings across a set of studies (Connell & Ramscar, 2001), and LSA-derived semantic associations between category names and features allowed for the correct categorisation of birds, fish, flowers and trees (Louwerse, 2011), as are other examples of successful semantic clustering based on word co-occurrence vectors (e.g.,. Bullinaria & Levy, 2007; Burgess, 1998).

Therefore, while recent accounts have argued connectionist predict models are preferable to count-vector models (e.g., Baroni et al., 2014; but see: Wingfield & Connell, 2019), it is the success of in particular early count-vector models such as LSA and HAL that led cognitive researchers to suggest that in conceptual representation, much (if not all) of the heavy lifting can be done by language. Indeed, throughout our lives, we are continuously exposed to staggering amounts of language (adults are exposed to around 100 000 words a day; Bohn & Short, 2009). Research has shown that people are sensitive to its statistical regularities, for example the fact that words are not distributed randomly, but cluster together (Wingfield & Connell, 2019). For example, words with the same meaning occur more frequently in the same context (e.g., *poodle* and *Labrador* may both frequently occur with words such as *bowl*, *walk*, *bone* and *pet*). As a consequence, the strongest of these accounts have argued that linguistic distributional information is an alternative type of mental representation (Burgess & Lund, 1997; Landauer & Dumais, 1997; Lund & Burgess, 1996).

1.3.2. Simulated accounts of concepts

The question of semantic content, or what it is that features are 'made of', has not received much attention in traditional feature-based accounts of categorisation (Meteyard et al., 2012). By contrast, simulated (or grounded, embodied) accounts of conceptual representation have focused specifically on the nature of semantic representations, and have argued against the division between semantic representations and perceptual experience. That is, simulated accounts have argued that concepts comprise direct information about our perceptual experience, bodily state and action (Barsalou, 1999; Glenberg & Robertson, 2000; Wilson, 2002; Zwaan, 1999). In doing so, they depart from amodal theories of cognition, which argue that perceptual experience is represented symbolically (Fodor, 2008; Fodor & Pylyshyn, 1988; Pylyshyn, 1984). The crucial difference is that where amodal accounts would argue that perceptual information about a given object (e.g., *dog*) is translated (transduced) into new representational units which allow for computation (e.g., features such as *has four legs*, has *a tail*, *barks* etc. may be combined into the concept *dog*, and compared to other concepts), simulated accounts argue that a concept is represented offline (i.e., in absence of the referent) through the (partial) reactivation of the brain areas involved in processing online perceptual experience with its referent. For example, the representation of *dog* might consist of the reactivation of visual (e.g., the colour of its coat) auditory (e.g., the sound of its bark), haptic (e.g., the touch of its wet nose), as well as action (e.g., playing fetch) experience.

Evidence for a simulated view of cognition comes from behavioural experiments that show a strong relationship between perceptual and conceptual processing across various domains, such as the effects of incongruent perceptual information on sentence and word processing or motor action, and vice versa. Zwaan and Taylor (2006) found that participants were faster to perform a motor action (e.g., turning a dial left or right to indicate the colour change of a rotating cross) when that action was congruent with visual information (i.e., rotation direction of the stimulus) and with sentences describing an action. Similarly, Zwaan et al. (2002) found that participants were faster to verify a pictured object (e.g., *eagle*) was mentioned in a preceding sentence (e.g., *the ranger saw the eagle in the sky*) when the shape of that object matched the state described in the sentence (e.g., *eagle with closed or outstretched wings*). Other research has shown that descriptions of motion (upwards or downwards) affect how objects in ambiguous images are perceived (Dils & Boroditsky, 2010).

In addition to visual information, research has also explored other sensory modalities. For example, Connell and Lynott (2010) found that the perceptual disadvantage people have in processing of tactile information (e.g., people are slower to detect vibrations than flashing lights or noise; (Spence et al., 2001) extends to conceptual processing, participants were slower to judge whether rapidly flashed words corresponding to the various sensory modalities (auditory, gustatory, tactile, olfactory, haptic and visual) when those words belonged to the tactile modality (e.g., itchy, stinging). Moreover, Pecher et al. (2003, 2004) found that verifying modality-specific sentences incurs a processing cost when switching from one modality to another (e.g., verifying the auditory sentence *leaves are rustling* followed by the gustatory sentence *lemons are tart*). Additional evidence comes from neuroimaging studies, which show shared activation between areas involved in processing perceptual experience and conceptual processing (Aziz-Zadeh et al., 2006; Carota et al., 2012; Goldberg et al., 2006; Hauk et al., 2004). Further evidence comes from patient studies (e.g., Boulenger et al., 2008; Fernandino et al., 2013), showing selective impairment of action-verb processing in patients with motor disorders (Parkinson's disease). Taken together, these findings suggest a strong link between conceptual and sensorimotor processing.

Meteyard et al. (2012) describe the varying views on (simulated) conceptual processing as a continuum, ranging from unembodied, to secondary, weak and strong embodiment. The simulated accounts described above fall mostly in the category of weak or strong embodiment. They hold that sensorimotor information may (partially) represent concepts in semantic memory, and that sensorimotor activation as observed in neuroimaging studies is meaningful. By contrast, the accounts that Meteyard et al. categorise under secondary embodiment argue that concepts may draw upon, but are not directly represented by, sensorimotor information. These accounts argue concepts are amodal, and bear only an indirect relation to perceptual and motor information, for example, because they form the basis from which abstracted amodal concepts are formed (e.g., Rogers & Patterson, 2007).

Some secondary embodiment accounts have dismissed the neuroimaging evidence that weak and strong embodiment accounts take to prove concepts may partially be represented through sensorimotor information. They have argued that motor resonance (e.g., the activation of motor areas involved in performing actions when reading sentences involving those actions) does not exclude the activation of amodal representations (Mahon & Caramazza, 2008). Mahon and Caramazza argue that the fact that fast, automatic and somatotopic activation is observed does not conclusively rule out alternative forms of conceptual representation. Instead, much like the cascading information activation of phonological information in object naming, sensorimotor information may be activated following the activation of an amodal concept.

Neuroimaging studies typically offer correlational rather than causal evidence and may as such not be able to rule out the explanation that sensorimotor activation is *downstream* activation, cascading from the activation of an amodal representation. However, there is more evidence that contradicts the criticism from secondary embodiment accounts. For example, patient studies (Boulenger et al., 2008), which *do* show a causal link between impairment of the motor system and the processing of action verbs (but not control items). Mahon and Caramazza (2008) note that Boulenger et al. (2008) did not test any other verbs. However, further evidence for the link between sensorimotor information and conceptual processing comes from studies on aphasia patients with reduced grey matter density, affecting the auditory cortex, which showed a selective impairment in processing sound-related words only (Bonner & Grossman, 2012). If conceptual processing is amodal, and activation of sensorimotor information is epiphenomenal, what explains these selective impairments? In addition to evidence from patient studies, there is also behavioural evidence which contradicts the epiphenomenal view. Connell et al. (2012) performed an experiment in which participants judged the size of manipulable objects (e.g., *wallet*, *key*) while tactile stimulation was applied to their hands or feet. When participants' hands were stimulated, their ability to judge the size of the objects improved. However, this effect did not occur to the same extent when participants received stimulation of the feet, nor did it occur for non-manipulable objects. Connell et al. argue that this effect cannot be explained by referring to the prior activation of an amodal symbol.

In summary, simulated accounts of cognition argue that concepts are inherently grounded in perceptual and action experience of the world. They contrast with their amodal predecessors in that they do not assume an arbitrary relationship between concepts and the perceptual information on which they are based. While there is considerable empirical evidence for a simulated account of cognition, that does not mean this view is without criticism. One apparent weakness of strong accounts of sensorimotor simulation is that some of the effects typically associated with simulated conceptual processing may also be explained by language.

1.3.3. Sensorimotor-linguistic accounts of concepts

Strong simulated accounts contrast with linguistic distributional accounts of conceptual processing. This is not surprising, as a crucial aspect of simulated accounts was the need to break away from symbols that bear no relationship to perceptual experience as the basis for conceptual representations. As Glenberg and Robertson (2000) argue by referring to the Chinese Room argument (Searle, 1981), models such as LSA are not actually representing knowledge: they are translating Chinese into Chinese without ever knowing what it is they are translating. Instead, Glenberg and Robertson argue that words are inherently mapped onto perceptual simulations, derived from interaction with the environment. To illustrate this, they demonstrated that LSA did not match participant's performance in judging sensibility of various sentences containing substitutes for everyday objects (e.g., filling up an old sweater with leaves vs. filling it with water to replace a pillow).

The work by Glenberg and Robertson illustrates the way early simulated accounts regarded the role of language in conceptual processing. However, language has not been fully discounted as a source of information underpinning concepts. Recently, researchers have suggested that while concepts are principally grounded in sensorimotor experience, they may not need to be fully engaged every time (Barsalou et al., 2008; Connell, 2018; Connell & Lynott, 2014; Louwerse, 2011). Indeed, a number of studies have illustrated that the linguistic distributional relations between words help to explain behaviour that was previously thought to be exclusively indicative of simulated conceptual representations. For example, Louwerse and Connell (2011) showed that modality switching effects, or the delays in verifying a sentence that has an incongruent perceptual modality compared to the one that preceded it (Pecher et al., 2003), may also be predicted from language statistics. Crucially, Louwerse and Connell noted that linguistic distributional information best explained the shorter RTs, whereas simulated information accounted mostly for the longer RTs. Furthermore, Connell and Lynott (2013) found that linguistic distributional information may inform rapid/shallow sensibility judgments of novel conceptual combinations better than interpretation judgments (which require slower and deeper processing). Other research has shown linguistic distributional information may inform semantic relatedness and spatial iconicity judgments (Louwerse & Hutchinson, 2012; Louwerse & Jeuniaux, 2010). as well as SNARC effects (Hutchinson & Louwerse, 2014).

Overall, sensorimotor-linguistic accounts that argue the conceptual system comprises both simulated and linguistic information. Several sensorimotor linguistic accounts exist and are differentiated by the levels of relative importance that they assign to each type of information. For example, language as situated simulation (LASS; Barsalou et al., 2008) assigns language a superficial role compared to simulations. They argue that linguistic processing may precede simulations in time depending on task constraints (e.g., when processing linguistic stimuli), and that linguistic processing may support superficial conceptual tasks (e.g., lexical decision making), but that deeper processing relies more heavily on simulation (deviating from previous two-system accounts such as Dual Coding Theory; Paivio, 1971). Throughout conceptual processing simulated, and linguistic information interact. A simulated concept may activate its associated label, which activates other linguistic forms, which in turn activate simulations. Vigliocco et al. (2009) expand on this and argue that once language has been grounded in sensorimotor and affective experience, as well as experience of our inner states, it may serve as a rich source from which meaning can be learned, especially for abstract concepts.

By contrast, the Symbol Interdependency Hypothesis (SIH; Louwerse, 2011) assigns a much stronger role to language. Louwerse argues that the idea that language meaning is always represented through the activation of perceptual information is unrealistic. In his view, the communicative function of language would be impossible to fulfil if every word were to activate a perceptual representation. Instead, SIH holds that the meaning of words may be represented through sensorimotor simulations but may also be bootstrapped from the relationships between words in language, which is thought to encode sensorimotor information. For example, Louwerse (2008) showed that word-order frequency predicted participants' response times in a spatial iconicity judgment task (e.g., judging whether the word *attic* positioned above the word *basement* is more iconic than the word *basement* positioned above the word attic). In this example, word-order encodes the perceptual information that *attics* appear above *basements*, with *attic-basement* being more frequent in language than *basement-attic*. Consequently, the SIH suggests that the relationships between words in language (e.g., word-order frequency) can provide meaning, without the need for simulated representations to be activated every time.

The Linguistic Shortcut Hypothesis (LSH; Connell, 2018; Connell & Lynott, 2014) builds on these theories. Like LASS, it assumes that sensorimotor and linguistic distributional information are part of a dynamic system. Furthermore, like SIH, it assumes that conceptual processing does not always require sensorimotor simulation. However, where both LASS and SIH work from the implicit assumption that sensorimotor and linguistic information are two sides of the same coin (i.e., they are multiple ways of representing the same information about the world around us), LSH assumes that the statistical distribution of words in language allows for the extraction of associative information that goes beyond mirroring the relative distribution of objects in the real world. That is, where simulations are limited to reflect perceptual, motor and affective experience, linguistic distributional information might consider the larger linguistic context in which a concept appears. Language may describe things that have no sensorimotor correlate. For example, when thinking of complex relations between abstract concepts such as *democracy* and *human rights,* linguistic distributional information may be a richer information source (see also Andrews et al., 2009). Crucially, where LASS and SIH assign a dominant role to either sensorimotor (LASS) or linguistic (SIH) information, LSH does not a priori favour one information type over the other. Rather, it holds that the relative importance of linguistic distributional versus sensorimotor simulation in each task response depends on a variety of factors (e.g., task goals, available resources, participant motivations, etcetera).

Central to LSH is the notion of different processing speeds for linguistic distributional information and sensorimotor information. Connell argues that while linguistic distributional information is typically argued to peak in activation prior to simulation information (Barsalou et al., 2008; Louwerse & Hutchinson, 2012), the real advantage lies in the relative importance of linguistic information at the point of response to a stimulus. For certain tasks, linguistic associations can serve as a shorthand to conceptual decision making, whereby it is not simply the fastest in terms of activation, but the fastest in terms of producing a usable outcome, for example in property verification. Moreover, in addition to being faster to produce a usable outcome, linguistic distributional information can also aid in determining whether a given conceptual task is worth spending additional cognitive resources. For example, when judging sensibility of novel compound nouns, linguistic distributional information may help to flag up when a stimulus (e.g., *onion bus*) is nonsensical and processing should stop, pending more information (Connell & Lynott, 2013).

Efforts to model semantic memory, which have run in parallel to the development of sensorimotor-linguistic theories of conceptual representation, have also stressed the multimodal nature of concepts (e.g., Rogers & McClelland, 2004). Therefore, some researchers have made attempts to expand existing connectionist models of semantic memory with linguistic distributional knowledge (e.g., Hoffman et al., 2018) or to develop distributional models that incorporate both sensorimotor and linguistic distributional count-vector information (Andrews et al., 2009). Such models provide a useful illustration of how a conceptual system based on sensorimotor and linguistic distributional information might process and store semantic content. However, both approaches rely on discrete and binary features as a substitute for sensorimotor knowledge, which may be derived from feature norms. Such norms do not yield grounded (i.e., perceptual or action) features for every concept, and tend to describe a limited set of concepts (Banks, et al., 2021) which necessarily limits feature-based models in their application. Similarly, in the case of recent connectionist approaches, the linguistic distributional component consisted of a limited set of verbal descriptors, which may provide a plausible process model, but may not fully capture linguistic distributional knowledge.

1.4. A sensorimotor-linguistic account of taxonomic and typicality effects/processes

Crucially, traditional explanations for behavioural effects in object categorisation (e.g., taxonomic, typicality) are closely tied to the development of (feature-based) accounts of conceptual representation. Network-models, which became a significant attempt at determining the structure of semantic memory, were initially framed as an explanation for the finding that participants are faster at categorising instances as members of a near-superordinate than a far-superordinate category (e.g., faster at categorising *Labradors* as *dogs* than as *animals*). Prototype theory was a response to a range of findings that showed distinct asymmetries in the way participants rated exemplariness, or how typical a given instance is of its category, and the fact that existing views (i.e., definitional accounts) could not explain such behaviour. Exemplar theory relied on behavioural and modelling work that showed that for specific categorisation tasks, remembering the idiosyncratic characteristics of exemplars was a viable strategy for categorising new objects that in some contexts outperformed abstract summary representation. In doing so, each of these theories developed distinct ideas about the underlying structure and nature of concepts.

By contrast, current views on conceptual processing have largely developed away from explanations of object categorisation behaviour, and as such have yet to develop a formal explanation for key behavioural effects in categorisation. It is the aim of this thesis to explore such an explanation. That is, if concepts indeed comprise sensorimotor and linguistic distributional information, where does that leave categories? By what mechanism are objects grouped together? What represents these categories in semantic memory? As outlined above, these mechanisms are typically observed indirectly, by attempting to predict categorisation *behaviour*, for example by accounting for taxonomic and typicality effects.

To my knowledge, no comprehensive simulated-linguistic account of categorisation behaviour exists, However, Barsalou (1999) outlines how simulated information might underpin categorisation. He argues that categorisation might be based around comparing a perceived instance (e.g., *dog*) to a simulated representation comprised of perceptual symbol systems. Barsalou argues this system is dynamic, as it may be updated when novel instances are successfully categorised. Expanding on this, Barsalou (2003a) argues that people may use perceptual simulations abstracted from experience with members to represent category concepts. He argues that such abstractions might contain simulations of features and their relationships. For example, for the category *tables*, people might form perceptual simulations of *legs*, which might comprise information of their experience with tables and their legs on all sensory modalities. In addition to this, people might represent relationships, such as spatial relationships between features (e.g., a table's *legs* are usually below the *tabletop*). Such features are not represented in isolation, but in the company of others, or even as part of a whole. Crucially, Barsalou argues, simulated information is used to generalise across category members on the fly and is not stored in a static summary representation (although abstractions may be reactivated over time).

Similarly, Connell and Lynott (2014) argue that a concept (e.g., *dog*) may arise through multiple experiences with instances of *dog and* may change over time because of repeated experience with dogs. As Barsalou illustrates, categorical inferring may build on the wide range of multimodal information available to the simulated representation, for example our sensorimotor experience with *dogs* allows us to approach a previously unseen instance of *dog* in a particular way (*careful, he may bite!*), guiding our actions based on our existing knowledge of *dogs*. Unlike traditional feature-based accounts, the smallest unit of representation is not an amodal feature (e.g., <u>*has wings*</u>), but multimodal perceptual experience with previous instances belonging to a concept. As such, it circumvents the issues faced by definitional (necessary features), prototype (constraints on similarity, constraints on what counts as a feature), exemplar (constraints on similarity within and between exemplars), and theory-theory (constraints on what constitutes a theory).

The role of linguistic distributional information in categorisation has also not extensively been documented. In a recent study, Banks et al. (2021) showed that the overlap in sensorimotor experience and the linguistic context in which category and member words appear independently affected the frequency and mean rank with which members are named for a given category. The more a given member label appeared in the same context as the category label (measured in a large corpus of subtitles), the more likely it was that people listed it, and the earlier they did so. Moreover, a computational model based on this theory was an excellent fit to human response patterns. Banks et al., took this as evidence for the linguistic shortcut hypothesis (see above). Other evidence for the role of linguistic distributional information in categorisation comes from Connell and Ramscar (2001), who illustrated that the LSA model of language co-occurrence successfully predicted object typicality. Since object typicality forms the basis of prototype theory, this finding lends further support for the notion that distributional relationships between words in language may aid categorisation.

A simulated-linguistic account might build on the role of sensorimotor simulations in categorisation as outlined by Barsalou (1999, 2003a) and Connell and Lynott (2014) and integrate them with linguistic distributional information, following recent views on conceptual processing. On this view, both category and member concepts comprise sensorimotor and linguistic distributional information, and their relationship may be expressed in terms of the degree of overlap between representational information of the member and category, as illustrated by Banks et al. (2021). Crucially, task demands may dictate the extent to which people rely on one form of information over the other. For example, when categorising an object under time constraints, participants may rely more on the computationally cheap and fast linguistic distributional information (i.e., the Linguistic

Shortcut Hypothesis; Connell, 2018). The goal of the present thesis is to determine the extent to which a simulated-linguistic account of categorisation is possible, and the degree to which the relative reliance on simulated and linguistic distributional information may vary depending on task constraints.

1.5. Thesis roadmap

Traditional explanations of behavioural effects in object categorisation (e.g., taxonomic and typicality effects) assume that categories are represented in semantic memory as feature-summaries or lists of exemplars, and that categorisation is the process of deriving feature-similarity between an instance and either of these representations (Murphy & Brownell, 1985; Rosch, 1978). By contrast, recent accounts of conceptual processing argue that concepts comprise (partial) reactivations of sensorimotor experience (i.e., perception-action and affective experience of the world) as well as linguistic distributional knowledge (Barsalou et al., 2008; Louwerse, 2011; Connell & Lynott, 2014). Crucially, however, where feature-based accounts of conceptual representation were heavily tied to behavioural effects in object categorisation, no explicit sensorimotor-linguistic account of categorisation has been formulated. Therefore, the aim of this thesis is to investigate key effects in object categorisation, most notably the effects of taxonomic (i.e., basic-level advantage) and to a lesser degree object typicality on performance in category verification from a sensorimotor-linguistic perspective on conceptual representation.

To this end, Chapter 2 describes results from a label-picture category-verification experiment, whereby response times (RT) and accuracy were recorded and modelled using predictors of overlap in sensorimotor (i.e., perception-action experience of the world) and linguistic distributional (i.e., knowledge about the statistical distribution of words in language) information between category and member concepts. This chapter also includes additional analyses of RT and accuracy data using adjusted measures of sensorimotor and linguistic distributional overlap, intended to capture the interaction between taxonomic level and object typicality (e.g., Murphy & Brownell, 1985; Jolicoeur et al., 1984), and their comparison to traditional measures of object typicality. Chapter 3 describes results from a forced-choice label-label categorisation task, whereby participants chose between two correct alternatives for a given category label. This study primarily aimed to contrast predictions from traditional accounts which attribute a special status to the basic level with predictions from a sensorimotor-linguistic account of categorisation, which holds that categorisation performance is improved for concepts that are closer in sensorimotor experience and linguistic distributional knowledge. Chapter 4 addresses a weakness of the experiments described in Chapter 2, namely the fact that the implicit name participants gave to an image had to be assumed. Therefore, Chapter 4 describes the collection and validation of a large normed set of photographs and their associated names. Finally, Chapter 5 describes a rapid-categorisation study, which aimed to expand on the findings of Chapter 2, using the normed dataset from Chapter 4, by testing the task-dependent reliance on either linguistic-distributional or sensorimotor information that sensorimotor-linguistic accounts of conceptual representation have argued for. To this end, the study reported in Chapter 5 limited the amount of perceptual processing available to participants by rapidly displaying an image followed by a perceptual mask at various intervals (ultra-rapid categorisation), with the aim of testing whether the effect of overlap in sensorimotor experience between category and member concepts would be reduced relative to the effect of overlap in linguistic distributional overlap when less time was available for perceptual processing. Together, these Chapters report a series of novel experiments that test whether modern sensorimotor-linguistic models can explain traditional categorisation behavioural effects, and to what extent.

1.6. References

Aerts, D., Broekaert, J., Gabora, L., & Sozzo, S. (2016). Generalizing Prototype Theory: A Formal Quantum Framework. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00418

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. https://doi.org/10.1037/a0016261

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*(3), 263–308. https://doi.org/10.1016/0010-0277(83)90012-4

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

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469. https://doi.org/10.1016/j.visres.2005.01.004

Banks, B., Wingfield, C., & Connell, L. (2021). Linguistic Distributional Knowledge and Sensorimotor Grounding both Contribute to Semantic Category Production. *Cognitive Science*, *45:* e13055. https://doi.org/10.1111/cogs.13055.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. Context-predicting semantic vectors. In K. Toutanova & H. Wu (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (pp. 238–247). Association for Computational Linguistics. doi:0.3115/v1/P14-1023

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*(1), 82–93. https://doi.org/10.3758/BF03197629

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227. https://doi.org/10.3758/BF03196968

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629–654. https://doi.org/10.1037/0278-7393.11.1-4.629

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

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1435), 1177–1187. https://doi.org/10.1098/rstb.2003.1319

Barsalou, L. W., Santos, A., Simmons, K. W., & Wilson, C. D. (2008). Language and Simulation in Conceptual Processing. In M. de Vega, A. M. Glenberg, & A. Graesser (Eds.), *Symbols, Embodiment and meaning* (pp. 245–283). Oxford University Press.

Bohn, R., & Short, J. (2009). *How Much Information?* University of California.

Bonner, M. F., & Grossman, M. (2012). Gray Matter Density of Auditory Association Cortex Relates to Knowledge of Sound Concepts in Primary Progressive Aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991. https://doi.org/10.1523/JNEUROSCI.6241-11.2012

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

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Brooks, L. (1978). Nonanalytic concept formation. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Lawrence Erlbaum Associates, Publishers.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526. https://doi.org/10.3758/BF03193020

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instruments, & Computers*, *30*(2), 188–198. https://doi.org/10.3758/BF03200643

Burgess, C., & Lund, K. (1997). Representing abstract words and emotional connotation in a high-dimensional memory space. In *Proceedings of the Cognitive Science Society* (pp. 61–66). Erlbaum.

Carota, F., Moseley, R., & Pulvermüller, F. (2012). Body-part-specific Representations of Semantic Noun Categories. *Journal of Cognitive Neuroscience*, *24*(6), 1492–1509. https://doi.org/10.1162/jocn_a_00219

Clarke, A., & Tyler, L. K. (2015). Understanding What We See: How We Derive Meaning From Vision. *Trends in Cognitive Sciences*, *19*(11), 677–687. https://doi.org/10.1016/j.tics.2015.08.008

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

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. (2010). Look but don't touch: Tactile disadvantage in processing modality-specific words. *Cognition*, *115*(1), 1–9. https://doi.org/10.1016/j.cognition.2009.10.005

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & 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

Connell, L., & Ramscar, M. (2001). Using Distributional Measures to Model Typicality in Categorization. In *Proceedings of the Twenty-Third Annual Meeting of the Cognitive Science Society* (pp. 226–231). Lawrence Erlbaum Associates.

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, *111*(2), 291–303. https://doi.org/10.1037/0033-2909.111.2.291

Crick, F. (1989). The recent excitement about neural networks. *Nature*, *337*(6203), 129–132.

Crutch, S. J., & Warrington, E. K. (2008). Contrasting patterns of comprehension for superordinate, basic-level, and subordinate names in semantic dementia and aphasic stroke patients. *Cognitive Neuropsychology*, *25*(4), 582–600. https://doi.org/10.1080/02643290701862290

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

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, *18*(4), 500–549. https://doi.org/10.1016/0010-0285(86)90008-3

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013). Parkinson's Disease Disrupts Both Automatic and Controlled Processing of Action Verbs. *Brain and Language*, *127*(1), 65–74. https://doi.org/10.1016/j.bandl.2012.07.008

Fodor, J. A. (2008). Modularity of the mind. In J. E. Adler & L. J. Rips (Eds.), *Reasoning. Studies of Human Inference and Its Foundations* (pp. 878–914). Cambridge University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Forde, E. M. E., & Humphreys, G. W. (2002). The Role of Semantic Knowledge in Short-term Memory. *Neurocase*, *8*(1–2), 13–27. https://doi.org/10.1093/neucas/8.1.13

Geeraerts, D. (2016). Prospects and problems of prototype theory. *Diacronia*, *4*, 1–16. https://doi.org/10.17684/i4A53en

Glenberg, A. M., & Robertson, D. A. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, *43*(3), 379–401. https://doi.org/10.1006/jmla.2000.2714

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

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*(2), 125–157. https://doi.org/10.1016/0010-0277(94)90065-5

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-Space Models of Semantic Representation From a Cognitive Perspective: A Discussion of Common Misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Hampton, J. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 441–461. https://doi.org/10.1016/S0022-5371(79)90246-9

Hampton, J. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition*, *9*(2), 149–156. https://doi.org/10.3758/BF03202329

Hampton, J. (1993). Prototype models of concept representation. In I. van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67–95). Academic Press.

Hampton, J. (1995). Testing the Prototype Theory of Concepts. *Journal of Memory and Language*, *34*(5), 686–708. https://doi.org/10.1006/jmla.1995.1031

Hampton, J. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, *65*(2), 137–165. https://doi.org/10.1016/S0010-0277(97)00042-5

Hampton, J. (2007). Typicality, Graded Membership, and Vagueness. *Cognitive Science*, *31*(3), 355–384. https://doi.org/10.1080/15326900701326402

Hampton, J., & Jönsson, M. L. (2012). Typicality and Compositionality: The Logic of Combining Vague Concepts. In M. Werning, W. Hinzen, & E. Machery (Eds.), *The Oxford handbook of compositionality* (pp. 385–402). Oxford University Press.

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

Heit, E., & Barsalou, L. W. (1996). The Instantiation Principle in Natural Categories. *Memory*, *4*(4), 413–452. https://doi.org/10.1080/096582196388915

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: A unique clinicopathological syndrome. *The Lancet Neurology*, *6*(11), 1004–1014. https://doi.org/10.1016/S1474-4422(07)70266-1

Hodges, J. R., Patterson, K., & Tyler, L. K. (1994). Loss of semantic memory: Implications for the modularity of mind. *Cognitive Neuropsychology*, *11*(5), 505–542. https://doi.org/10.1080/02643299408251984

Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, *125*(3), 293. https://doi.org/10.1037/rev0000094

Hofmann, B. (2017). Do health professionals have a prototype concept of disease? The answer is no. *Philosophy, Ethics, and Humanities in Medicine*, *12*(1), 6. https://doi.org/10.1186/s13010-017-0047-7

Humphreys, G. W., & Forde, E. M. E. (2005). Naming a giraffe but not an animal: Base-level but not superordinate naming in a patient with impaired semantics. *Cognitive Neuropsychology*, *22*(5), 539–558. https://doi.org/10.1080/02643290442000176

Jefferies, E., Thompson, H., Cornelissen, P., & Smallwood, J. (2020). The neurocognitive basis of knowledge about object identity and events: Dissociations reflect opposing effects of semantic coherence and control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20190300. https://doi.org/10.1098/rstb.2019.0300

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*(3), 248–277. https://doi.org/10.1037/0096-3445.126.3.248

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243–275. https://doi.org/10.1016/0010-0285(84)90009-4

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, *112*(3), 500–526. https://doi.org/10.1037/0033-2909.112.3.500

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What categories reveal about the mind*. The University of Chicago Press.

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Laurence, S., & Margolis, E. (1999). Concepts and Cognitive Science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core Readings* (pp. 3–81). The MIT Press.

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, *21*(6), 335–346. https://doi.org/10.1038/s41583-020-0277-3

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

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*(1), 28–35. https://doi.org/10.3758/PBR.15.1.28

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*(3), 551–569. https://doi.org/10.1037/a0039184

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), 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

Mandler, J. M., & Bauer, P. J. (1988). The cradle of categorization: Is the basic level basic? *Cognitive Development*, *3*(3), 247–264. https://doi.org/10.1016/0885-2014(88)90011-1

Mandler, J. M., & McDonough, L. (2000). Advancing Downward to the Basic Level. *Journal of Cognition and Development*, *1*(4), 379–403. https://doi.org/10.1207/S15327647JCD0104_02

Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an Integration of Deep Learning and Neuroscience. *Frontiers in Computational Neuroscience*, *10*. https://www.frontiersin.org/article/10.3389/fncom.2016.00094

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 54–70. https://doi.org/10.1037/0278-7393.23.1.54

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310. https://doi.org/10.1038/nrn1076

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, *6*(4), 462–472. https://doi.org/10.3758/BF03197480

McRae, K., Cree, G. S., Seidenberg, M. S., & Mcnorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

Medin, D. L. (1989). Concepts and Conceptual Structure. *American Psychologist*, *44*(12), 1469–1481. https://doi.org/10.1037/0003-066X.44.12.1469

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*(2), 158–190. https://doi.org/10.1016/0010-0285(88)90018-7

Medin, D. L., & Smith, E. E. (1984). Concepts and Concept Formation. *Annual Review of Psychology*, *35*(1), 113–138.

Mervis, C. B., & Crisafi, M. A. (1982). Order of Acquisition of Subordinate-, Basic-, and Superordinate-Level Categories. *Child Development*, *53*(1), 258–266. https://doi.org/10.2307/1129660

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

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv:1301.3781 [Cs]*. http://arxiv.org/abs/1301.3781

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775–799. https://doi.org/10.1037/0278-7393.27.3.775

Murphy, G. L. (1982). Cue validity and levels of categorization. *Psychological Bulletin*, *91*(1), 174–177. https://doi.org/10.1037/0033-2909.91.1.174

Murphy, G. L. (2002). *The Big Book of Concepts*. MIT Press.

Murphy, G. L. (2016). Is there an exemplar theory of concepts? *Psychonomic Bulletin & Review*, *23*(4), 1035–1042. https://doi.org/10.3758/s13423-015-0834-3

Murphy, G. L., & Allopenna, P. D. (1994). The Locus of Knowledge Effects in Concept Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 904–919. https://doi.org/10.1037/0278-7393.20.4.904

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 70–84. https://doi.org/10.1037/0278-7393.11.1.70

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316. https://doi.org/10.1037/0033-295X.92.3.289

Murphy, G. L., & Ross, B. H. (2005). The two faces of typicality in category-based induction. *Cognition*, *95*(2), 175–200. https://doi.org/10.1016/j.cognition.2004.01.009

Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 1–20. https://doi.org/10.1016/S0022-5371(82)90412-1

Murphy, G. L., & Spalding, T. L. (1995). Knowledge, Similarity, and Concept Formation. *Psychologica Belgica*, *35*(2–3), 127–144. https://doi.org/10.5334/pb.882

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 572–586. https://doi.org/10.1037/0278-7393.15.4.572

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M., & Palmeri, T. J. (1997). An Exemplar-Based Random Walk Model of Speeded Classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Stanton, R. D. (2005). Speeded Classification in a Probabilistic Category Structure: Contrasting Exemplar-Retrieval, Decision-Boundary, and Prototype Models. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 608–629. https://doi.org/10.1037/0096-1523.31.3.608

Palmer, S. E. (1977). Hierarchical Structure in Perceptual Representation. *Cognitive Psychology*, *9*(4), 441–474. https://doi.org/10.1016/0010-0285(77)90016-0

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying Different-Modality Properties for Concepts Produces Switching Costs. *Psychological Science*, *14*(2), 119–124. https://doi.org/10.1111/1467-9280.t01-1-01429

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin & Review*, *11*, 164–167. https://doi.org/10.3758/BF03206477

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363. https://doi.org/10.1037/h0025953

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. MIT Press.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407. https://doi.org/10.1016/0010-0285(72)90014-X

Rey, G. (1983). Concepts and stereotypes. *Cognition*, *15*(1–3), 237–262. https://doi.org/10.1016/0010-0277(83)90044-6

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General*, *122*(4), 468–486. https://doi.org/10.1037/0096-3445.122.4.468

Rips, L. J., Shoben, E., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20. https://doi.org/10.1016/S0022-5371(73)80056-8

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. The MIT Press. https://doi.org/10.7551/mitpress/6161.001.0001

Rogers, T. T., & McClelland, J. L. (2008). Précis of Semantic Cognition: A Parallel Distributed Processing Approach. *Behavioral and Brain Sciences*, *31*(6), 689–714. https://doi.org/10.1017/S0140525X0800589X

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition & Categorization*. Lawrence Erlbaum Associates, Publishers.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502. https://doi.org/10.1037/0096-1523.2.4.491

Roth, E. M., & Mervis, C. B. (1983). Fuzzy set theory and class inclusion relations in semantic categories. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 509–525. https://doi.org/10.1016/S0022-5371(83)90310-9

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, *15*(3), 346–378. https://doi.org/10.1016/0010-0285(83)90012-9

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Rumelhart, D. E., McClelland, J. L., & Group, P. R. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. https://doi.org/10.7551/mitpress/5236.001.0001

Rumelhart, D. E., & Norman, D. (1988). Representation in Memory. In R. Atkinson, R. Herrnstein, G. Lindzey, & R. Luce (Eds.), *Steven's handbook of experimental psychology: Vol. 2. Learning and cognition* (p. 137). Wiley.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Siew, C. S. Q., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive Network Science: A Review of Research on Cognition through the Lens of Network

Representations, Processes, and Dynamics. *Complexity*, *2019*, e2108423. https://doi.org/10.1155/2019/2108423

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Harvard University Press.

Smith, E. E., Shoben, E., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214–241. https://doi.org/10.1037/h0036351

Smith, J. D. (2002). Exemplar Theory's Predicted Typicality Gradient Can Be Tested and Disconfirmed. *Psychological Science*, *13*(5), 437–442. https://doi.org/10.1111/1467-9280.00477

Smith, J. D. (2014). Prototypes, exemplars, and the natural history of categorization. *Psychonomic Bulletin & Review*, *21*, 312–331. https://doi.org/10.3758/s13423-013-0506-0

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3–27. https://doi.org/10.1037/0278-7393.26.1.3

Smits, T., Storms, G., Rosseel, Y., & De Boeck, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. *Psychonomic Bulletin & Review*, *9*(4), 836–844. https://doi.org/10.3758/BF03196343

Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*, 330–336.

Storms, G., Boeck, P. D., & Ruts, W. (2000). Prototype and Exemplar-Based Information in Natural Language Categories. *Journal of Memory and Language*, *42*, 51–73.

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Voorspoels, W., Vanpaemel, W., & Storms, G. (2008). Exemplars and prototypes in natural language concepts: A typicality-based evaluation. *Psychonomic Bulletin & Review*, *15*, 630–637. https://doi.org/10.3758/PBR.15.3.630

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, *27*(4), 635–657. https://doi.org/10.1080/14640747508400525

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

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

Wisniewski, E. J., & Medin, D. L. (1994). *On the Interaction of Theory and Data in Concept Learning*. *18*(2), 221–281. https://doi.org/10.1016/0364-0213(94)90002-7

Wittgenstein, L. (1958). *Philosophical Investigations* (2nd ed.). Basil Blackwell Ltd.

Zadeh, L. A. (1996). Fuzzy sets. In G. J. Klir & B. Yuan (Eds.), *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh* (pp. 394–432). World Scientific. https://doi.org/10.1142/9789814261302_0021

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*(1), 81–88. https://doi.org/10.1080/01638539909545070

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., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Chapter 2: The Effects of Sensorimotor and Linguistic information on the Basic-Level Advantage

2.1. Chapter introduction

The basic-level advantage, the general preference and processing advantages for categorisation at an intermediate level of abstraction (e.g., *dog*) is one of the best-known effects in categorisation research. Traditional accounts of categorisation have argued that the basic-level advantage arises because basic-level categories offer a maximally distinct and informative match to an object's features. In the series of studies described in this chapter, I tested whether measures of sensorimotor and linguistic overlap between a category- and member-concept explained differences in categorisation performance (accuracy and response time), to test the hypothesis that overlap in sensorimotor experience and linguistic distributional knowledge form the basis for categorisation. Moreover, I investigated whether taxonomic level interacted with object typicality, such that subordinate-level categorisation was preferred for atypical items, and tested whether sensorimotor and linguistic distributional information capture the graded structure of categories (i.e., typicality gradients).

The Effects of Sensorimotor and Linguistic Information on the Basic-Level Advantage

Rens van Hoef, Louise Connell, and Dermot Lynott

Department of Psychology, Lancaster University

Author Note

Rens van Hoef https://orcid.org/0000-0003-1355- 1541

Louise Connell https://orcid.org/0000-0002-5291-5267

Dermot Lynott https://orcid.org/0000-0001-7338-0567

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 682848) to LC.

Correspondence concerning this article should be addressed to Rens van Hoef, Department of Psychology, Fylde College, Lancaster University, Bailrigg, Lancaster, LA1 4YF, UK. Email: r.vanhoef@lancaster.ac.uk

2.2. Abstract

The basic-level advantage is one of the best-known effects in human categorisation. Traditional accounts argue that basic- level categories present a maximally informative or entry-level into a taxonomic organisation of concepts in semantic memory. However, these explanations are not fully compatible with most recent views on the structure of the conceptual system such as linguistic-simulation accounts, which emphasise the dual role of sensorimotor (i.e., perception- action experience of the world) and linguistic information (i.e., statistical distribution of words in language) in conceptual processing. In three pre-registered word→ picture categorisation studies, we hypothesised that our novel measures of sensorimotor and linguistic distance would contribute to categorical decision making, would map onto the graded structure of categories and would outperform traditional taxonomic levels (i.e., subordinate, basic, superordinate) in predicting the basic-level advantage. Results showed that, overall, our measures predicted the basic-level advantage at least as well as taxonomic level. Sensorimotor information best explained processing speed, whereas taxonomic level best explained participant’s choices.

2.3. Introduction

Categorisation is critical to our everyday cognitive functioning. Representative categories and concepts[4] allow us to adequately perceive, think about, perform actions with and speak about our day-to-day experience (Lakoff, 1987). Without categories, we would have to treat every object, action or event as a unique instance, rendering us overwhelmed and low on cognitive resources (E. E. Smith & Medin, 1981). Instead, the use of categories allows us to organise the environment and the objects encountered within it into groups we judge to be meaningfully similar, thus enabling us to infer knowledge and potential actions, even if we have never encountered a particular instance before. While the fundamental importance of categorisation to our cognitive abilities is evident, the precise definition of categories and how categorical information is cognitively structured remains under debate.

In traditional, feature-based accounts of categorisation and conceptual structure, natural categories are classes that group entities together according to their shared features or properties. While feature-based theories differ in their details, they generally agree that concepts comprise discrete, binary features (e.g., a concept either has, or has not, the feature <u>*can fly*</u>), and that categorisation is possible because certain features occur together more frequently than others (e.g., if it <u>*has wings*</u>, <u>*lays eggs*</u> and <u>*can fly*</u>, it is likely a member of the category *bird*; (Cree & McRae, 2003; Hampton, 1993; Malt & Smith, 1984; Posner & Keele, 1968; Rogers & Patterson, 2007; Rosch, 1973; E.E. Smith et al., 1974; Tyler et al., 2000). In one popular view, categories are stored in semantic memory through an abstracted summary of how features are shared by category members (Posner & Keele, 1968; Rogers & Patterson, 2007; Rosch & Mervis, 1975).

[4] In the present paper, we use the term *concept* to refer to an aggregated, canonical aspect of experience that can be mentally represented offline (i.e., in the absence of its referent: Connell & Lynott, 2014), and the term *category* to refer to a class that groups entities together.

Any given concept may be categorised at multiple, inclusive levels of abstraction (e.g., that small brown creature may simultaneously be categorised as *house sparrow*, *sparrow*, *bird*, and *animal*), reflecting a taxonomy-like hierarchy from very specific lower levels to very abstract higher levels. Crucially, while any given concept may be categorised at any taxonomic level, the basic level (e.g., *bird*) is generally privileged (Rosch, 1978; Rosch, Mervis, et al., 1976). First demonstrated in an extensive series of experiments by Rosch et al. (1976), objects are categorised faster and more accurately when they are preceded by their name at the basic level (e.g., *dog*; *bird*), compared to superordinate (e.g., *animal*) or subordinate (e.g., *Labrador*, *sparrow*) level names. This basic-level advantage is one of the most fundamental effects in categorisation research, and has repeatedly been replicated in later work on potential moderating factors, such as object typicality (Jolicoeur et al., 1984; Murphy & Brownell, 1985), context (Murphy & Wisniewski, 1989), subject expertise (K. E. Johnson & Mervis, 1997; Tanaka & Taylor, 1991) and neurological disorders (Rogers & Patterson, 2007).

While the basic-level advantage is a robust effect, the mechanisms underlying it have not been conclusively explained. Traditional hierarchical accounts argue that taxonomic structure is integral to how categorical knowledge is represented in semantic memory, where feature information is stored only once, at the highest possible level, and generalised to all subordinate levels (Collins & Loftus, 1975), thus avoiding redundancy. Jolicoeur et al. (1984) argue that objects are most easily categorised at the basic level because it is the usual level at which the taxonomic structure is entered, and so categorisation at a level different to the entry level incurs a cost in response time and/ or accuracy. Other feature-based accounts do not assume that semantic memory is structured hierarchically, but rather that a taxonomy is implicit in how features are interrelated. The differentiation account (Markman & Wisniewski, 1997; Murphy & Brownell, 1985; Murphy & Lassaline, 1997; Murphy & Smith, 1982), argues that basic-level categories are quite distinct from contrasting categories, (e.g., *dogs* and *birds* share few features) while also being quite informative in how they group concepts together (e.g., *Labradors* and *collies* share many features). An object can therefore be categorised most quickly and accurately at the basic level because it provides the maximally distinctive and informative match to the object's features, whereas other taxonomic levels are disadvantaged because they match few features of the object (i.e., superordinate categories are distinctive without being informative) or there are several competitors that match some of the object's features (i.e., subordinate categories are informative without being distinct). That is, the basic level is implicitly advantaged in how it best matches features of the object to be categorised (see also Rogers & Patterson, 2007).

Since the advent of traditional, feature-based views of concepts and categorisation, an alternative view of the conceptual system has emerged that may offer a different explanation for the processing advantages in categorisation. Linguistic-simulation accounts of the conceptual system emphasise the importance of both sensorimotor and language experience in conceptual processing (Barsalou et al., 2008; Connell, 2018; Connell & Lynott, 2014; Louwerse, 2011). Both simulated and linguistic distributional information are essential to the operation of the conceptual system, but they interact flexibly to allow reliance on one form of information over another, depending on the exact context or cognitive task (Connell, 2018). Simulated representations emerge from sensorimotor experience with our environment, whereby the neural activations across brain areas involved in processing this experience are represented as partial replays upon retrieval (Barsalou, 1999). These comprise perceptual, motor, affective and other information in direct and vicarious experience (e.g., the concept *dog* might be represented by its smell, the sound it makes when it barks, the touch of its fur etc.). Evidence for the role of simulated representations comes from neuroimaging studies, showing shared activation between areas involved in perceptual experience and their equivalent in conceptual processing (Aziz-Zadeh et al., 2006; Carota et al., 2012; Goldberg et al., 2006; Hauk et al., 2004), as well as from behavioural studies that reveal intricate relationships between perceptual and conceptual processing (Connell & Lynott, 2010; Dils & Boroditsky, 2010; Zwaan & Taylor, 2006). Linguistic distributional knowledge, meanwhile, reflects our vast experience with language, where our sensitivity to statistical properties (Aslin & Newport, 2012; Landauer & Dumais, 1997; Lund & Burgess, 1996) has allowed us to develop knowledge of how words and phrases have specific patterns in their distribution relative to each other (see Wingfield & Connell, 2019, for a review). Certain words occur in the same or similar contexts more often than others (e.g., the contexts in which people mention *dog* and *animal* are more alike than those of *dog* and *cup*), and such linguistic distributional information has been shown to be powerful enough to predict conceptual processing in a wide range of tasks (Connell, 2018; Connell & Lynott, 2013; Goodhew et al., 2014; Louwerse, 2011).

If concepts are indeed represented as a combination of sensorimotor simulation and linguistic distributional knowledge, it follows that such sensorimotor and/or linguistic information may also underlie categorisation. Category membership may be a product of representational similarity between a category concept (e.g., *dog*) and a potential member concept (e.g., *Labrador*), based on sensorimotor experience of the referent concepts and linguistic experience of the concept labels across language. In sensorimotor terms, many feature-based theories emphasise that categorical distinctions emerge at least in part from commonalities in the way we perceive and interact with the word around us (Cree & McRae, 2003; Tyler et al., 2000). However, sensorimotor experience may also be considered as the extent to which a concept is experienced via each perceptual modality or action effector (i.e., sensorimotor strength: Lynott et al., 2020), where the overlap in sensorimotor experience between a category concept (e.g., *dog*) and a member concept (e.g., *Labrador*) predicts how readily people name the member as an example of that category (Banks et al., 2021). In linguistic distributional terms, the relationship between member-concept labels and category-concept labels in corpus-derived linguistic space is also an effective predictor of category membership (Connell & Ramscar, 2001; Riordan & Jones, 2011; Wingfield & Connell, 2019). When a category label (e.g., *dog*) appears in very similar context to a member concept label (e.g., *Labrador*), people tend to judge the member concept as an excellent example of its category (i.e., graded structure of concepts, Connell & Ramscar, 2001).

Compared to traditional accounts of categories and concepts that rely on discrete features to compute and explain advantage effects in categorisation, linguistic-simulation theories take a very different approach in some key respects. Firstly, while traditional hierarchical views assume that semantic memory is structured into discrete taxonomic levels, with the basic level accorded a preferential status (e.g., Jolicoeur et al., 1984), linguistic-simulation accounts do not share that assumption. Rather, although linguistic-simulation accounts have not yet addressed the basic-level advantage directly, they treat all category concepts (i.e., *Labrador*, *dog* and *animal*) with equal status, with no assumption of hierarchy nor preference for the basic level, and assume that task goals and available resources will determine which concept is activated first (Connell & Lynott, 2014). In lack of an explicit *a priori* basic-level preference, linguistic-simulation views are similar to feature-based differentiation accounts (e.g., Murphy & Brownell, 1985; Rogers & Patterson, 2007) but differ in that they do not assume the basic level is implicitly advantaged by optimally distinct and informative binary features. Secondly, where feature-based views draw a distinction between concepts and features, linguistic-simulation views do not. In these views, there is no *a priori* difference between *tail, barks* and *dog*. They may be related, in that a sensorimotor simulation of *dog* involves the activation of the concepts *tail* and *barks*, or that the word label *dog* occurs frequently in the same or similar linguistic contexts as *tail* and *barks* (which enables the extraction of semantic relations; Wingfield & Connell, 2019), but feature-concepts are not qualitatively different or subsidiary representations to object-concepts. Whether overlaps in sensorimotor and linguistic distributional experience between member concepts and category concepts can contribute to the basic-level advantage – without a fixed taxonomy or subsidiary features – is the subject of the present paper.

The current study

In this paper, we report three pre-registered experiments, one exploratory study and two exploratory analyses. All three studies use a label → picture categorisation task similar to that used by Rosch et al. (1976), where participants judged (yes/no) whether the pictured item belonged to the category named in the preceding label. We investigated whether categorisation performance (response time, accuracy) can be predicted by sensorimotor (i.e., perception-action experience of the world) and linguistic distributional information (i.e., statistical distribution of words in language). Critically, we used a novel measure of sensorimotor information that was fully grounded in perceptual and action experience alone (i.e., without the use of abstracted features), based on multidimensional ratings of sensorimotor strength from Lynott and colleagues (2020).[5] Our measure of linguistic distributional information was derived from co-occurrence frequencies in a large corpus of English. Together, these measures allowed us to distinguish whether representational similarity between a category and member concept was due to overlap in sensorimotor information (e.g., the concepts *animal* and *dog* both involve similar perception and action experience) or linguistic distributional information (e.g., the words *animal* and *dog* both appear in similar contexts across language).

[5] The present experiments were developed in parallel with a separate investigation using the same measure of sensorimotor overlap between concepts (Banks et al., 2021); since both studies used this new measure at the same time, both reports can legitimately describe its use as novel.

We expected that sensorimotor and linguistic information would contribute to categorical decision making, and that the latter in particular would contribute to the basic level advantage in categorisation. That is, we expected people to categorise pictured objects more quickly and accurately when the member concept (e.g., *dog*) was close to the category concept (e.g., *animal*) in both sensorimotor experience and linguistic distributional knowledge.

2.4. Experiment 1: The Basic-Level Advantage

In our first study (pre-registration, data, analysis code and results available at https://osf.io/vdka2/?view_only=ac6c78793a354a1f95da6c36b5ac6163), we examined the basic-level advantage in a classic label → picture category verification task. Participants first saw a category name at one of various levels of specificity (e.g., general *animal*, basic *dog* or specific *Labrador*), followed by a picture (e.g., photograph of a *Labrador*), and their task was to decide yes/no whether the pictured item belonged to the specified category. We expected responses to be faster and more accurate for basic-level category labels (i.e., the basic-level advantage), and aimed to contrast two competing explanations.

If traditional accounts of the basic-level advantage are correct, then the effect emerges from explicit or implicit levels of categorical representations in a taxonomic hierarchy (i.e., subordinate, basic, superordinate). Within this hierarchy, the basic level of *dog* is either the usual point of entry into the taxonomic structure of semantic memory (where accessing other levels of *Labrador* or *animal* incurs a processing cost: Glass & Holyoak, 1974; Jolicoeur et al., 1984) or is the category best differentiated by the pictured object's features (where *animal* matches few features and *Labrador* has too many competitors that match many features: (Markman & Wisniewski, 1997; Murphy & Brownell, 1985; Rogers & Patterson, 2007). As a result, categorical decisions are easier to make at the basic level.

By contrast, we hypothesised that the basic-level advantage emerges from representational overlap of linguistic and sensorimotor information between a category and member concept. Extensive research on picture naming has shown that when participants see a picture of a dog (e.g., a *Labrador, poodle* or *collie*), they label it with the most frequent and earliest-acquired name: *dog* (e.g., Bates et al., 2003; Belke et al., 2005). Since the most frequent, earliest-acquired name tends to be the basic-level label (Rosch, Mervis, et al., 1976), we therefore expected categorisation performance to be fastest and most accurate when the basic-level category label was presented before the picture (e.g., *dog* → [picture of a *dog*]). That is, the match between the presented category label (*dog*) and the name automatically activated for the pictured object (*dog*) facilitates fast and accurate category verification. Simultaneously, the overlap between the sensorimotor representation of the category referent (*dog*) and the pictured object (*dog*) also facilitates responding, but likely to a lesser extent because linguistic activation tends to operate faster than sensorimotor activation (Barsalou et al., 2008; Connell, 2018). The extent of a basic-level advantage would then depend on the overlap in sensorimotor and linguistic distributional experience between the alternative category labels and the picture name.

For example, when a category label such as *Labrador* is presented, it automatically activates a sensorimotor simulation of the referent (e.g., perceptual and action experience of a *Labrador*) and linguistic distributional neighbours of the label (e.g., words that appear in similar contexts to *"Labrador"*). Next, the picture is presented and is automatically labelled as *dog*, which activates the linguistic distributional neighbours of *dog* and a more detailed sensorimotor simulation of the pictured *dog*. The more similar the sensorimotor experience and linguistic contexts are between *Labrador* and *dog*, the more they facilitate fast and accurate category verification, and the closer their response latency and accuracy will be to the basic-level case (e.g., label *dog* → [picture of a *dog*]). On the other hand, the more distant the category label and picture are in sensorimotor and linguistic distributional experience (e.g., *animal* → [picture of a *dog*]), the slower and more error-prone category verification will be.

Specifically, we predicted that both sensorimotor and linguistic distributional information would inform categorisation, and that linguistic distributional information would contribute above and beyond sensorimotor alone. We also predicted that a combination of sensorimotor and/or linguistic distributional information would explain categorisation performance (RT and accuracy) better than traditional accounts of the basic-level advantage, which are based on discrete levels in a taxonomic hierarchy (i.e., subordinate, basic, superordinate).

2.4.1. Method

Participants. Thirty native speakers of English (23 female; $M_{\text{age}} = 22$ years, $SD = 3.69$) were recruited from Lancaster University in return for partial course credits or a sum of money (£3.50). We determined sample size via sequential hypothesis testing using Bayes Factors (Schönbrodt et al., 2017), which allows evidence for/against the hypothesis to accumulate until a pre-specified threshold of evidence is reached, and thus enables flexible sampling without increasing Type 1 error. We stopped sampling at the minimum bound of Nmin = 30, when analysis A for RT (see Data Analysis) cleared the specified grade of evidence $BF_{10} \geq 3$ (actual $BF_{10} = 1998.20$). This threshold indicated that a basic-level advantage could be detected in our data (i.e., categorical decisions were made faster when the displayed word label was at the basic level compared to the superordinate and subordinate levels).

The pre-registered accuracy threshold of 80% correct answers on fillers we established based on pilot testing proved to be too strict, and would have led to the exclusion of 12 participants. As a result, we decided to deviate from the pre-registration and lower this threshold to 70%; one participant did not pass this new threshold and was replaced.

Materials. Test items consisted of 216 label → picture items, comprising 72 target pictures (depicting natural objects and artefacts), each of which was paired with three labels that correctly described it at the subordinate, basic and superordinate level (e.g., the picture of a *Labrador* was pared with labels *animal*, *dog* and *Labrador* respectively). We sourced all pictures through online image search, ensuring they were labelled for reuse with modification and had a minimum size of 1024x768 pixels. We then edited all pictures to display only target objects on a white background. All 72 subordinate labels were uniquely paired with pictures (e.g., label *Labrador* → picture *Labrador*), and 24 basic-level categories were paired with three different images (e.g., label *dog* → pictures *Labrador*, *collie*, and *poodle*). Finally, these 24 basic-level categories were grouped into superordinate categories at 2-7 members apiece, meaning that nine superordinate labels were paired with each between 6-21 different pictures (e.g., label *animal* → pictures *Labrador, collie, poodle, chimpanzee, gorilla, orangutan*, etc.). We ensured that all labels were present in Lynott et al.'s (2020) sensorimotor norms to allow for the calculation of sensorimotor distances (see Design and Analysis). Finally, we divided all 216 test items into three stimulus lists of 72 items, where each list featuring 24 subordinate, 24 basic and 24 superordinate labels and included each picture only once.

Filler items consisted of 116 label → picture pairs, containing similar object pictures and labels to test items. Of these, 71 false fillers were seen by all participants, and featured 23 superordinate (e.g., label "*publication*" → picture *eggplant*), 24 basic-level (e.g., label *"horse"* → picture *zebra*) and 24 subordinate (e.g., label *"anchovy"* → picture *sunglasses*) labels. Forty-one of these fillers were easily recognisable as false (i.e., pictured object clearly unrelated to the label e.g., *label "frog"* → picture *shamrock*) and thirty were more challenging (i.e., pictured object belonged to the same superordinate category as the label, e.g., label *"cow"* → picture *buffalo*). A further 11 unique filler items were added to each stimulus list, featuring labels that appeared once among the items of that list, in order to ensure that repeated labels amongst test items could not cue participants to respond "yes" to category membership (e.g., the label "*animal*" appears in multiple true test items). Of these fillers, five were superordinate (3 true; 2 false) and six were basic-level (3 true; 3 false). Finally, to balance the true/false proportion per category type, we added 12 fillers that were the same for all lists, with unique subordinate labels (6 true; 6 false). As a result, the final stimulus lists each contained 166 label → picture pairs, divided evenly between true and false (72 true test items, 11 true fillers and 83 false fillers).

Procedure. Trials were presented on a white background, using PsychoPy (version 1.84.1; Peirce, 2007). Each trial began with a blank screen displayed for 200 ms followed by a fixation cross for 300 ms, the label (centred, black lowercase Arial, 52 px) for 1000 ms, another blank screen for 200 ms, a fixation cross for 300 ms, and finally the picture which remained onscreen until a response key was pressed. Participants sat in front of a computer with a keyboard. They were told they would see a series of word-picture pairs where the word represented a category and the picture a potential member of that category. They were asked to press YES (z-key on the keyboard) when the picture showed a valid category member and NO (m-key) when it did not. Response times were measured from the onset of the picture to the onset of a valid keypress, and accuracy of each decision was also recorded. Participants were randomly assigned to a stimulus list. Test and filler items appeared in random order with a self-paced break every 60 trials. Testing took approximately 20 minutes, including informed consent and debriefing.

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.

Critical predictors. As well as a specified taxonomic level (subordinate, basic, superordinate), each label → picture test item had an associated value in two critical predictors that captured the overlap in sensorimotor and linguistic distributional experience between category concept and member concept.

Linguistic distance. Using a subtitle corpus consisting of 200 million words in British English (see van Heuven et al., 2014), we calculated log co-occurrence frequencies around each word with a context radius of five. Each word in the corpus was represented as a vector of log co-occurrence frequencies, allowing us to compare two words by calculating the cosine distance between their vectors (i.e., $1 - \cos(\theta(u,v))$). For example, the words *dog* and *animal* generally appear in relatively similar contexts across language, therefore distance between their vectors in linguistic space is smaller than the distance between two words that appear in very different contexts, such as *dog* and *spaghetti* (.23 for the former compared to .46 for the latter example).

Previous research suggests that pictures tend to be implicitly named with the most frequent, earliest-acquired word (e.g., Bates et al., 2003; Belke et al., 2005), which is usually the basic level (Rosch, Mervis, et al., 1976). We based our calculations on this assumption (i.e., we used the basic label *dog* as the name for all three pictures of dogs, regardless of whether it contained a *Labrador*, *collie* or *poodle*). As a result, the corresponding linguistic distance for basic-level label → picture items was always zero (e.g., *dog* → *dog* distance = 0), and the ability of linguistic distance to predict categorisation performance depended on the presence of a systematic relationship between our dependent variables (i.e., RT, accuracy) and the linguistic distances of superordinate (e.g., *animal* → *dog*) and subordinate items (e.g., *Labrador* → *dog*). While this approach assumes which word label will be implicitly activated by a picture, it ensured fair comparison with the taxonomic category predictors, in which the basic-level category took the reference level of 0 in the dummy coding of sub- and superordinate categories (see data analysis). The final linguistic distance measure for each label → picture pair ranged in theory from -1 to +1 (actual range = [.00, .83], *M* = .25, *SD* = .21), with higher values indicating greater distance in linguistic space (i.e., less overlap in the linguistic distributional experience of each word).

Sensorimotor distance. To compare how two concepts overlapped in terms of sensorimotor experience, we took the novel approach of calculating sensorimotor distance based on multidimensional ratings of sensorimotor strength. We used Lynott et al.'s (2020) sensorimotor norms for 40,000 concepts, in which people rated the extent to which they experienced a particular concept via six perceptual modalities (auditory, gustatory, haptic, interoceptive, olfactory, visual) and by performing an action with five action effectors (foot, hand, head, mouth, torso), where each dimension was separately rated on a scale from 0 (not at all) to 5 (greatly). Each concept was therefore represented by an 11-dimensional vector of grounded sensorimotor experience, allowing us to compare two words by calculating the cosine distance between their vectors (as for linguistic distance).

The final sensorimotor distance measure for each label → picture pair ranged in theory from -1 to +1 (actual range [.00, .29], M = .03, SD = .05), with higher values indicating greater distance in sensorimotor space (i.e., less overlap in the sensorimotor experience of each concept). Linguistic and sensorimotor distance measures were moderately correlated, $r = .38$, BF_{10}=903110.10 (14.44% shared variance).

Data analysis. We planned three sets of analyses to test our hypotheses. Analysis A tested whether a classic basic-level advantage could be distinguished in the data. We ran a mixed effects linear regression of RT (correct trials only) with crossed random effects of participants and items, and fixed effects of taxonomic level (dummy-coded as superordinate and subordinate variables with basic as the reference level). We also ran a mixed effects logistic regression (binomial, logit link) of accuracy (incorrect = 0, correct =1; all trials included), with crossed random effects of participants and items, and fixed effects of taxonomic level (coded as above). For both analyses, we used Bayesian model comparisons (Bayes Factors calculated from BIC; Wagenmakers, 2007) to test whether the data favoured a model containing the above fixed effects over a null model containing only random effects.

Analysis B tested whether variance in RT and accuracy could be explained by sensorimotor and linguistic distance. Model comparisons tested whether the data favoured a model containing both sensorimotor and linguistic fixed effects over a model containing only a sensorimotor fixed effect. Although not specified in the pre-registered analysis plan due to an error of omission, we also tested whether the data favoured a model with only sensorimotor distance over a null model containing only random effects (i.e., reflecting our pre-registered hypothesis that sensorimotor distance contributes to categorical decision making).

Finally, Analysis C tested whether RT and accuracy were best explained by traditional taxonomic levels or by sensorimotor-linguistic information. Model comparisons tested whether the data favoured the best-fitting sensorimotor-linguistic model from Analysis B over the taxonomic model from Analysis A. Linear models of RT and logistic models of accuracy were compared separately.

For each analysis, we report the coefficients and NHST statistics of fixed effects in the best-fitting model.

2.3.2. Results and Discussion

We removed as outliers 55 trials from the RT analysis (2.81 % of 1960 correct responses) and 63 trials from the accuracy analysis (2.92% of 2160 responses) that had RTs more than 2.5SD from the participant’s mean. Table 1 shows results of all model comparisons.

Table 1

Model comparisons for linear mixed effect regressions of RT and logistic mixed effects regressions of accuracy in Experiment 1 showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Analysis	Model comparison	RT		Accuracy	
		ΔR^2	BF_{10}	ΔR^2	BF_{10}
	Null model (random effects)	.264	-	.270	-
A	Taxonomic levels vs. null	.012	1998.19	.065	49020.80
B	Sensorimotor distance vs. null	.011	9414.44	.016	12.81
	Sensorimotor + Linguistic distance vs. null	.012	383.75	.041	298.87
	Sensorimotor + Linguistic distance vs. Sensorimotor-only	.001	0.04	.025	21.11
C	Best Sensorimotor-Linguistic model vs. taxonomic levels.	-	4.71	-	0.01

Relationship between dependent variables. An exploratory Bayesian correlation analysis using JASP (v.0.14.0.0, with default beta prior width =1) between a per-item average RT and percentage of correct answers provided moderate evidence *against* a correlation (BF_{01} = 3.72), suggesting that accuracy and response time were not meaningfully related (see OSF for analysis code and results).

Taxonomic levels. Bayesian model comparisons in Analysis A showed very strong evidence for models containing taxonomic levels over a null model containing only random effects of participant and item on participant's RT (BF_{10} = 1998.20) and accuracy (BF_{10} = 49020.80). In RT, categorisation decisions made at the subordinate level were 37 ms slower than at the basic level [unstandardized β = 36.99, 95% CI = ±29.41, t(1822.17) = 2.46, p = .014]. Furthermore, categorisation decisions at the superordinate level were 84 ms slower than at the basic level [β = 83.81, 95% CI = ±29.68, t(1828.05) = 5.53, p <.001]. Accuracy was overall high. Participants were most likely to answer correctly when an image was preceded by a label at the basic level (predicted probability of a correct answer = 97.7%). Compared to the basic level, participants were 2.74 times more likely to respond incorrectly when an image was labelled at the subordinate level, [β =-1.01, 95% CI = ±0.45, z = -4.37, p < .001] (predicted probability of a correct answer = 93.9%). Finally, participants were up to 3.71 times more likely to respond incorrectly at the superordinate level [β = -1.31, 95% CI = ±0.44, z = -5.83, p < .001] (predicted probability of a correct answer = 91.8%). Our data thus replicates the classic basic-level advantage in categorisation.

Sensorimotor-linguistic predictors. In RT, Analysis B model comparisons showed very strong evidence for the effect of sensorimotor distance over a null model containing only random effects (BF_{10} = 9414.44). While a model containing both sensorimotor and linguistic distance was also better than the null (BF_{10} = 383.75), model comparisons indicated strong evidence *against* the inclusion of linguistic distance in RT models (i.e., the data were BF_{10} = 24.53 times more likely under a model containing only sensorimotor distance compared to a model containing both sensorimotor and linguistic distances). Hence, the best sensorimotor-linguistic model of RT was sensorimotor distance alone, where RT increased with sensorimotor distance (β = 722.97, 95% CI = ±277.49, t (1690.37) = 5.11, p <.001), by up to 210 ms.[6]

In accuracy, there was positive evidence for the effect of sensorimotor distance alone (BF_{10} = 12.81), but very strong evidence for the inclusion of linguistic distance alongside sensorimotor distance (BF_{10} = 289.87). Hence, the best sensorimotor-linguistic model of accuracy included both sensorimotor and linguistic distance. Coefficients showed that participants were more likely to respond incorrectly as sensorimotor distance increased (β = -3.51, 95% CI = ±3.21, z =-2.15, p = .030) *and* as linguistic distance increased (β = -1.63, 95%

[6] This value reflects the change in the dependent variable at the maximum sensorimotor (.29) or linguistic (.83) distance between categories and members in our dataset, calculated as a proportion of the beta-coefficient (e.g., .29*722.97 = 209.66).

CI = ±0.82, z = -3.87, p <.001) between categories and member concepts. That is, participants were 2.77 times[6] more prone to error for items at the greatest sensorimotor distance (.29) than for items at the smallest sensorimotor distance (zero). Simultaneously, for items at the greatest linguistic distance (0.83) participants were 4.66 times[6] more prone to error than for items at the smallest linguistic distance (zero).

As predicted, sensorimotor and linguistic distributional information contribute to categorical decision making. Overlap in sensorimotor experience between category concept and member concept (e.g., between *animal* and *Labrador*) facilitates categorisation RT and accuracy. Similarly, overlap in linguistic experience between the distributional patterns of category and member label also facilitates categorisation accuracy; however, categorisation RT was not influenced by linguistic distance.

Best model. Analysis C showed mixed results as to whether the taxonomic or sensorimotor-linguistic model best explained the data. As predicted, Bayesian model comparisons found positive evidence that sensorimotor distance was BF_{10} = 4.71 times better than taxonomic levels in predicting RT. By contrast, and contrary to predictions, there was strong evidence that taxonomic levels were BF_{01}= 164.02 times better than sensorimotor and linguistic distance at predicting accuracy.

Summary. Overall, while both sensorimotor and linguistic distributional information contribute to categorical decision making, they did not systematically do better than a taxonomic hierarchy of subordinate, basic, and superordinate levels. That is, sensorimotor information predicted response times best, while taxonomic level predicted accuracy best. We address a possible cause in experiment 2 below.

2.4.2. Exploratory analysis: Psycholinguistic characteristics

Research has shown compelling evidence for the influence of a number of psycholinguistic characteristics of objects and their names on a range of tasks which involve lexical and semantic processing. For example, various researchers have found an effect of the age at which objects and their names are typically learned (age of acquisition, AoA) on performance in object recognition (Catling et al., 2008; Dent et al., 2007; Urooj et al., 2014), word recognition, picture naming (Belke et al., 2005; Morrison et al., 2003) and object classification (Catling & Johnston, 2006). However, research has also uncovered effects of the frequency of words in written and spoken language, the degree to which people are familiar with these words as well as their length.

In the present work, we have thus far not addressed the effects psycholinguistic characteristics of in particular the category label may have had on performance in our task. However, it may be of interest to do so for two reasons: firstly, the wealth of evidence from psycholinguistic research which suggests in particular age of acquisition and word frequency affect performance in word recognition as well as object recognition, which may carry over to a semantic decision-making process. In other words, it may be the case that the delay that is observed for recognising words and objects with names that are less frequent and are acquired later in life also causes a delay in the category verification process. Secondly, previous categorisation research has suggested that the basic-level advantage may extend to their frequency in everyday language, as well as to the order in which category labels are learned. For example, early categorisation research suggests that that children acquire more differentiated categories (i.e., categories that maximise within-category similarity and between category-distinction, typically basic-level categories) before less differentiated categories (Mervis & Crisafi, 1982b; Mervis & Rosch, 1981). More recently, some have nuanced this by arguing that children follow a broad-to-narrow pattern, displaying appropriate distinguishing behaviour between animals and non-animals before displaying appropriate distinguishing behaviour between rabbits and birds (Mandler & McDonough, 2000). While these accounts ultimately find each other in the belief that adults display a basic-level advantage, they make seemingly different predictions for the relationship between the age at which basic-level labels are acquired and later performance differences. This in turn might result in different predictions for the effects a measure of AoA might have on categorisation performance in our task. Early categorisation researchers might argue that since basic-level labels are most differentiated, they are learned earlier and therefore AoA effects should mirror the classic basic-level advantage. If, by contrast, basic-level labels are learned *after* superordinate-level label, AoA effects should not match the classic basic-level advantage.

In this exploratory analysis, our aim was twofold. Firstly, we aimed to test whether the psycholinguistic characteristics of the category label indeed affect categorisation performance, and if so in which direction. Where an effect would be observed, we expected that RT would increase with age of acquisition and word length, and decrease as word frequency and familiarity increased. On the other hand, we expected accuracy to decrease as age of acquisition and word length increased, and to increase with word frequency and familiarity. Secondly, we aimed to test whether the previously observed effects of taxonomic level, sensorimotor and linguistic distributional information could still be observed after the inclusion of psycholinguistic characteristics of the category label. Specifically, we investigated whether psycholinguistic characteristics (i.e., AoA, word frequency, familiarity and word length) of the category labels used in our task were better predictors of categorisation performance (accuracy, RT) than taxonomic level. Furthermore, we tested whether the inclusion of psycholinguistic characteristics changed the previously observed effects of sensorimotor and linguistic distributional information on RT and accuracy.

To test this, we carried out a set of additional mixed effects regressions of RT and accuracy. Analysis D tested whether the taxonomic level of the category predicted RT and accuracy above and beyond the psycholinguistic characteristics of the category label. To this end we ran two linear mixed effects regressions of RT, with random effects as per analysis A. The first regression included fixed effects of subjective estimates of AoA (retrieved from: Kuperman et al., 2012), Zipf log word frequency (retrieved from: van Heuven et al., 2014), subjective familiarity ratings (retrieved from: Scott et al., 2019; Stadthagen-Gonzalez & Davis, 2006; and the MRC Psycholinguistic database) and word length in characters excluding spaces where available for all category labels (calculated in R). All psycholinguistic characteristics were centred around their respective means. A second regression included the additional fixed effects of taxonomic level (dummy-coded as outlined in analysis A). Furthermore, we ran two mixed effects logistic regressions of accuracy (binomial, link = logit), with the same fixed and random effects structure as for the analysis of RT. We used Bayesian model comparisons to test whether the data favoured a psycholinguistic-only model over a model containing only random effects, and whether the data favoured models including taxonomic levels over those including only psycholinguistic variables (i.e., whether taxonomic information predicted additional variance to psycholinguistic information).

Analysis E tested whether sensorimotor and or linguistic distributional information predicted RT and accuracy above and beyond the psycholinguistic characteristics of the category label. To this end, we ran three linear mixed effects regressions of RT with random effects as per analysis B. The first regression included fixed effects of AoA, Zipf log word frequency, average familiarity rating and word length in characters excluding spaces. A second regression added a fixed effect of sensorimotor distance between the category label and implicit image name, while a third regression added a fixed effect of linguistic distributional distance between a category label and implicit image name. We also ran three mixed effects logistic regressions of accuracy (binomial, link = logit) with the same fixed and random effects structure as for the analysis of RT. Bayesian model comparisons tested whether the data favoured a model including sensorimotor and/or linguistic distributional information over that including only psycholinguistic characteristics of the category label.

Exploratory results

We could not retrieve information on all category labels used in our task for all psycholinguistic variables. As a consequence, we limited the dataset for this analysis to complete cases only. In the analysis of accuracy, we included 1687 out of 2097 trials (excluding 19.55 % of trials). In the analysis of RT, we included 1531 out of 1905 correct trials (excluding 19.63 % of trials). Because of this, these exploratory results should be interpreted with care. We also note here that we found evidence of moderate to strong correlations between psycholinguistic predictors (see supplementals on OSF) which is in line with previous research (e.g., Brysbaert & Ghyselinck, 2006). Nevertheless, we found no evidence of multicollinearity (i.e., VIF < 10) for any of the models tested (see supplemental materials on OSF).

Psycholinguistic characteristics vs. taxonomic level. In analysis D of RT, we found evidence that a model including psycholinguistic effects fit the data worse than a model containing only random effects, with Bayesian model comparisons suggesting the data was BF_{01} = 16.73 times more likely under a random effects model. Interestingly, we found evidence against the inclusion of dummy-coded predictors of taxonomic level (i.e., BF_{01} = 2.25), even when a taxonomic-only model was supported by the data far better than a psycholinguistic-only model (BF_{10} = 14568.12). Consequently, the psycholinguistic-taxonomic model did also not outperform a random-effects model (BF_{01} = 37.66).

In analysis D of accuracy, in contrast to the analysis of RT, we found evidence that a model containing psycholinguistic characteristics of the category label outperformed a model containing only random effects, with Bayesian model comparisons suggesting the data were BF_{10} = 126076.40 times more likely under the psycholinguistic model. Furthermore, in contrast to the analysis of RT, we found evidence that a model including taxonomic levels outperformed a psycholinguistic-only model (BF_{10} = 15.37) in predicting accuracy. Inspection of the coefficients of the psycholinguistic-taxonomic model showed that increasing age of acquisition associated with the category label reduced the likelihood that participants categorised an image correctly by up to 19.25 times[6] [β = -.46, 95% CI = ±.25, z = -3.58, p <.001]. In contrast to our confirmatory analyses (see analysis A above), we observed only a partial basic-level advantage. That is, in line with our confirmatory analyses, compared to the basic level, participants were less likely to correctly categorise an image when it followed a superordinate-level label [β = -1.29, 95% CI = ±.61, z = -4.12, p <.001]. However, accuracy was not significantly worse at the subordinate compared to the basic level [β = .13, 95% CI = ±.95, z = -.28, p =.78]. Here too, we note that a model we ran on the same truncated dataset including only taxonomic predictors outperformed the psycholinguistic model (BF_{10} = 32.35). Interestingly, inspection of this model showed the classic basic-level advantage over both subordinate and superordinate-level labels.

Psycholinguistic characteristics vs. sensorimotor-linguistic information. In analysis E of RT, we found evidence for the inclusion of sensorimotor distance to a model containing random effects and psycholinguistic characteristics (BF_{10}= 140.04). This model outperformed a model containing only random effects (BF_{10} = 8.39). In line with confirmatory analysis B, we found evidence against the inclusion of linguistic distributional information to the sensorimotor-psycholinguistic model (BF_{01}=13.21). Participants were up to 219.95 ms slower to categorise objects with the largest sensorimotor distance (.29) between the category label and basic-level image name. We note here that the truncated dataset strongly supported a sensorimotor-only model over a sensorimotor-psycholinguistic model (BF_{01} = 802.76).

In analysis E of accuracy, we found equivocal evidence for the inclusion of sensorimotor distance to a model containing random effects and psycholinguistic characteristics (BF_{10}= 1.37). Contrasting the results from confirmatory analysis B, we found equivocal evidence for the inclusion of linguistic distributional information to a sensorimotor-psycholinguistic model (BF_{10} = 1.79). Here, we note that a non-psycholinguistic sensorimotor-linguistic model of the truncated data greatly outperformed both a non-psycholinguistic sensorimotor model (BF_{10} = 65657.21) as well as a psycholinguistic-sensorimotor-linguistic model (BF_{10} = 41.07), suggesting that the reduced effect of linguistic distributional information may have been a consequence of the inclusion of psycholinguistic predictors and not truncating the dataset.

In analysis F of RT, Bayesian model comparisons showed that the data were BF_{10} = 14.04 times more likely under the best-performing psycholinguistic model from analysis E (psycholinguistic + sensorimotor information) than under the best-performing psycholinguistic model from analysis D (psycholinguistic information only). By contrast, analysis F of accuracy showed that the data were more BF_{10} = 6.30 times more likely under the best-performing analysis E model (psycholinguistic + taxonomic level) than under the best-performing analysis D model (psycholinguistic + sensorimotor + linguistic information).

These results are exploratory, and must be interpreted with caution. When taken at face value, the fact that a model including psycholinguistic characteristics of the label did not predict RT above and beyond random effects suggests that the current data do not support a psycholinguistic interpretation of the time course in object categorisation. As such, they are in line with previous categorisation research (e.g., Murphy & Medin, 1982) that suggests little effect of order of learning and frequency. The fact that the inclusion of taxonomic level to a psycholinguistic model did not improve model fit, while a taxonomic-only model greatly outperformed both a psycholinguistic and random effects model is puzzling, but possibly speaks to the highly exploratory nature of these analyses. This is particularly likely as in the case of RT, truncation of the data as a result of limited availability of psycholinguistic measures, led to a reduction of the basic-level advantage over subordinate-level labels. Future research may address this particular problem by ensuring the use of only items for which psycholinguistic characteristics are available.

The accuracy results suggest a psycholinguistic interpretation is possible, yet not most likely. While the inclusion of psycholinguistic characteristics improved model fit over random effects, the best-fitting psycholinguistic model included taxonomic levels. Interesting are the fact that models including psycholinguistic characteristics reduced both the advantage of basic-level labels over subordinate-level labels, and the ability of linguistic distributional information to predict accuracy over sensorimotor information. It is possible that the inclusion of psycholinguistic characteristics resulted in net suppression effects in our model. However, overall, our exploratory analyses seem to indicate that simpler, non-psycholinguistic models still outperform their psycholinguistic counterparts when tested on a similarly truncated dataset.

Finally, we need to be cautious in interpreting results from models in which predictors are strongly related. This is a well-documented problem in psycholinguistic research, and is typically addressed during study design (e.g., by using orthogonal designs). These exploratory results illustrate this difficulty, and suggest that any future studies that aim to explore a combination of psycholinguistic effects and sensorimotor-linguistic effects would need to take care to ensure an orthogonal design, and might use positive pointwise mutual information (PPMI; Bullinaria & Levy, 2007) rather than log co-occurrence to minimise the relationship between linguistic distributional information measures and psycholinguistic characteristics such as word frequency. Investigating psycholinguistic characteristics alone may ultimately not bring us closer to understanding why certain category labels have a performance advantage over others, particularly given fact that the previously observed effect of sensorimotor information is largely unaffected. That is, if the basic-level advantage is indeed explained by word frequency and age of acquisition (i.e., basic-level labels are learned earlier and are more frequent in language), then why are these category labels more frequent, and why are they learned earlier?

2.5. Experiment 2: Typicality and the Basic-Level advantage

In experiment 1, we assumed all pictures to be implicitly named at the basic level. However, categories can be graded in terms of the "goodness" of membership: that is, how members range in typicality of their respective categories, which affects category processing and production (Armstrong et al., 1983; Rips et al., 1973; Rosch, 1973; Rosch, Mervis, et al., 1976; Rosch, Simpson, et al., 1976; Rosch & Mervis, 1975; E.E. Smith et al., 1974). That is, categorising typical items (e.g., *sparrow*, for the category *bird*) tends to be faster than categorising atypical items (e.g., *penguin*). However, unusual, atypical members tend to be named at the specific, subordinate level rather than at the more general basic level (e.g., a picture of a penguin is more likely to be named as a *penguin* rather than a *bird*: (Rosch, Mervis, et al., 1976; Snodgrass & Vanderwart, 1980). Jolicoeur et al. (1984) interpreted this finding to mean that the subordinate level, rather than the basic level, acts as the entry point into the taxonomic hierarchy of semantic memory for atypical category members. Alternatively, differentiation accounts proposed that a picture of an atypical bird like a *penguin* is more easily categorised at the subordinate level of *penguin* because its features are better matched at this specific level (i.e., the subordinate level is maximally informative and distinctive) compared to more general levels like *bird* or *animal* (Murphy & Brownell, 1985). In both accounts, categorisation of typical category members would therefore show the traditional basic-level advantage (i.e., basic level faster and more accurate than subordinate and superordinate levels), but categorisation of atypical members would show a different pattern (i.e., subordinate level faster and more accurate than basic level, followed by superordinate level).

If it is indeed the case that unusual category members are implicitly named at the specific, subordinate level, then it also affects how sensorimotor and linguistic information should be operationalized. In the previous experiment, we calculated sensorimotor and linguistic distance from the category name to the basic-level label of the pictured object. For example, in the item *animal* → picture of *poodle*, we assumed the picture would be implicitly labelled as the basic-level *dog*, hence sensorimotor and linguistic distance was calculated from *animal → dog*. However, for unusual category members, sensorimotor and linguistic distance should instead be calculated from the category name to the *subordinate* label. If a poodle is an unusual dog, then its picture would be implicitly labelled as *poodle*, and the item *animal* → picture of *poodle* should have its sensorimotor and linguistic distance calculated from *animal → poodle*. Since previous work has shown a close relationship between typicality and overlap of linguistic distributional experience (e.g., Connell & Ramscar, 2001), we opted to implement an internally-consistent adjustment for categorical gradedness by using linguistic distance to determine whether a member concept should be considered a good or poor example of its category. Member concepts that were close to their category concept (e.g., *salmon* and *fish* appear in very similar linguistic contexts, and have a cosine distance of .28) were considered good category members whose pictures would activate basic-level labels and corresponding sensorimotor information, whereas member concepts that were distant from their category concept (e.g., *sailfish* and *fish* appear in rather different linguistic contexts, with a cosine distance of .67) were considered poor/unusual members that would activate subordinate labels and corresponding sensorimotor information. With this adjustment for categorical gradedness in place, we could characterise as before the representational overlap of linguistic and sensorimotor information between a category and member concept.

Hence, in this study (data, analysis code, results, and preregistration available at https://osf.io/vdka2/?view_only=ac6c78793a354a1f95da6c36b5ac6163), we collected typicality ratings for each of our subordinate-level stimuli as a member of its basic-level category (e.g., typicality of *sailfish* as a *fish*) and examined its influence on categorical decision making using the dataset of Experiment 1. In line with previous research on the role of object typicality in categorisation (e.g., Jolicoeur et al., 1984; Murphy & Brownell, 1985), we expected typicality to enhance the ability of traditional taxonomic accounts to explain the basic-level advantage in categorisation, and to interact with subordinate taxonomic level so that typical items would show a basic-level advantage but atypical items would show a subordinate-level advantage. From the sensorimotor-linguistic perspective, we also hypothesised that linguistic distributional information would capture the graded structure of categories, whereby linguistic distance would correlate negatively with typicality ratings (i.e., less typical = greater linguistic distance between category and member concept). Using gradedness-adjusted measures of sensorimotor and linguistic distance, we predicted – as before – that linguistic distance would predict categorisation performance above and beyond sensorimotor distance (i.e., greater category-member distance results in slower RT and poorer accuracy in categorical decision) and that the best sensorimotor/linguistic model would outperform the taxonomic-typicality model.

2.5.1. Method

Materials & dependent measures. We used the categorical decision dataset from Experiment 1, with new predictors as outlined below.

Critical predictors. As well as a specified taxonomic level (subordinate, basic, superordinate), each label → picture test item had an associated typicality rating of the pictured object as a member of its basic-level category. In addition, each label → picture test item had a gradedness-adjusted measure that captured the overlap in sensorimotor and linguistic distributional experience between category concept and member concept.

Typicality ratings. We collected typicality ratings from 12 naïve participants (all native speakers of English) for each of the 72 test items that comprised a basic-subordinate concept pair (e.g., *fish-salmon*). These ratings were collected as part of a larger study collecting typicality ratings for 2280 category-member items, where items were divided into lists of 120 items each (Banks & Connell, 2021); the present 72 category-member items were randomly spread across lists. Participants were asked to rate how good an example of the basic-level category (e.g., *fish*) they thought each subordinate category member (e.g., *salmon*) to be, on a scale from 1 (very poor) to 5 (very good); alternatively, they could select a "don't know" option if they were not familiar with the category or member concept in question. Data collection stopped when every item had 12 valid ratings. We then calculated the average typicality rating per category member, and used this typicality rating on every label → picture trial where it was presented (e.g., all trials with a *salmon* picture used the typicality rating for *salmon* as a kind of *fish*). Mean typicality rating across all items was 4.53 (SD = .36, range = [3.42, 5.00]).

Gradedness-adjusted linguistic and sensorimotor distance. For our original calculation of linguistic distance in Experiment 1, we assumed that all pictured objects were implicitly named at the basic level (i.e., we used the basic label *fish* as the name for all three pictures of fish). In order to incorporate categorical gradedness into linguistic and sensorimotor distance, where pictures of unusual category members would instead be implicitly named at the subordinate level (e.g., using the specific label *sailfish* as the name for the picture of a *sailfish*), we used the data to determine the tipping point of linguistic distance that distinguished good from unusual category members. We first examined the distribution of linguistic distance between all 72 subordinate member concepts and their basic-level category concept (e.g., *fish → sailfish*, *fish → salmon*) and visually established 10 potential thresholds beyond which we assumed member concepts to be unusual for their category. We then replaced the linguistic distances of all items that fell beyond each threshold with those calculated using the subordinate name as the picture label. For instance, if the linguistic distance for *fish → sailfish* exceeded the threshold, we replaced all linguistic distances of *sailfish* trials (originally calculated as *animal → fish*, *fish → fish*, *sailfish → fish*) with their gradedness-adjusted linguistic distances (i.e., *animal → sailfish*, *fish → sailfish, sailfish → sailfish*). For these same items, we likewise replaced the original sensorimotor distances with their gradedness-adjusted sensorimotor distance.

Finally, we examined which threshold was best supposed by the data by running mixed effect regression of RT per candidate threshold, with random effects of participant and item and fixed effects of gradedness-adjusted sensorimotor and linguistic distance. Model comparisons showed that the best-fitting model was based on a gradedness-adjusted linguistic distance threshold of .33, and that the data favoured this model $BF_{10} = 1908.20$ times more strongly than the original model used in Experiment 1. In short, a linguistic distance of .33 acted as a tipping point between good category members ($N = 23$) that were implicitly named at the basic level and unusual category members ($N = 49$) that were implicitly named as the specific, subordinate level. We therefore used these optimal gradedness-adjusted linguistic and sensorimotor distances in subsequent analyses. Gradedness-adjusted linguistic distance ($M = .30$, $SD = .25$, range = [0, .83]) and sensorimotor distance ($M = .04$, $SD = .05$, range = [0, .29]) were moderately correlated at $r = .53$ (i.e., 28.30% shared variance).

Data analysis. Five sets of analyses were planned to test our hypotheses: two to test the predictions of traditional taxonomic accounts of the basic-level advantage, and three to test our sensorimotor-linguistic account. Analysis A tested whether including item typicality would predict categorical decision RT and accuracy better than taxonomic information alone. RT (correct trials only) and Accuracy (all trials) were analysed using the model specifications of Analysis A in Experiment 1, with an additional fixed effect of typicality (variable centred). Bayesian model comparisons tested whether the data favoured random effects only, taxonomic level, or taxonomic level and typicality. Analysis B tested whether categorisation at the subordinate versus basic level differed for typical and atypical category members. RT and accuracy were analysed as per Analysis A, with the additional fixed effect of interaction between typicality and subordinate taxonomic level (i.e., where basic level is coded as the reference level). Model comparisons tested whether the data favoured this interaction model over the final model of Analysis A.

In order to test our sensorimotor-linguistic predictions, Analysis C investigated if linguistic distance and typicality were correlated, using a Bayesian correlation analysis in JASP (version 0.9.2: JASP Team, 2019) with default beta prior width =1 and a directional hypothesis of negative correlation (i.e., higher distance = less typical). Analysis D examined whether variance in RT and accuracy could be explained by gradedness-adjusted sensorimotor and linguistic distance. Model comparisons tested whether the data favoured a model containing sensorimotor distance alone, or a model containing the additional fixed effect of linguistic distance. Finally, in Analysis E, we investigated whether RT and accuracy were best explained by traditional taxonomic-typicality information or by sensorimotor-linguistic information. Therefore, in analysis E, Bayesian model comparisons determined whether our data favoured the best-performing taxonomic-typicality model from analysis B or the best-performing sensorimotor-linguistic model from analysis D.

2.5.2. Results and discussion

Outlier trials were removed from the dataset in Experiment 1. In addition, we removed one item (*gavel*) from both the RT (21 out of 1905 trials, 1.10%) and accuracy (28 out of 2097 trials, 1.33%) data because it had an exceptionally low typicality rating (2.2 on a 1-5 scale) that made it an outlier more than five standard deviations below the mean item typicality (*M* = 4.53, *SD* =.45). While removing this item deviated from the pre-registered analysis plan, we felt it was necessary to avoid compromising the robustness of our analyses. The classic basic-level advantage remained intact with this item excluded: there was very strong evidence for including taxonomic levels in analysis of both RT (BF_{10} = 10938.02) and accuracy (BF_{10} = 84965.45). Categorisation at the basic level was faster than at subordinate [unstandardized β = 42.11, 95% CI = ±29.32, t (1803.90) = 2.81, p = .005] and superordinate levels [β = 88.15, 95% CI = ±29.61, t (1809.75) = 5.83, p <.001]. Similarly, accuracy was greatest at the basic level, followed by subordinate [β =-1.06, 95% CI = ±0.47, z = -4.44, p < .001] and superordinate [β = -1.36, 95% CI = ±0.45, z = -5.84, p < .001] levels[7].

Table 2 shows model comparisons for analyses A, B, D and E.

[7] Full coefficient statistics for all models including and excluding the *gavel* item are included in supplementals. In brief, analysis including this outlier item did not affect inferences based on Bayesian model comparisons, but in did create a weak but significant coefficient effect of typicality in Analysis B that disappeared when the item was excluded, which suggested we were correct to remove it.

Table 2

Model comparisons for linear mixed effect regressions of RT and logistic mixed effects regressions of accuracy in Experiment 2 showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Analysis	Model comparison	RT		Accuracy	
		ΔR^2	BF_{10}	ΔR^2	BF_{10}
	Null model (random effects)	.262	-	.261	-
A	Taxonomic levels vs. null	.013	10938.02	.070	84965.45
	Taxonomic levels + Typicality vs. null	.013	270.43	.079	7331.97
	Taxonomic levels + Typicality vs. Taxonomic levels	<.001	0.02	<.001	0.09
B	Taxonomic levels + Typicality + Interaction vs. null	.014	12.81	.079	190.57
	Taxonomic levels + Typicality + Interaction vs. Taxonomic levels + Typicality	<.001	0.05	<.001	0.02
D	Graded Sensorimotor distance vs. null	.018	32605775.72	.027	772.78
	Graded-Sensorimotor + Graded-Linguistic distance vs. null	.017	380788.74	.033	134.29
	Graded-Sensorimotor + Graded-Linguistic distance vs. Graded-Sensorimotor only	<.001	0.02	.005	0.17
E	Best Graded Sensorimotor-Linguistic model vs. best Taxonomic-Typicality model	-	120571.71	-	0.10

Taxonomic levels and typicality. In Analysis A, model comparisons showed evidence *against* adding typicality to a model containing taxonomic levels. In RT analysis, the data strongly favoured the taxonomic-only model (BF_{01} = 40.45) and typicality had very little effect [β = 9.48, 95% CI = ±47.36*,* t (75.05) = 0.39, p = .700]. In accuracy analysis, the data again favoured the taxonomic-only model (BF_{01} = 11.59), with typicality weakly trending in the predicted direction [β =0.59, 95% CI = ±0.66, z = 1.76, p = .080]. In short, typicality ratings did not affect overall categorisation performance in our data, contrary to predictions of some traditional accounts (e.g., Rosch, 1973; Rosch, Simpson, et al., 1976; Rosch & Mervis, 1975).

Interaction of typicality with taxonomic level. Analysis B model comparisons showed strong evidence *against* an interaction between typicality and subordinate taxonomic level. In analysis of RT, the interaction had little effect [$\beta = 46.16$, 95% CI = ±75.71, $t(1812.60) = 1.19$, $p = .230$], and data were $BF_{01} = 21.11$ times more likely under the model without an interaction. In analysis of accuracy, we also found strong evidence against adding the interaction between subordinate taxonomic level and typicality ($BF_{01} = 38.47$), where it had little effect on categorical decision [$\beta = 0.25$, 95% CI = ±0.88, $z = 0.56$, $p = .570$]. The best taxonomic-typicality model in this analysis therefore contained taxonomic levels and typicality ratings, but no interaction. Contrary to taxonomic accounts that hold typicality affects the basic-level advantage (Jolicoeur et al., 1984; Murphy & Brownell, 1985), we found no evidence that atypical items were preferentially categorised at the specific, subordinate level. Rather, all category members showed the traditional basic-level advantage where the basic level was faster and more accurate than subordinate and superordinate levels (but see general discussion).

Linguistic distance and typicality. Analysis C found equivocal evidence for a negative correlation between our original measure of linguistic distance between subordinate items and their basic-level label, and average typicality rating [$r = -.176$, $BF_{10} = 0.79$]. That is, linguistic distance did not decisively capture the graded structure of categories, where more unusual, atypical examples of a category were associated with greater linguistic distance (i.e., less overlap in linguistic contexts) between category and member concept.

Gradedness-adjusted sensorimotor and linguistic distance. In RT, Analysis D model comparisons showed very strong evidence for the effect of gradedness-adjusted sensorimotor distance over a null model of random effects, $BF_{10} = 32605775.72$. However, we found strong evidence *against* adding gradedness-adjusted linguistic distance to a model containing sensorimotor distance alone. That is, as we found for unadjusted distance measures in

Experiment 1, the data were BF_{01} = 15.64 times more likely under a model containing only sensorimotor distance compared to a model containing both linguistic and sensorimotor distance. In the best-fitting sensorimotor-only model, categorisation took up to 244.38 ms longer at the greatest gradedness-adjusted sensorimotor distance (0.29, reflecting the least overlap in sensorimotor experience) than for items at the smallest sensorimotor distance (zero), [β = 842.68, 95% CI = ±235.21, $t(1532.40) = 6.52$, $p < .001$].

In accuracy analysis, we again found very strong evidence for the effect of gradedness-adjusted sensorimotor distance over a null model, BF_{10} = 772.78. However – and unlike Experiment 1 – we also found evidence *against* the inclusion of gradedness-adjusted linguistic distance, whereby the data favoured the model containing sensorimotor distance alone (BF_{01} = 5.75). Hence, as for RT, the best gradedness-adjusted sensorimotor-linguistic model of accuracy was sensorimotor distance alone, where the greatest distance made errors up to 6.90 times[6] more likely than the shortest distance [β =-6.66, 95% CI = ±2.70, z = -4.82, $p < .001$].

Best model. In analysis E, results followed the mixed pattern of Experiment 1 regarding whether the data were more likely under a model using traditional taxonomic levels or sensorimotor/linguistic distance. As predicted, Bayesian model comparisons between the best-performing models from analyses B and D found very strong evidence that gradedness-adjusted sensorimotor distance was BF_{10} = 120571.71 times better at predicting RT than a model containing taxonomic levels and typicality. However, against predictions but consistent with Experiment 1, there was evidence that the taxonomic-typicality model was BF_{01} = 9.49 times better (i.e., BF_{10} = 0.10) at predicting accuracy than a model containing gradedness-adjusted sensorimotor distance.

Summary. In this study, we examined whether accounting for categorical gradedness affected the ability of sensorimotor and linguistic distributional information to contribute to categorical decision making, relative to traditional taxonomic levels. In contrast to our predictions, linguistic distributional information did not decisively capture the graded structure of categories: unusual member concepts were not more linguistically distant from their category concept (e.g., *sweatpants* and *trousers* overlap little in linguistic experience) than member concepts that were good examples of their category (e.g., *jeans* and *trousers* overlap a lot in linguistic experience).

Nevertheless, adjusting for categorical gradedness improved the predictive ability of sensorimotor and linguistic distance measures compared to the unadjusted measures of Experiment 1. Gradedness-adjusted sensorimotor information contributed to categorical decision making, and outperformed traditional predictors of taxonomic levels and typicality in fitting RT (but not accuracy). Against our predictions, however, gradedness-adjusted linguistic distributional information was not an effective predictor of either RT or accuracy. That is, even though linguistic distance formed the basis for adjusting categorical gradedness, when these gradedness-adjusted measures are used, it appears that sensorimotor distance between the category and member concepts is more relevant to the time course and decision outcome than the linguistic distributional relationship between the category and member labels.

Nonetheless, the analysis with typicality produced some unexpected results. Contrary to previous findings in the literature (Rosch, Simpson, et al., 1976; Rosch & Mervis, 1975), typicality did *not* affect categorical decisions, nor did it impact on the basic-level advantage as predicted by the entry-level (Jolicoeur et al., 1984) and differentiation (Murphy & Brownell, 1985) accounts. That is, the least typical items in our dataset tended to be categorised at the basic rather than subordinate level (e.g., a picture of a *sailfish* was categorised as a *fish* more quickly and accurately than as a *sailfish*), the same as the most typicality items (e.g., a *Labrador* as a *dog*). One possible explanation for the absence of typicality effects is that our items did not span a wide enough range of typicality ratings (i.e., range [3.42, 5.00] on a 1-5 scale), and hence were not sufficiently atypical to trigger the mechanisms that should cause unusual category members to be categorised at the subordinate rather than basic level. However, this explanation cannot account for the fact that categorical gradedness *did* affect categorisation when it was modelled via linguistic distance: gradedness-adjusted measures of sensorimotor and linguistic distributional information outperformed the unadjusted measures used in Experiment 1, and gradedness-adjusted sensorimotor distance outperformed taxonomic-typical models in predicting RT. Such a pattern of findings suggests that typicality may not be the best measure of the graded structure of categories, and that the goodness-of-membership may be better captured by an implicit measure derived from distributional patterns in language use. In the next experiments, we examine this possibility, and the robustness of our reported findings, via replication studies.

2.5.3. Exploratory analysis: Psycholinguistic characteristics

As for experiment 1, we ran a set of exploratory analyses to verify that the effects we observed would not be subject to significant change when introducing psycholinguistic characteristics of the category label into the regression models. To this end, we carried out a set of additional mixed effects regressions of RT and accuracy. Analysis F tested whether object typicality predicted RT and accuracy above and beyond the taxonomic level and psycholinguistic characteristics of the category label. To this end we ran two linear mixed effects regressions of RT, with random effects as per analysis A. The first regression included fixed effects of AoA ratings, Zipf log word, familiarity rating, word length in characters excluding spaces (as per exploratory analyses of experiment 1), as well as taxonomic level (dummy-coded as per experiment 1, analysis A). All psycholinguistic characteristics were centred around their respective means. The second regression included an additional fixed effect of mean typicality. Furthermore, we ran two mixed effects logistic regressions of accuracy (binomial, link = logit), with the same fixed and random effects structure as for the analysis of RT. A third regression included the interaction between the subordinate taxonomic level and object typicality. We used Bayesian model comparisons to test whether the data favoured a psycholinguistic-taxonomic-typicality model over a psycholinguistic-taxonomic model.

Analysis G tested whether categorisation at the subordinate versus basic level differed for typical and atypical category members, when controlling for psycholinguistic characteristics of the category label. RT and accuracy were analysed as per Analysis F, with the additional fixed effect of interaction between typicality and subordinate taxonomic level (i.e., where basic level is coded as the reference level). We used Bayesian model comparisons to test whether the data favoured this interaction model over the final model of Analysis F.

Analysis H tested whether gradedness-adjusted sensorimotor and or linguistic distributional information predicted RT and accuracy above and beyond the psycholinguistic characteristics of the category label. To this end, we ran three linear mixed effects regressions of RT with random effects as per analysis A. The first regression included fixed effects of AoA-rating, Zipf log word frequency, average familiarity rating and word length in characters excluding spaces. A second regression added a fixed effect of gradedness-adjusted sensorimotor distance between the category label and implicit image name, while a third regression added a fixed effect of gradedness-adjusted linguistic distributional distance between a category label and implicit image name. We also ran three mixed effects logistic regressions of accuracy (binomial, link = logit) with the same fixed and random effects structure as for the analysis of RT. Bayesian model comparisons tested whether the data favoured a model including sensorimotor and/or linguistic distributional information over that including only psycholinguistic characteristics of the category label.

Finally, in analysis I, we used Bayesian model comparisons to test which model from analysis G and H fit RT and accuracy data best. As in the exploratory analysis for Experiment 1, the data was limited to complete cases only. As in the confirmatory analyses, all trials concerning the item *gavel* were removed. The final RT dataset consisted of 1510 observations; the final accuracy dataset consisted of 1659 observations.

Psycholinguistic vs. taxonomic level and average typicality. In Analysis F of RT, as in the exploratory analyses of Experiment 1, neither a psycholinguistic (BF_{01} = 20.99) nor a psycholinguistic-taxonomic model (BF_{01} = 9.52) outperformed a model containing only random effects of participant and item. Furthermore, model comparisons showed evidence *against* adding typicality to a model containing taxonomic levels and psycholinguistic characteristics of the category label (BF_{01} = 38.45). In the analysis of accuracy, the data did favour both a psycholinguistic (BF_{10} = 346717.70) and a psycholinguistic-taxonomic model (BF_{10} = 2445518.00) over a model containing only random effects. As in or confirmatory analyses, the inclusion of average typicality ratings did not improve model fit over a psycholinguistic-taxonomic model (BF_{01} = 24.96). Inspection of the coefficients showed that accuracy decreased with age of acquisition [β = -0.48, 95% CI = ±0.26, z = -3.66, p <.001] and was lower for superordinate [β = -1.35, 95% CI = ±0.66, z = -4.01, p <.001] but not subordinate [β = -.08, 95% CI = ±0.98, z = -.17, p = .86] compared to basic-level category labels, matching exploratory analyses of Experiment 1.

Interaction between typicality and subordinate level labels. In analysis G of RT, we found evidence against the hypothesis that the addition of an interaction between subordinate category labels and average typicality improved model fit over a psycholinguistic-taxonomic-typicality model (BF_{01} = 34.64). Similarly, in the analysis of accuracy, we found evidence against the addition of the interaction between subordinate category labels and average typicality (BF_{01} = 39.30). These results mirrored the confirmatory findings from experiment 2.

Psycholinguistic characteristics vs. gradedness adjusted distances. In analysis H of RT, we found strong evidence for the inclusion of gradedness-adjusted sensorimotor information to a model containing fixed effects of psycholinguistic characteristics (BF_{10} = 2718619.62). In line with confirmatory analyses, we found evidence against the inclusion of gradedness-adjusted linguistic distributional information to a psycholinguistic-sensorimotor model (BF_{01} = 35.55). Inspection of the coefficients of the psycholinguistic-sensorimotor model once showed that RT increased with age of acquisition [β =28.71, 95% CI = ±17.03, t (292.65) = 3.30, p = .001], word frequency [β =44.89, 95% CI = ±35.79, t(335.66) = 2.46 , p = .014] and gradedness-adjusted sensorimotor distance [β =1021.91, 95% CI = ±324.43, t(478.21) = 6.17 , p < .001]. We note here that a non-psycholinguistic model including only random effects and gradedness-adjusted sensorimotor distance was strongly supported over a psycholinguistic-sensorimotor model (BF_{10} = 783.96).

In analysis H of accuracy meanwhile, we found strong evidence for the inclusion of gradedness-adjusted sensorimotor distance (BF_{10} = 545.55). Moreover, we found decisive evidence against the inclusion of gradedness-adjusted linguistic distributional information (BF_{01} = 18.31). Inspection of the coefficients of the psycholinguistic-sensorimotor model showed that participants' accuracy decreased as age of acquisition [β = -.50, 95% CI = ±.24, z = -4.09, p <.001], word frequency [β = -.64, 95% CI = ±0.58, z = -2.16, p =.031] and gradedness-adjusted sensorimotor distance increased [β = -10.22, 95% CI = ±4.25, z = -4.71, p <.001]. Crucially, this model outperformed a non-psycholinguistic model (BF_{10} = 102944.80).

Best model. In analysis I of RT, Bayesian model comparisons showed that the data were BF_{10} = 47441150.00 times more likely under the best-fitting model from analysis H (sensorimotor-psycholinguistic) than under the best-fitting model from analysis G (taxonomic-typicality- psycholinguistic). In analysis I of accuracy meanwhile, model comparisons showed that the data were $BF_{10} = 1913.99$ times more likely under the best-fitting model from analysis H (sensorimotor-psycholinguistic) than under the best-fitting typicality model from analysis G (taxonomic-typicality psycholinguistic).

Taken together, these exploratory findings are broadly in line with the confirmatory analysis of experiment 2, as well as the exploratory analyses of experiment 1, with two exceptions. Firstly, we found equivocal evidence that a taxonomic-only model predicted accuracy better than a taxonomic-psycholinguistic model, secondly, we found overwhelming evidence that a sensorimotor-psycholinguistic model outperformed a non-psycholinguistic model in predicting accuracy. This suggested that for this data, the best-fitting model included both a measure of gradedness-adjusted sensorimotor distance and psycholinguistic characteristics, in particular age of acquisition and word frequency. Together with the exploratory results in Experiment 1, this strengthens the evidence that age of acquisition may indeed affect categorical decision making.

2.6. Experiment 3a: Replication of Experiment 1

In an attempt to replicate the effects found in Experiment 1 (i.e., sensorimotor and linguistic distance predicted the basic level advantage in categorisation, and outperformed taxonomic level as a predictor in RT but not accuracy), we set out to investigate the same hypotheses in a replication study run online (i.e., via a web-based experimental platform) rather than in the lab. Our predictions remained the same (pre-registration, data, analysis code, and results available at https://osf.io/vdka2/?view_only=ac6c78793a354a1f95da6c36b5ac6163).

2.6.1. Method

The method was identical to Experiment 1 with the following exceptions.

Participants. 25 participants (21 female, M_{age}, 37.76, SD = 10.86) were recruited through Prolific.co (formerly Prolific.ac), an online crowdsourcing platform. Through Prolific's recruitment filter settings, we ensured that all participants were native speakers of English, had no dyslexia, and had a minimum number of 10 submissions with a Prolific approval rating of >95%. Participants were required to achieve an accuracy threshold of 70% accuracy on filler items; no participants were removed for failing to reach this threshold.

Sample size was determined through sequential hypothesis testing of analysis A as per Experiment 1. Since we stopped sampling at our lower bound of N = 30 in Experiment 1, we used sequential analysis to determine the number of lab-based participants at which evidence for the basic-level advantage in RT began to consistently exceed our evidence threshold of $BF_{10} \geq 3$, which occurred at N = 14 (BF_{10} = 71.88). Recent studies (Crump et al., 2013; Hilbig, 2016; Semmelmann & Weigelt, 2017) suggest web-based data collection yields comparable results to lab-based testing, but may still be subject to noise. In order to allow for a higher level of noise in our dataset, we therefore set our present lower bound to be 50% higher at Nmin = 21, and also raised our threshold of evidence to a more conservative $BF_{10} \geq 10$. Due to technical error, an additional 4 participants above Nmin were tested before online recruitment automatically closed; we opted to include all tested participants in data analysis rather than arbitrarily exclude the final four. At 25 participants, the specified grade of evidence for the presence of a basic-level advantage in our data was comfortably cleared at BF_{10} = 14650719.43.

Procedure. The experiment ran on web-based platform Gorilla.sc (Anwyl-Irvine et al., 2020), which handled both collection of informed consent and experimental data collection. Trial presentation was identical to Experiment 1, except the font used to display labels was lowercase Open Sans. Furthermore, in contrast to Experiment 1, we specified a more conservative Bayes Factor threshold at BF ≥ 10, as we expected web-based data to be noisier than lab-based data.

Ethics and consent. The study received ethical approval from the Lancaster University Faculty of Science and Technology Research Ethics Committee. As well as the terms specified in Experiment 1, participants consented to take part on condition that they passed a series of attention checks (i.e., 70% accuracy on filler items as per Experiment 1).

Data analysis. We repeated confirmatory analyses A through C, as well as exploratory analyses D through F as specified in Experiment 1.

2.6.2. Results and discussion

One participant had a very long tail of slow RTs, which indicated inattention but were not otherwise excluded by the pre-registered criteria for outlier removal. As a result, we decided to remove all trials with RT > 10000 ms. We removed 7 trials from the accuracy analysis and 2 trials from the RT analysis. We removed an additional 49 trials from the RT analysis and 57 trials from the accuracy analysis for having RTs that were more than 2.5 SD from the participant's mean. No responses were removed due to motor error. In total we thus removed 51 outliers from the RT analysis (3.07% of 1663 correct responses) and 64 outliers from the accuracy analysis (3.55% of all 1800 responses). Table 3 shows model comparisons for all analyses.

Table 3

Model comparisons for linear mixed effect regressions of RT and logistic mixed effects regressions of accuracy in Experiment 3a showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Analysis	Model comparison	RT		Accuracy	
		ΔR^2	BF_{10}	ΔR^2	BF_{10}
	Null model (random effects)	.197	-	.262	-
A	Taxonomic levels vs. null	.024	14650719.43	.129	9.72×10^{10}
B	Sensorimotor distance vs. null	.022	93175931.44	.048	12088.39
	Sensorimotor + Linguistic distance vs. null	.024	26695351.30	.049	314.19
	Sensorimotor + Linguistic distance vs. Sensorimotor-only	.002	0.28	.001	0.03
C	Best Sensorimotor-Linguistic model vs. taxonomic levels.	-	6.36	-	<0.01

Relationship between dependent variables. An exploratory Bayesian correlation analysis using JASP (v.0.14.0.0, with default beta prior width =1) between a per-item average RT and percentage of correct answers provided anecdotal evidence against their correlation (BF_{01} = 2.01), suggesting that accuracy and response time were not meaningfully associated.

Taxonomic levels. Bayesian model comparisons for Analysis A showed very strong evidence for models containing taxonomic levels over models containing only random effects. In RT, categorisation at the subordinate [unstandardized β = 35.91, 95% CI = ±28.94, t(1528.67) = 2.43, p = .015] and superordinate [β = 105.85, 95% CI = ±30.09, t(1538.60) = 6.89, p <.001] levels was slower than at the basic level. In contrast to the findings from Experiment 1, the basic-level advantage did not appear in accuracy: compared to the basic level, participants were 4.57 times more likely to respond incorrectly at the superordinate level [β = -1.52, 95% CI = ±.53, z = -5.67, p < .001]. However, people were 1.49 times more likely to respond *correctly* at the subordinate level than at the basic level (i.e., in the opposite direction to the predicted basic-level advantage), but with a small effect that in NHST terms was not significant [$\beta = 0.41$, 95% CI = ±.69, $z = 1.16$, $p = .245$]. Predicted probabilities of a correct answer were highest at the subordinate-level at 98.6%, followed by basic-level (97.9%) and finally the superordinate-level (91.2%). In other words, accuracy was approximately equal at the subordinate and basic levels, and worse at the superordinate level. The present study therefore largely but not completely replicates the classic basic level advantage.

Sensorimotor-linguistic predictors. In RT, Analysis B model comparisons showed very strong evidence for the effect of sensorimotor distance over the null model (BF_{10} = 93175931.44), where RT increased with up to 271 ms[6] for the largest sensorimotor distance compared to the smallest one [$\beta = 935.20$, 95% CI = ±272.66, $t\,(1461.16) = 6.72$, $p < .001$]. However, model comparisons did not support an effect of linguistic distance above and beyond sensorimotor distance. Evidence for the sensorimotor-only model was $BF_{01} = 3.49$ times stronger than for the model including linguistic distance, which was below the specified threshold for this experiment ($BF \geq 10$), and hence constitutes equivocal evidence against the inclusion of linguistic distance. We therefore conclude that the best sensorimotor-linguistic model of RT was most likely sensorimotor distance alone (replicating Experiment 1), but acknowledge an additional effect of linguistic distance may still be possible. In accuracy, there was positive evidence for the effect of sensorimotor distance alone, but this time – unlike Experiment 1 – there was strong evidence *against* the inclusion of linguistic distance alongside sensorimotor distance ($BF_{01} = 38.47$). That is, the best sensorimotor-linguistic model of accuracy was sensorimotor distance alone. People were up to 15.47 times[6] more likely to respond incorrectly as sensorimotor distance increased [$\beta = -9.44$, 95% CI = ±3.58, $z = -5.17$, $p < .001$].

As predicted, and replicating Experiment 1, sensorimotor information contributes to categorical decision making. Overlap in sensorimotor experience between category and member concept facilitates categorisation RT and accuracy. However, contrary to what we predicted, and not replicating the results from lab-based testing, we found no positive evidence for the effect of linguistic distributional information.

Best model. Analysis C again showed mixed results as to which model best explained the data. Unlike Experiment 1, where sensorimotor distance outperformed taxonomic levels in explaining RT, model comparisons in the present analysis showed they performed approximately equivalently. The data favoured the model with sensorimotor distance BF_{10} = 6.35 times more than the model including taxonomic levels, which was below the specified threshold for this experiment ($BF \geq 10$), and thus constitutes equivocal evidence. Model comparisons for accuracy showed overwhelming evidence that taxonomic levels were BF_{01} = 8040485.30 times better at fitting the data than sensorimotor distance, against predictions but replicating Experiment 1.

Summary. These results are similar but not identical to the findings of Experiment 1. Nevertheless, they provide further evidence for the effects (or lack thereof) of sensorimotor and linguistic distributional information on picture categorisation. When it comes to the time course of categorical decision making, the overlap in sensorimotor experience between the category and member concepts was at least as good as discrete taxonomic levels (i.e., subordinate vs. basic vs. superordinate) in predicting performance. Linguistic distributional information might have a small effect on RT but the evidence is equivocal. When it comes to the accuracy of categorical decisions, however, taxonomic levels outperform the ability of sensorimotor information to predict performance. Crucially, the datasets for Experiments 1 and 3a yielded slightly different results, to gain a more complete understanding of the effects of sensorimotor and linguistic distance on categorisation performance across both our experiments, we repeat the analyses described here for a combined lab- and web-based dataset further on in this paper (see Experiment 4a).

2.6.3. Exploratory results

As per the exploratory analyses of experiment 1, we could not retrieve information on all category labels used in our task for all psycholinguistic variables. As a consequence, we opted to limited this analysis to complete cases only. In the analysis of accuracy, we included 1392 out of 1736 trials (excluding 19.81 %). In the analysis of RT, we included 1277 out of 1612 correct trials (excluding 20.78 %). Because of this, these exploratory results should be interpreted with care.

Psycholinguistic characteristics vs. taxonomic level. In analysis D of RT, we found strong evidence that a model including psycholinguistic characteristics of the label did not fit the data better than a model including only random effects (BF_{01} = 263.79). However, contrasting the exploratory analysis of lab-based Experiment 1, we found strong evidence that a model including taxonomic information fit the data better than a model containing only psycholinguistic characteristics of the label (BF_{10} = 28849.18). Inspection of the coefficients confirmed no significant effect of AoA, word frequency, familiarity or word length. As in our confirmatory analysis (A) participants were up to 87.74 ms slower to categorise objects following subordinate level labels [β = 87.74, 95% CI = ±57.06, t (842.09) = 3.01, p =.003], and up to 89.30 ms slower to categorise objects following a superordinate-level label [β = 111.04, 95% CI = ±38.20, t (1197.77) = 5.70, p < .001] compared to a basic-level label, showing the classic basic-level advantage. As in our previous exploratory analyses, we found evidence that a taxonomic-only model greatly outperformed a psycholinguistic-only model (BF_{10} = 3864454127.00)

In analysis D of accuracy, we found evidence that a model including psycholinguistic characteristics of the category label outperformed a model containing only random effects

(BF_{10} = 58.40). Furthermore, we found evidence that a model including taxonomic levels outperformed a psycholinguistic model (BF_{10} = 196.87). Inspection of the coefficients showed that increasing age of acquisition associated with the category label reduced the likelihood that participants categorised an image correctly by up to 11.94 times[6] [β = -.39, 95% CI = ±.37, z = -2.04, p = .041]. In line with our confirmatory analyses (see analysis A above) and the exploratory analysis of Experiment 1, we observed only a partial basic-level advantage. That is, compared to the basic level, participants were less likely to correctly categorise an image when it followed a superordinate-level label [β = -1.48, 95% CI = ±.68, z = -4.23, p <.001]. However, accuracy was non-significantly *better* at the subordinate compared to the basic level [β = .72, 95% CI = ±1.35, z = 1.05, p =.29]. Unlike our exploratory analyses however, this effect was also observed in a non-psycholinguistic taxonomic model of the truncated dataset.

Psycholinguistic characteristics vs. sensorimotor-linguistic information. In analysis E of RT, we found evidence for the inclusion of sensorimotor distance to a model containing random effects and psycholinguistic characteristics (BF_{10}= 80000.29). However, in contrast to our confirmatory analysis B, we also found strong evidence for the inclusion of linguistic distributional information (BF_{10} = 74.65). Inspection of the coefficients for this model confirmed no significant effects of any of the psycholinguistic variables. Participants were up to 180.99 ms slower to categorise an object with a greater sensorimotor [β = 624.12, 95% CI = ±392.24, t(543.38) = 3.12, p =.002] distance and up to 138.52 ms slower with greater linguistic distributional distance [β = .72, 95% CI = ±99.42, t(1038.63) = 4.02, p <.001] between the category label and implicit image name.

In analysis E of accuracy, we found equivocal evidence for the inclusion of sensorimotor distance to a model containing random effects and psycholinguistic characteristics (BF_{10}= 2.27). Furthermore, we found equivocal evidence for the inclusion of linguistic distributional information to a sensorimotor-psycholinguistic model ($BF_{10} = 1.26$). This contrasted our confirmatory analyses, where evidence suggested the data strongly favoured a sensorimotor model.

In analysis F of RT, Bayesian model comparisons showed that the data were $BF_{10} = 207.03$ times more likely under the best-performing model from analysis E (psycholinguistic-sensorimotor-linguistic) than under the best-performing model from analysis D (psycholinguistic-taxonomic). By contrast, analysis F of accuracy showed that the data were more $BF_{10} = 69.02$ times more likely under the best-performing analysis D model (psycholinguistic-taxonomic) than under the best-performing analysis E model (psycholinguistic-sensorimotor-linguistic).

These findings showed no significant effects of psycholinguistic characteristics on the time course in object categorisation, but like our previous exploratory analyses, we observed a small effect of age of acquisition on the likelihood that participants correctly categorised an object. Unlike the exploratory analyses of our lab-based data, the inclusion of psycholinguistic characteristics did not reduce the effect of linguistic distributional information relative to that of sensorimotor-linguistic information. In fact, unlike our confirmatory analyses, the inclusion of linguistic distributional information to a psycholinguistic-sensorimotor model of RT improved model fit. However, given the fact that this effect was also observed in non-psycholinguistic models, this may have been the result of truncation of the dataset.

2.7. Experiment 3b: Exploratory Replication of Experiment 2.

In Experiment 2, we found that adjusting measures of sensorimotor and linguistic distance to reflect the graded structure of categories considerably increased their ability to predict categorisation performance. Unexpectedly, however, typicality ratings (i.e., the traditional measure of categorical gradedness) did not alter the ability of taxonomic levels to predict categorisation. That is, while there was no evidence that *atypical* member concepts (e.g., relatively poor examples of their category) were categorised at the subordinate (rather than basic) level, there was indeed evidence that *unusual* category members (i.e., that appear in quite dissimilar linguistic contexts to their category concept) were categorised at the specific, subordinate level.

We therefore opted to run an exploratory (i.e., not pre-registered)[8] replication of Experiment 2, using the new dataset collected in Experiment 3a. That is, we investigated the same hypotheses as Experiment 2, based on analysis of the RT and accuracy data collected as part of Experiment 3a.

2.7.1. Method

All materials, critical predictors, and data analysis, including exploratory analyses, were identical to Experiment 2. Dependent measures of RT and accuracy were taken from Experiment 3a, as well as its threshold for inferring of $BF \geq 10$.

2.7.2. Results and discussion

With outliers already removed in Experiment 3a, we further removed all trials corresponding to the item *gavel* as per Experiment 2, which excluded 24 trials from the accuracy dataset (1.38% of 1736 trials) and 17 trials from the RT dataset (1.05% of 1611 trials).

With this item excluded, the pattern of basic-level advantage remained the same as in Experiment 3a: there was very strong evidence for including taxonomic levels in analysis of both RT ($BF_{10} = 13936195.41$) and accuracy ($BF_{10} = 4.84 \times 10^9$). RT showed the classic basic-level advantage, with categorisation at the basic level faster than at subordinate

[8] Since there were no new data to collect to run this analysis, we opted not to pre-register it and instead included it as an exploratory analysis.

[unstandardized β = 37.28, 95% CI = ±29.23, $t(1512.22) = 2.50$, $p = .012$] and superordinate levels [β = 106.66, 95% CI = ±30.31, $t(1521.65) = 6.90$, $p <.001$]. Accuracy, however, was equally good at the basic and subordinate levels [β =0.45, 95% CI = ±0.71, $z = 1.24$, $p = .22$], but worse at the superordinate level [β = –1.46, 95% CI = ±0.53, $z = –5.36$, $p < .001$].

Table 4 shows all model comparisons.

Taxonomic levels and typicality. In RT, Analysis A model comparisons showed strong evidence *against* adding typicality to a model containing taxonomic levels ($BF_{01} = 40.45$) as typicality had very little effect [β = -1.00, 95% CI = ±46.24, $t\ (71.64) =-.04$, $p = .97$]. In accuracy, the data also favoured the taxonomic-only model ($BF_{01} = 21.11$), with again very little effect of typicality [β =0.49, 95% CI = ±0.79, $z = 1.21$, $p = .22$]. These results replicated our findings from Experiment 1, and do not support predictions of traditional accounts of categorical gradedness based on typicality (e.g., Rosch, Simpson, et al., 1976) .

Table 4

Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy Experiment 3b, showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Analysis	Model comparison	RT		Accuracy	
		ΔR^2	BF_{10}	ΔR^2	BF_{10}
	Null model (random effects)	.196	-	.257	-
A	Taxonomic levels vs. null	.024	13936195.41	.126	4.84×10^9
	Taxonomic levels + Typicality vs. null	.024	344551.90	.132	2.29×10^8
	Taxonomic levels + Typicality vs. Taxonomic levels	<.001	0.02	.007	.05
B	Taxonomic levels + Typicality + Interaction vs. null	.025	23155.80	.150	14650719.43
	Taxonomic levels + Typicality + Interaction vs. Taxonomic levels + Typicality	.001	.07	.017	0.06
D	Graded-Sensorimotor distance vs. null	.018	268337.29	.084	1.87×10^9

	Graded-Sensorimotor + Graded-Linguistic distance vs. null	.018	8103.08	.089	59411596.94
	Graded-Sensorimotor + Graded-Linguistic distance vs. Graded Sensorimotor-only	<.001	.03	.005	.03
E	Best Graded Sensorimotor-Linguistic model vs. best Taxonomic- typicality model.	-	.78	-	8.17

Interaction of typicality with taxonomic level. Analysis B model comparisons again showed strong evidence *against* including an interaction between subordinate taxonomic level and typicality. In the analysis of RT, the interaction had little effect [β = 52.37, 95% CI = ±71.91, t(1521.83) = 1.42, p = .15], and the data were BF_{01} = 14.88 times more likely under the model *without* an interaction. In the analysis of accuracy, we again found evidence *against* adding the interaction between subordinate taxonomic level and typicality [β =1.03, 95% CI = ±, z = 1.38, p = .17] to a model including only taxonomic level and typicality (BF_{01} = 15.64). The best taxonomic-typicality model in this analysis therefore contained both taxonomic levels and typicality, but no interaction. These results replicate the findings from Experiment 2, and run contrary to predictions of taxonomic accounts that hold typicality affects the basic-level advantage (Jolicoeur et al., 1984; Murphy & Brownell, 1985).We therefore found no evidence that atypical items were preferentially categorised at the subordinate level. Instead, regardless of typicality, category members tended to show a traditional basic-level advantage in terms of processing speed (where the basic level was faster than subordinate and superordinate levels), and a partial basic-level advantage in terms of accuracy, where the basic-level was more accurate than the superordinate level but not the subordinate level.

Gradedness-adjusted sensorimotor and linguistic distance. In RT, Analysis D model comparisons showed very strong evidence for the effect of gradedness-adjusted sensorimotor distance over a null model (BF_{10} = 268337.29). However, as in Experiment 2, we found strong evidence *against* adding gradedness-adjusted linguistic distance to a model containing sensorimotor distance alone: the data were BF_{01} = 33.11 times more likely under a model containing only sensorimotor distance compared to a model containing both linguistic and sensorimotor distance. In accuracy analysis, we again found very strong evidence for the effect of gradedness-adjusted sensorimotor distance over a null model (BF_{10} > 1.87 billion), and evidence *against* the inclusion of gradedness-adjusted linguistic distance, whereby the data strongly favoured the model containing sensorimotor distance alone (BF_{01} =31.50). This again contrasted our findings with the unadjusted measures of sensorimotor and linguistic distance in Experiment 1. Hence, the best gradedness-adjusted sensorimotor-linguistic model of both RT and accuracy was one containing the fixed effect of sensorimotor distance alone, where higher distance slowed down categorisation by up to 216.80 ms[6] [β = 747.58, 95% CI = ±255.32, t(1424.47) = 5.74, p <.001] and made errors up to 31.90 times[6] more likely [β =-11.94, 95% CI = ±3.37, *z* = -6.94, *p* <.001].

Best Model. In RT, model comparisons between the best-performing models from analyses B and D showed equivalent performance between a taxonomic-typicality model (BF_{01} = 1.28) and a model containing gradedness-adjusted sensorimotor distance, contrasting our findings in experiment 2. We note here that since typicality had little effect on performance, the data overall favoured a model including *only* taxonomic levels (i.e., the best model from Analysis A): when we explored its performance in comparison to the gradedness-adjusted sensorimotor model, we found it performed BF_{01} = 51.93 times better.

In accuracy, model comparisons between the best-performing models from analyses B and D showed a weak reversal of the effects found in experiment 2, with evidence trending in favour of gradedness-adjusted sensorimotor distance (BF10 = 8.17). However, this level of evidence is below the threshold for this study (BF≥10), and thus constitutes equivocal evidence). Again, since typicality had little effect on accuracy, we explored a comparison with the best model from Analysis A (i.e., taxonomic levels only), and found that taxonomic levels and gradeness-adjusted sensorimotor distance performed equally well (BF_{10}=0.39)

Summary. Results of this exploratory replication were similar but not identical to the findings of Experiment 2. As predicted, sensorimotor distance contributed to both RT and accuracy. However, counter to our predictions, linguistic distance did not. Both findings are consistent with Experiment 2. Not consistent with our findings in experiment 2 was the finding that sensorimotor distance and taxonomic level performed equally well in explaining both accuracy and RT. That is, sensorimotor distance between a category and member concept explains the latency and accuracy of picture categorisation at least as well as the traditional combination of taxonomic levels and typicality. While this does not support our prediction that sensorimotor distance would outperform taxonomic levels and typicality, it does also not reject it. Finally, we note that taxonomic level alone outperformed sensorimotor distance in explaining RT data. This finding was not related to a specific prediction, and as a result cannot be compared to a related finding in Experiment 2.

2.7.3. Exploratory analysis: Psycholinguistic characteristics

As for experiment 1, 2 and 3a, we ran a set of exploratory analyses to verify that the effects we observed would not be different when controlling for psycholinguistic characteristics of the category label. To this end, we carried out analyses F, G, H and I as outlined in the exploratory analyses of Experiment 2, on the dataset collected in experiment 3a. As in previous exploratory analyses, we restricted the data to observations for which psycholinguistic characteristics were available, as per the confirmatory analyses in experiment 2 and 3b, we removed all trials pertaining to the item *gavel*. The final RT dataset contained 1260 observations, whereas the final accuracy dataset contained 1368 observations.

Psycholinguistic vs. taxonomic level and average typicality. In Analysis F of RT, model comparisons showed that a model containing psycholinguistic characteristics did not fit the data better than a model containing random effects (BF_{01} = 318.20). As before, the inclusion of taxonomic level improved model fit over a psycholinguistic-only (BF_{10} = 37307.53). and random-effects model (BF_{10} = 116.97). Furthermore, we found evidence *against* adding typicality to a model containing taxonomic levels and psycholinguistic characteristics of the category label, with comparisons suggesting the data was BF_{01} = 32.85 times more likely under the psycholinguistic-taxonomic model.

In analysis F of accuracy, model comparisons showed that a psycholinguistic model did outperform random effects (BF_{10} = 2228.18). However, we again found evidence against the inclusion of object typicality to a model containing psycholinguistic characteristics and taxonomic levels (BF_{01} = 21.29).

Interaction between typicality and subordinate level labels. In analysis G of RT, we found evidence against the addition of an interaction between subordinate category labels and average typicality improved model fit to a psycholinguistic-taxonomic-typicality model (BF_{01} = 21.40). In the analysis of accuracy, we found that the data were less likely under a model including the interaction between subordinate category labels and average typicality (BF_{01} = 4.45) than under a non-interaction model, however because this falls below the specified threshold of BF_{10} >= 10, this constitutes equivocal evidence.

Psycholinguistic characteristics vs. gradedness adjusted distances. In analysis H of RT, we found strong evidence for the inclusion of gradedness-adjusted sensorimotor information to a model containing fixed effects of psycholinguistic characteristics (BF_{10} = 261.37). In line with confirmatory analyses, we found evidence against the inclusion of gradedness-adjusted linguistic distributional information to a psycholinguistic-sensorimotor model (BF_{01}= 35.16).

In analysis H of accuracy meanwhile, we found strong evidence for the inclusion of gradedness-adjusted sensorimotor distance (BF_{10} = 157.87). Furthermore, we found evidence *against* the inclusion of gradedness-adjusted linguistic distributional distance ($BF_{01} = 35.36$). This matches confirmatory analyses.

Best model. In analysis I of RT, Bayesian model comparisons showed that the data were $BF_{10} = 4.34$ times more likely under the best-fitting model from analysis G (taxonomic-typicality- psycholinguistic) than under the best-fitting model from analysis H (sensorimotor-psycholinguistic). However, since this falls below the grade of evidence of we specified for this experiment ($BF_{10} >= 10$), this constitutes equivocal evidence.

In analysis I of accuracy meanwhile, model comparisons showed that the data were $BF_{10} = 4707.48$ times more likely under the best-fitting model from analysis H (sensorimotor-psycholinguistic) than under the best-fitting model from analysis G (taxonomic-typicality psycholinguistic). This contrasts confirmatory findings, where evidence for the performance of a model including gradedness-adjusted sensorimotor distance over a taxonomic-typicality model was in the equivocal zone.

2.8. Experiment 4a: Combined Analysis for Experiments 1 and 3a.

As pre-registered, we combined the datasets of Experiment 1 and 3a in order to examine whether the pattern of findings was consistent across experiments. That is, to examine the cross-experiment effects of taxonomic level (i.e., basic-level advantage) and sensorimotor and linguistic distance on accuracy and RT (i.e., a representational overlap advantage).

2.8.1. Data analysis

We combined the datasets analysed in Experiments 1 and 3a, resulting in 3517 responses for analysis of RT, and 3833 responses for analysis of accuracy.

We then ran Experiment 1’s analyses A through C for the combined dataset, with one change: a fixed effect of Experiment was included in all models immediately following the addition of random effects. In addition to this, we ran additional models for the interactions between the fixed effects of taxonomic level and sensorimotor and linguistic distance with Experiment. The experiment variable was dummy-coded with lab-based Experiment 1 = 0, web-based Experiment 3 = 1. All other analyses proceeded as outlined in Experiment 1.

2.8.2. Results and discussion

Table 5 shows all model comparisons.

Experiment effect (web-based versus lab-based). Bayesian model comparisons showed strong evidence against an effect of Experiment (i.e., web- vs. lab-based testing paradigm) on either RT (BF_{01} = 28.51, [β = 47.44, 95% CI = ±75.46, $t(55.00)$ =1.23, p =.22] or accuracy (BF_{01} = 28.51, [β = 0.31, 95% CI = ±.49, z = 1.28, p = .20]).

Table 5

Model comparisons for linear mixed effect regressions of RT and logistic mixed effects regressions of accuracy in Experiment 4a (web- and lab-based taxonomic) and 4b (web- and lab-based typicality).

Analysis	Model comparison	RT		Accuracy	
		ΔR^2	BF_{10}	ΔR^2	BF_{10}
	Null model (random effects)	.244	-	.265	-
	Experiment type vs. null	.006	0.03	.005	0.03
A	Taxonomic levels vs. null	.022	$4.81\text{x}10^{11}$	.073	$9.69\ \text{x}10^{11}$
	Taxonomic levels + Interaction Experiment type and subordinate level vs. Taxonomic levels	.023	0.02	.104	2440.62
	Taxonomic levels + Interaction Experiment type and subordinate level + Interaction Experiment type and superordinate level vs. Taxonomic levels	.023	<0.01	.105	44.70
B	Sensorimotor distance vs. null	.023	$7.94\text{x}10^{11}$	.032	18033.74
	Sensorimotor + Linguistic distance vs. null	.023	$1.02\text{x}10^{11}$	.044	31257.04
	Sensorimotor + Linguistic distance vs. Sensorimotor-only	<.001	0.13	.012	1.73
	Sensorimotor + Linguistic distance + Interaction Sensorimotor and Experiment vs. Sensorimotor + Linguistic distance.	.024	0.02	.003	0.04
	Sensorimotor + Linguistic distance + Interaction Sensorimotor distance and Experiment + Interaction Linguistic distance and Experiment vs. Sensorimotor + Linguistic distance.	.024	<0.01	.002	0.01
C	Best Sensorimotor-Linguistic model vs. taxonomic levels.	-	1.65	-	<0.01

Taxonomic levels. In RT, Analysis A model comparisons showed very strong evidence for models containing taxonomic levels over a model containing only random effects (BF_{10} = 4.81x10^{11}). Categorisation at the subordinate [unstandardized β = 36.59, 95% CI = ±20.68, t(3399.17) = 3.47, p <.001] and superordinate [β = 94.82, 95% CI = ±21.18, t(3409.42) = 8.78, p <.001] levels was slower than at the basic level. Likewise, in accuracy, model comparisons showed strong evidence for the inclusion of taxonomic levels (BF_{10} =

9.69×10^{11}). However, we also found strong evidence for including the interaction between the dummy-coded subordinate taxonomic level and Experiment type over a model including only taxonomic levels and Experiment type. Accuracy was high, predicted probabilities of a correct answer were highest at the basic level (97.8%), followed by the subordinate (96.2%) and the superordinate level (91.5%). Compared to the basic level, participants were 4.05 times more likely to respond incorrectly at the superordinate level [β = -1.40, 95% CI = ±.33, z = -8.28, p < .001]. While experiment type had little effect [β = -0.07, 95% CI = ±.52, z = -0.27, p =.787], the interaction between Experiment type and subordinate category labels meant that participants were up to 2.91 times more likely to respond incorrectly at the subordinate level in the lab-based study, [β = -1.07, 95% CI = ±.36, z = -5.32, p <.001], but slightly more likely to answer *correctly* at the subordinate level in the web-based study, [β = 1.55, 95%CI = ±.64, z = 4.75, p <.001]. In the analysis of RT, combining both datasets yields a clear replication of the basic-level advantage. In accuracy, the basic-level has a consistent advantage over superordinate but not subordinate categories.

Sensorimotor-linguistic predictors. In RT, Analysis B model comparisons showed very strong evidence for the effect of sensorimotor distance over the null model (BF_{10} = 1.02×10^{11}), Model comparisons did not support an effect of linguistic distance above and beyond sensorimotor distance, in line with the outcomes from experiment 3a: evidence for the sensorimotor-only model was BF_{01} = 7.76 times stronger than for the model including linguistic distance, which was below the specified threshold for this experiment (BF ≥ 10), and therefore constitutes equivocal evidence. Neither a model including the interaction between sensorimotor distance and Experiment, nor a model including the interaction between sensorimotor and linguistic distance and Experiment outperformed the sensorimotor-linguistic model (see Table 5). RT increased with up to 217 ms for the largest sensorimotor distance[6], compared to the smallest one [β = 749.93, 95% CI = ±220.28, t(2647.86) = 6.67, p

<.001], and with up to 39 ms for the largest linguistic distance compared to the smallest one [β = 47.34, 95% CI = ±75.89, t(3143.00) = 1.22, p =.043]. We conclude from this combined analysis that while picture categorisation RT is robustly explained by sensorimotor information, there is a possibility that linguistic distributional information plays a weaker role.

In accuracy, model comparisons showed strong evidence for the effect of sensorimotor distance alone (BF_{10} =18033.74). However, unlike our findings in Experiment 1, comparisons showed equivocal evidence for the inclusion of linguistic distance alongside sensorimotor distance (BF_{10} = 1.73). Again, neither a model including the interaction between sensorimotor distance and experiment type, nor a model including the interactions between sensorimotor and linguistic distance and experiment type outperformed the sensorimotor-linguistic model (see Table 5). Evidence for the sensorimotor-linguistic model over a sensorimotor-only model did not surpass the specified threshold of (BF ≥ 10), making the evidence for linguistic distance equivocal. Participants were up to 5.19 times more likely to respond incorrectly as sensorimotor distance increased [β = -5.74, 95% CI = ±2.49, z =-4.53, p <.001], and up to 2.39 times more likely to respond incorrectly as linguistic distance increased [β = -1.07, 95% CI = ±.66, z =-3.16, p <.001]. We conclude from this combined analysis that picture categorisation accuracy is robustly predicted by sensorimotor information, and predicted somewhat equivocally by linguistic distributional information where effects range from very strong (Experiment 1) to non-existent (Experiment 3a).

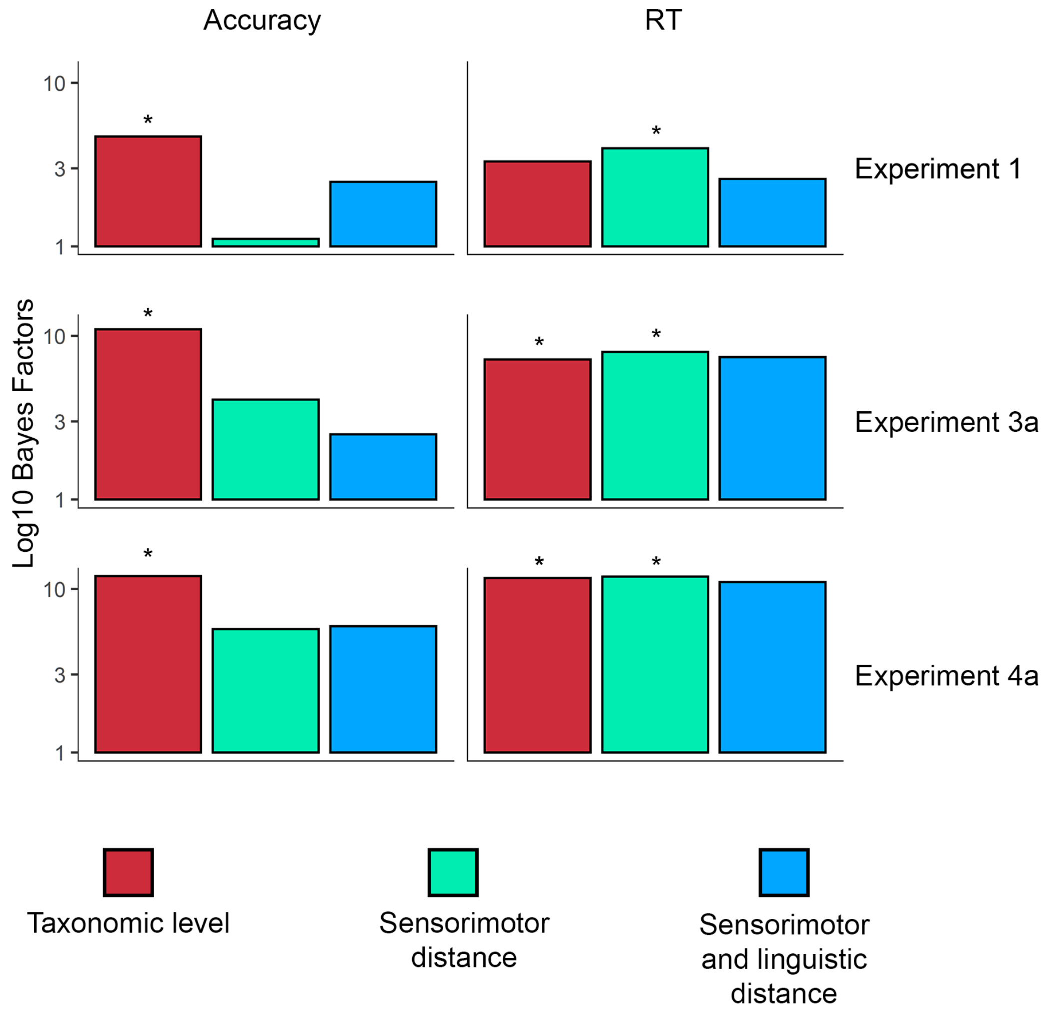
Best model. Analysis C showed that sensorimotor distance and taxonomic levels predicted RT equally well. The data favoured a model with sensorimotor distance BF_{10} = 1.65 times more than a model with taxonomic levels, which constitutes equivocal evidence. In accuracy, evidence overwhelmingly showed that taxonomic levels were BF_{01} = 1982759.26

times better at fitting the data than sensorimotor distance, against our predictions but consistent with Experiment 1.

Summary. A combined analysis Experiments 1 and 3a overall partially supported our hypotheses that sensorimotor and linguistic information would inform categorisation (see figure 1). As predicted, sensorimotor distance between a category- and member-concept informed the time course of categorisation, and did so at least as well as taxonomic level across two experiments. However, counter to our prediction, linguistic distributional distance did not predict RT above and beyond sensorimotor distance. Moreover, while linguistic distributional distance predicted accuracy at least as well as sensorimotor distance across two experiments, neither distance measure outperformed a simple division into taxonomic levels. Finally, we found that web-based participants were more accurate at categorising objects at the subordinate compared to the basic level.

Figure 1.

Bayes factors for models of accuracy and RT compared to null (random effects), for Experiments 1, 3a and 4a.



Note: best-fitting models are marked with an asterisk. In the case of equivocal performance between two models both models are marked. Experiment 4a denotes the analysis of a combined web- and lab-based dataset rather than data from a new experiment.

2.9. Experiment 4b: Combined Analysis for Experiments 2 and 3b

As an exploratory analysis, we combined the datasets of Experiment 2 and 3b, in order to examine whether the pattern of findings was consistent across experiments. That is, to explore the cross-experiment effects of object typicality and its interaction with subordinate-level category labels, as well as of gradedness-adjusted sensorimotor and linguistic distance on RT and accuracy.

2.9.1. Data analysis

We combined the datasets analysed in Experiments 2 and 3b, resulting in 3479 responses for analysis of RT, and 3781 responses for the analysis of accuracy. We ran Experiment 2's analyses A, B, D and E this the combined dataset, adding a fixed effect of experiment, immediately following the addition of random effects (coded as per Experiment 4a). In addition to this, we added the interactions between all fixed effects and Experiment type. All other analyses proceeded as outlined in experiment 2, except we retained the stricter Bayes factor threshold for inferring from Experiment 3b ($BF \geq 10$).

2.9.2. Results and discussion

Table 6 shows all model comparisons.

Experiment effect (web-based versus lab-based). As in Experiment 4a, Bayesian model comparisons showed strong evidence against an effect of experiment type (i.e., web- vs. lab-based testing paradigm) on either RT (BF_{01} =25.79, [β =50.19, 95% CI = ±75.17, $t(55.01)$ =1.31, p =.20]) or accuracy (BF_{01} =23.33, [β = .35, 95% CI = ±.48, z = 142, p = .16]).

Table 6.

Model comparisons for linear mixed effect regressions of RT and logistic mixed effects regressions of accuracy in Experiment 4b.

Analysis	Model comparison	RT		Accuracy	
		ΔR^2	BF_{10}	ΔR^2	BF_{10}
	Null model (random effects)	24.44	-	.256	-
	Experiment type vs. null	.007	0.04	.007	0.04
A	Taxonomic levels vs. null	.024	3.56×10^{12}	.075	2.06×10^{11}
	Taxonomic levels + Typicality vs. null	.024	6.20×10^{10}	.081	8.82×10^{9}
	Taxonomic levels + Typicality + Interaction subordinate and Experiment type vs. null	.024	1386432863.11	.114	1.44×10^{13}
	Taxonomic levels + Typicality + Interaction subordinate and Experiment type + Interaction superordinate and Experiment type vs. null	.025	36034955.09	.114	2.39×10^{11}
	Taxonomic levels + Typicality + Interaction subordinate and Experiment type + Interaction superordinate and Experiment type + Interaction Typicality and Experiment type vs. null	.025	660003.22	.113	4378622438.03
	Taxonomic levels + Typicality vs. Taxonomic levels	<.001	0.02	.006	0.04
B	Taxonomic levels + Typicality + Interaction vs. null	.025	5.62×10^{9}	.082	2.41×10^{8}
	Taxonomic levels + Typicality + Interaction + Interaction subordinate and Experiment type + Interaction superordinate and Experiment type + Interaction Typicality and Experiment type vs. null	.025	56954.05	.115	139002155.75
	Taxonomic levels + Typicality + Interaction vs. Taxonomic levels + Typicality	<.001	0.09	<.001	0.03
D	Graded Sensorimotor distance vs. null	.019	4.32×10^{14}	.054	1.08×10^{10}
	Graded Sensorimotor + Linguistic distance vs. null	.020	3.73×10^{13}	.055	3.97×10^{8}
	Graded Sensorimotor + Linguistic distance vs. Graded Sensorimotor-only	.001	0.09	<.001	0.04
	Graded Sensorimotor + Linguistic distance + interaction Graded Sensorimotor and Experiment Type vs. Graded Sensorimotor + Linguistic Distance	.027	3.24×10^{10}	.059	<0.01

	Graded Sensorimotor + Linguistic distance + interaction Graded Sensorimotor and Experiment Type + interaction Linguistic distance and Experiment Type vs. Graded Sensorimotor + Linguistic distance	.027	840914039.1	.059	<0.01
E	Best graded Sensorimotor Linguistic model vs. best Taxonomic- typicality model.	-	6974.39	-	1.22

Taxonomic levels and typicality. Bayesian model comparisons showed strong evidence for including taxonomic levels in analysis of both RT ($BF_{10} = 3.56x10^{12}$), and accuracy ($BF_{10} = 2.06x10^{11}$). However, in accuracy, as in Experiment 4a, we also found strong evidence for including the interaction between the dummy-coded subordinate taxonomic level and experiment type over a model including only taxonomic levels and experiment type. **In RT, results matched our findings in Experiments 2 and 3b, and categorisation at the basic level was faster than at both the subordinate** [unstandardized $\beta = 40.14$, 95% CI = ±20.71, $t(3362.51) = 3.80$, $p = .00015$] and superordinate [unstandardized $\beta = 97.49$, 95% CI = ±21.18, $t(3372.55) = 9.02$, $p <.001$] levels. In accuracy, results mirrored the findings in Experiments 2 and 3b. That is, the interaction between subordinate-level category labels and experiment type meant that for lab-based data accuracy was worse at both the subordinate [$\beta = –.57$, 95% CI = ±0.37, $z = –3.04$ $p =.002$] *and* the superordinate level [$\beta = –1.41$, 95% CI = ±0.34, $z = –8.10$, $p < .001$] compared to the basic level, whereas for web-based data accuracy was slightly better at the subordinate compared to the basic level [$\beta = 1.50$, 95% CI = ±0.82, $z = 3.57$, $p < .001$]. The effect of superordinate level was the same regardless of experiment type. In summary, as for Experiment 4a, analyses show a classic basic-level advantage in RT and lab-based, but not web-based, accuracy.

In analysis A, we found strong evidence *against* adding typicality to a model containing taxonomic levels and experiment, with little effect of average typicality ratings on either RT ($BF_{01} = 57.80$, [$\beta = -3.53$, 95% CI = ±41.34, $t(72.06) =.17$, $p = .87$]) or accuracy

(BF_{01} =23.34, [β =0.44, 95% CI = ±0.60, z = 1.43, p = .15]). In addition to this, we found no evidence for an interaction between typicality and experiment type (see Table 6). These results replicate our findings from Experiments 2 and 3b, and provide no support for an overall effect of categorical gradedness.

Interaction of typicality with taxonomic level. In RT, analysis B model comparisons showed strong evidence *against* including an interaction between subordinate taxonomic level and typicality, the data were BF_{01} = 11.59 times more likely under the model without an interaction. In the analysis of accuracy, we found further evidence against adding the interaction between subordinate taxonomic level and typicality [β =.39, 95% CI = ±.72, z = 1.05, p = .29] to a model containing taxonomic level and typicality (BF_{01} = 36.60). The best taxonomic-typicality model in this analysis therefore once more contained both taxonomic levels and typicality ratings, but no interaction. This replicates the findings from Experiment 2 and 3b and further contrasts with taxonomic accounts that hold typicality moderates the basic-level advantage by shifting to subordinate levels for atypical items (Jolicoeur et al., 1984).

Gradedness-adjusted sensorimotor and linguistic distance. In RT, analysis D model comparisons for both datasets combined showed very strong evidence for the effect of gradedness-adjusted sensorimotor distance over a null model. As in Experiment 2 and 3b, we found evidence *against* adding gradedness-adjusted linguistic distance to a model containing sensorimotor distance alone; the data were BF_{01} = 11.59 times more likely under a model containing only sensorimotor distance compared to a model containing both linguistic and sensorimotor distance. We found no evidence for an interaction between adjusted sensorimotor and or linguistic distance and experiment type (see Table 6). Categorisation took up to 237.56 ms longer[2] as sensorimotor distance increased [β = 819.17, 95% CI = ±183.67, t(2926.13) = 8.74, p <.001].

In accuracy, we again found strong evidence for the effect of gradedness-adjusted sensorimotor distance over a null model (BF_{10}=1.08x10^{10}). As in Experiment 2 and 3b, we found evidence *against* the inclusion of gradedness-adjusted linguistic distance, whereby the data strongly favoured the model containing sensorimotor distance alone (BF_{01} = 27.11). Here too, we found no evidence for an interaction between adjusted sensorimotor and or linguistic distance and experiment type (see Table 6). As a result, the best gradedness-adjusted sensorimotor-linguistic model of both RT and accuracy was one containing the fixed effect of sensorimotor distance alone. Greater sensorimotor distance reduced accuracy up to 12.83 times [2] [β =-8.80, 95% CI = ±2.14, z = -8.05, p <.001]. This contrasted with our findings regarding unadjusted measures of sensorimotor and linguistic distance in Experiment 1, where the data favoured a model including both distance measures, and the analysis of web- and lab-based data in Experiment 4a, where evidence for both distance measures over sensorimotor distance alone was equivocal.

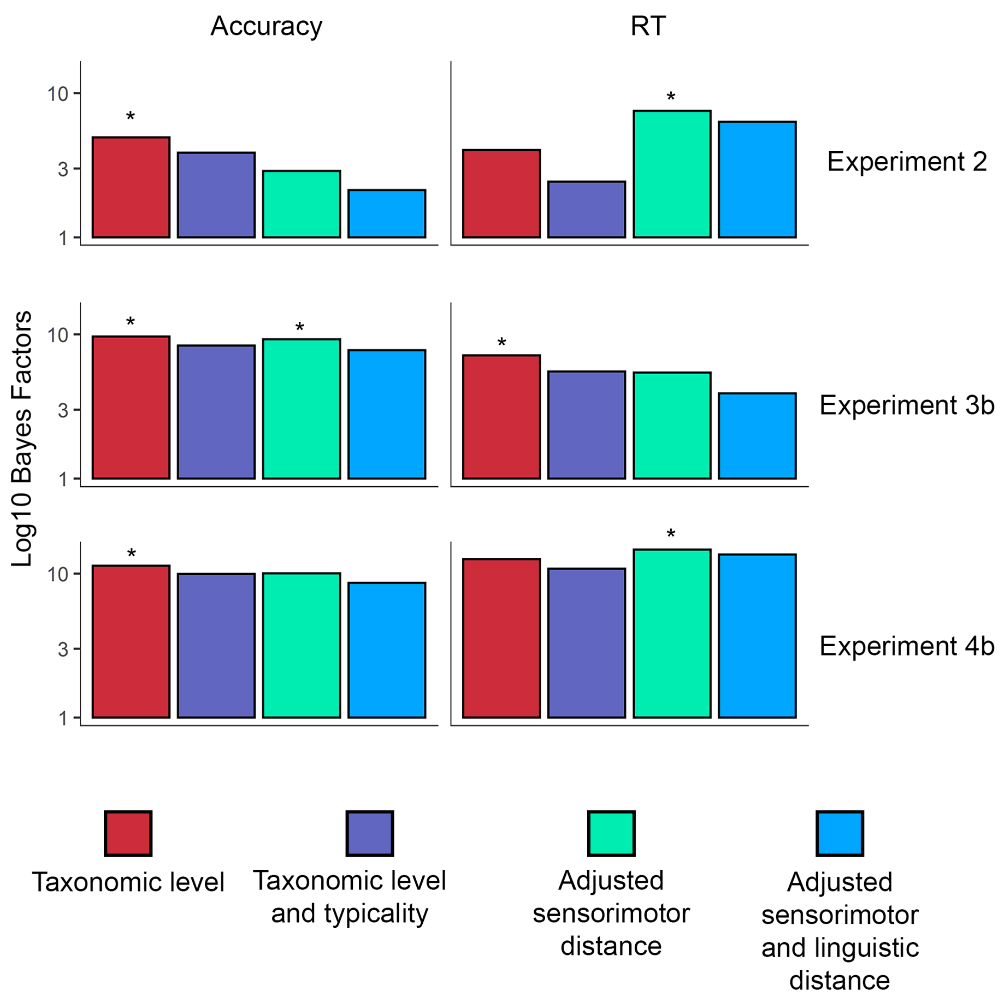
Best model. In analysis E for RT, model comparisons between the best performing models from analyses B and D showed strong evidence in favour of an adjusted sensorimotor model (BF_{10} = 601.84) over a model including taxonomic levels and typicality. We note here that the evidence grade for a model of adjusted sensorimotor distance over a model including only taxonomic levels was also strong (BF_{10} = 10.48), contrasting our findings in Experiment 3b, but matching those in Experiments 1 and 2.

Finally, in accuracy, model comparisons showed equivocal evidence for a model including gradedness-adjusted sensorimotor distance over a model containing taxonomic levels and average typicality (BF_{10} = 1.22). We note here that the data favoured a model including taxonomic levels and the interaction between subordinate category labels over gradedness adjusted sensorimotor distance (BF_{01} = 1339.43).

Summary. A combined analysis of the data in Experiments 2 and 3b showed that across two experiments, there is no evidence for an effect of basic-level typicality on categorisation RT or accuracy. Furthermore, while we found a weak effect of the interaction between typicality and subordinate-level labels on RT, model comparisons showed that contrary to our hypothesis there was no support for a subordinate-level advantage over the classic basic-level advantage. Finally, this analysis partially supported our hypotheses regarding the effects of linguistic and sensorimotor information, in that sensorimotor but not linguistic information did indeed inform categorisation performance. Participants made categorical decisions more quickly when sensorimotor experience of the category and member concept overlapped to a large degree. These decisions were also made more accurately when sensorimotor experience of a category and concept overlapped to a greater extent, however this effect was not stronger than that of the classic, basic-level advantage.

Figure 2.

Bayes factors for models of accuracy and RT, compared to null (random effects) for experiments 2, 3b, and 4b, showing evidence for models containing taxonomic level only, taxonomic level and object typicality, sensorimotor distance only and sensorimotor and linguistic distance.



Note: best-fitted models are marked with an asterisk. In the case of equivocal performance between two models both models are marked. Experiment 4b denotes the analysis of a combined dataset of experiments 2 and 3b, rather than data from a new experiment.

2.10. General discussion

In this paper, we examined the extent to which the basic-level advantage in picture categorisation is influenced by the overlap in sensorimotor experience and linguistic distributional knowledge between a given category and its member concepts. We found that, as expected, sensorimotor and to a lesser extent linguistic distributional information contributed to latency and accuracy in speeded picture categorisation. Furthermore, we found that adjusting sensorimotor and linguistic distance to accommodate graded structure enhanced their ability to predict categorisation performance. Critically, we found that sensorimotor distance outperformed taxonomic level as a predictor of response time. This finding lends support our hypothesis that overlap in sensorimotor and linguistic distributional information predicts categorisation performance, and contrasts the taxonomic prediction of a basic-level advantage.

Nevertheless, in contrast to our predictions, we found that taxonomic level consistently outperformed both sensorimotor and linguistic distributional distance in predicting the accuracy of the response. That is, we observed a classic, basic-level advantage whereby images were categorised more accurately at the basic compared to the superordinate and subordinate level. While sensorimotor and linguistic distributional information did affect accuracy in the expected direction, taxonomic level was the best predictor overall. This effect was stable across experiments, even when we adjusted our measures of sensorimotor and linguistic distance to reflect graded category structure, and suggests that sensorimotor and linguistic distributional information do not fully capture taxonomic structure. We note here that taxonomic level in and of itself is an empty notion, and the variance it accounts for may come from for example psycholinguistic characteristics. Indeed, the classic basic-level advantage partially disappeared with the inclusion of psycholinguistic variables. The present study was not designed to tackle the highly related effects of for example age of acquisition and word frequency, but future work might explore the effect taxonomic levels and psycholinguistic characteristics in studies with an orthogonal design.

Furthermore, contrasting previous findings (Jolicoeur et al., 1984; Murphy & Brownell, 1985) we found no effect of object typicality on either accuracy or response time. One possible explanation for this finding is that our stimuli were overall highly typical: no items received an average typicality rating lower than 3.42 on a scale from 1 to 5, with the exception of one item which was excluded as an outlier. The inclusion of a typicality predictor to this dataset was pre-registered as a secondary analysis, and as such our stimuli were not explicitly selected to range in object typicality. Related to this is that we found little evidence for a correlation between linguistic distance, which may have been caused by the low variance in typicality. This finding is even more surprising because adjusting sensorimotor and linguistic distance to reflect graded structure did improve their ability to predict categorisation performance. As such, these findings support the idea that graded categorical structure affects categorical judgments, but the lack of an interaction between typicality rating and subordinate-level labels suggest that typicality ratings may not be the best way to measure categorical gradedness. A future study with a specifically selected wider range of items, containing more atypical exemplars would be preferred.

Our data demonstrate that it is possible to express the relationship between category and member-concepts in terms of sensorimotor overlap. Across two experiments (1 and 3), categorical judgments were made faster and more accurately when a potential member was closer in in sensorimotor experience and knowledge to its category. Crucially the evidence for the effect of sensorimotor information on response times was stronger than or equal to the evidence for an effect of the taxonomic level of the category label, which also reliably predicted categorisation performance. Our findings thus suggest that in this particular task, sensorimotor overlap is able to capture more of the nuances in the time course of categorisation than a strict division into three taxonomic levels might.

As a consequence, the present findings support a sensorimotor-linguistic account of object categorisation and contrast traditional accounts of processing advantages in categorisation in a number of ways, which we outline below.

The present findings contrast with network entry-level accounts, which argue that the basic-level is the preferred entry-level into a semantic structure with discrete taxonomic levels (Collins & Loftus, 1975; Glass & Holyoak, 1974), which may shift depending on object typicality (Jolicoeur et al., 1984). With regards to processing advantages in categorisation, these accounts argue that moving from the basic entry-point to the abstract superordinate or specific subordinate level incurs a processing cost, which results in the basic-level advantage. If this were indeed the case, then the discrete division into taxonomic levels should have been the strongest predictor of categorisation performance across all our experiments. However, results from Experiments 1, 2, and 4b were inconsistent with this view. While our data show that taxonomic information indeed was a good predictor of categorisation performance, sensorimotor overlap proved a better predictor of response latencies than taxonomic level.

In our data, we also found no evidence for an ‘entry-level shift’. That is, some hierarchical accounts (Jolicoeur et al., 1984) predict that atypical category members are preferentially categorised at the subordinate rather than the basic level, which would result in an interaction between the subordinate taxonomic level and object typicality. We found no such interaction in Experiments 2, 3b or 4b. A potential explanation for this finding is the absence of strongly atypical category members in our stimulus set, meaning that our set lacked sufficient gradedness to affect categorisation performance. However, across two Experiments (2 and 3b), we found strong evidence that adjusting our measures of sensorimotor experience and linguistic distributional overlap to reflect categorical gradedness improved the former's power to predict categorisation speed, beyond unadjusted measures and models including the interaction between the subordinate taxonomic level and average typicality.

It is important to note that the gradedness-adjusted sensorimotor measure incorporated the idea that pictures of "good" category members are recognised as their basic category concepts (e.g., picture of *jeans* recognised as *trousers*) while pictures of unusual category members are recognised as the specific, subordinate member concept (e.g., picture of *sweatpants* recognised as *sweatpants*, and not *trousers*). As a result, "good" member concepts are judged more quickly and accurately when preceded by their basic-level label (e.g., *trousers* → [picture of *jeans*] processed as *trousers* → *trousers*), and all other judgements are slower and less accurate according to the sensorimotor distance between the category and member concepts (e.g., subordinate *jeans* → [picture of *jeans*] processed as *jeans* → *trousers*; superordinate *clothing* → [picture of *jeans*] processed as *clothing* → *trousers*). On the other hand, unusual member concepts are judged more quickly and accurately when preceded by their specific, subordinate label (e.g., *sweatpants* → [picture of *sweatpants*] processed as *sweatpants* → *sweatpants*), and all other judgements are slower and less accurate as per above (e.g., basic *trousers* → [picture of *sweatpants*] processed as *trousers* → *sweatpants*; superordinate *clothing* → [picture of *sweatpants*] processed as *clothing* → *sweatpants*). As a measure, the average typicality rating for an item of its basic-level category proved ineffective at confirming the predictions made by hierarchical accounts (e.g., Jolicoeur et al., 1984). However, given the increased performance of an adjusted measure of sensorimotor distance, it may be the case that the notion that unusual items are implicitly named at the subordinate rather than the basic level is in fact valid.

The present findings also contrast with other feature-based accounts, which hold that taxonomic information is implicit in feature relations (e.g., differentiation account Collins & Loftus, 1975; Markman & Wisniewski, 1997; Murphy & Brownell, 1985; Murphy & Lassaline, 1997). These accounts predict a performance advantage for the basic level, because it provides the maximally distinctive and informative match to an object's features. As a consequence, a differentiation account would also predict taxonomic level to be the best predictor of categorisation performance overall. However, sensorimotor overlap best predicted RT, whereas taxonomic level best predicted accuracy. The fact that both taxonomic information and sensorimotor information are strong predictors does not constitute evidence against a sensorimotor-linguistic account of categorisation performance. Feature-based differentiation accounts of categorisation suggest that taxonomic structure arises from the distribution of features, and not vice versa. A potential explanation for the fact that taxonomic and sensorimotor information both explain aspects of categorisation performance is therefore that taxonomic information emerges from sensorimotor experience and/or linguistic distributional knowledge. Indeed, recent research suggests that hierarchical information may emerge spontaneously from the latent structure in multidimensional sensorimotor experience, without the need for discrete features (Connell et al., 2020). That is, the relative extent to which different perceptual modalities and action effectors underlie sensorimotor experience of concepts is sufficient to extract hierarchical information. That is, within the latent structure in multidimensional sensorimotor experience, *sparrows* and *crows* are closer to one another than *sparrows* and *hammers*, but both are also closer to *bird* than they are to *animal*. This may indicate that while concepts are represented in semantic memory through sensorimotor experience and linguistic distributional knowledge, hierarchical information emerges implicitly through similarity relationships between concepts, depending on what concepts are compared. As such, semantic memory may not be organised in fixed hierarchies, but the latent structure of multidimensional sensorimotor experience may give the impression of taxonomic organisation, depending on what and how concepts are compared.

Our findings are not straightforward, and did not confirm our hypotheses completely. Contrary to our hypothesis, linguistic distributional information did not reliably predict categorisation performance above and beyond sensorimotor distance. That is, it predicted accuracy better than sensorimotor distance in Experiment 1, but not in Experiment 3a, However, the fact that the effect of sensorimotor information was more pronounced in our data may have been a result of the nature of the task. In our task, a classic label-picture categorisation task, participants verified whether the object described by the label was visible in the picture. It has been argued by linguistic-simulation theories (Barsalou et al., 2008; Connell, 2018; Connell & Lynott, 2014; Louwerse, 2011) that the systems governing linguistic and simulated information are interactive and as such, labels may activate sensorimotor representations, which facilitate picture verification (Boutonnet & Lupyan, 2015; Lupyan & Thompson-Schill, 2012). As such, we suggest the following process model: when participants see a label (e.g., *dog*), this activates a sensorimotor representation. When the participant sees the image, they verify whether the image matches this activated representation. The greater the overlap between the perceived image and the pre-activated perceptual simulation, the less additional activation is required and the faster and more accurate the response. Participants may have relied more strongly on perceptual simulations than linguistic information in the particular task we used here, as a label was always followed by an image. It may have been the case that access to in particular visual experience may have weighed heavier than linguistic distributional knowledge. To test this, future work may explore whether linguistic distributional information is a stronger predictor of performance in categorisation with limited access to visual information.

In summary, our results show that a sensorimotor-linguistic approach to concepts and categories may explain variations in categorisation performance, and thus provide a viable alternative to classic explanations of the basic-level advantage. Firstly, our measures of sensorimotor and/or linguistic distributional similarity successfully captured aspects of the relationship between categories and their members: we have demonstrated they not only predict categorisation performance, but also reflect the graded structure of categories. This shows that categories may be represented directly through sensorimotor and/or linguistic representations, and provides an alternative explanation to hierarchical accounts of the basic level advantage (e.g., Jolicoeur et al., 1984).

Overall, this study found support for the view that the relationship between category- and member-concepts may be expressed in terms of sensorimotor and/or linguistic distributional overlap, and therefore that categorisation may be the result of clustering concepts in terms of sensorimotor and/or linguistic experience. The limited effect of linguistic distributional information in this task is surprising, but is still in line with sensorimotor-linguistic accounts of conceptual processing, which argue a conceptual system that relies on both types of information is dynamic rather than static, and may rely more on one information type compared to the other depending on task constraints (Connell & Lynott, 2014). Our findings are thus largely in line with current sensorimotor-linguistic views on conceptual processing.

2.11 References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*(3), 263–308. https://doi.org/10.1016/0010-0277(83)90012-4

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning: From Acquiring Specific Items to Forming General Rules. *Current Directions in Psychological Science*, *21*(3), 170–176. https://doi.org/10.1177/0963721412436806

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

Banks, B., & Connell, L. (2021). Category production norms for 117 concrete and abstract categories in British English. *Manuscript in Preparation*.

Banks, B., Wingfield, C., & Connell, L. (2021). Linguistic Distributional Knowledge and Sensorimotor Grounding both Contribute to Semantic Category Production. *Cognitive Science*, *45:* e13055. https://doi.org/10.1111/cogs.13055.

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

Barsalou, L. W., Santos, A., Simmons, K. W., & Wilson, C. D. (2008). Language and Simulation in Conceptual Processing. In M. de Vega, A. M. Glenberg, & A. Graesser (Eds.), *Symbols, Embodiment and meaning* (pp. 245–283). Oxford University Press.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380. https://doi.org/10.3758/BF03196494

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical-semantic competition hypothesis. *Cognition*, *96*(2), B45–B54. https://doi.org/10.1016/j.cognition.2004.11.006

Boutonnet, B., & Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience*, *35*(25), 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, *13*(7–8), 992–1011. https://doi.org/10.1080/13506280544000165

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526. https://doi.org/10.3758/BF03193020

Carota, F., Moseley, R., & Pulvermüller, F. (2012). Body-part-specific Representations of Semantic Noun Categories. *Journal of Cognitive Neuroscience*, *24*(6), 1492–1509. https://doi.org/10.1162/jocn_a_00219

Catling, J. C., Dent, K., & Williamson, S. (2008). Age of acquisition, not word frequency affects object recognition: Evidence from the effects of visual degradation. *Acta Psychologica*, *129*(1), 130–137. https://doi.org/10.1016/j.actpsy.2008.05.005

Catling, Jon. C., & Johnston, R. A. (2006). The effects of age of acquisition on an object classification task. *Visual Cognition*, *13*(7–8), 968–980. https://doi.org/10.1080/13506280544000138

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

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. (2010). Look but don't touch: Tactile disadvantage in processing modality-specific words. *Cognition*, *115*(1), 1–9. https://doi.org/10.1016/j.cognition.2009.10.005

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & 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., & Ramscar, M. (2001). Using Distributional Measures to Model Typicality in Categorization. In *Proceedings of the Twenty-Third Annual Meeting of the Cognitive Science Society* (pp. 226–231). Lawrence Erlbaum Associates.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*(2), 163–201. https://doi.org/10.1037/0096-3445.132.2.163

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Dent, K., Catling, J. C., & Johnston, R. A. (2007). Age of acquisition affects object recognition: Evidence from visual duration thresholds. *Acta Psychologica*, *125*(3), 301–318. https://doi.org/10.1016/j.actpsy.2006.08.004

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

Glass, A., & Holyoak, K. (1974). Alternative conceptions of semantic theory. *Cognition*, *3*(4), 313–339. https://doi.org/10.1016/0010-0277(74)90002-X

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

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

Hampton, J. (1993). Prototype models of concept representation. In I. van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67–95). Academic Press.

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

Hilbig, B. E. (2016). Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*(3), 248–277. https://doi.org/10.1037/0096-3445.126.3.248

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243–275. https://doi.org/10.1016/0010-0285(84)90009-4

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

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What categories reveal about the mind*. The University of Chicago Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

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

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 250–269. https://doi.org/10.1016/S0022-5371(84)90170-1

Mandler, J. M., & McDonough, L. (2000). Advancing Downward to the Basic Level. *Journal of Cognition and Development*, *1*(4), 379–403. https://doi.org/10.1207/S15327647JCD0104_02

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 54–70. https://doi.org/10.1037/0278-7393.23.1.54

Mervis, C. B., & Crisafi, M. A. (1982). Order of Acquisition of Subordinate-, Basic-, and Superordinate-Level Categories. *Child Development*, *53*(1), 258–266. https://doi.org/10.2307/1129660

Mervis, C. B., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, *32*(1), 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513

Morrison, C. M., Hirsh, K. W., & Duggan, G. B. (2003). Age of Acquisition, Ageing, and Verb Production: Normative and Experimental Data. *The Quarterly Journal of Experimental Psychology Section A*, *56*(4), 1–26. https://doi.org/10.1080/02724980244000594

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 70–84. https://doi.org/10.1037/0278-7393.11.1.70

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical Structure in Concepts and the Basic Level of Categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts and Categories* (pp. 93–131). Psychology Press - Taylor & Francis.

Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 1–20. https://doi.org/10.1016/S0022-5371(82)90412-1

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 572–586. https://doi.org/10.1037/0278-7393.15.4.572

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363. https://doi.org/10.1037/h0025953

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

Rips, L. J., Shoben, E., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20. https://doi.org/10.1016/S0022-5371(73)80056-8

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition & Categorization*. Lawrence Erlbaum Associates, Publishers.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502. https://doi.org/10.1037/0096-1523.2.4.491

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. (2019). 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

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*(4), 1241–1260. https://doi.org/10.3758/s13428-016-0783-4

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Harvard University Press.

Smith, E. E., Shoben, E., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214–241. https://doi.org/10.1037/h0036351

Snodgrass, J., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of*

Experimental Psychology: Human Learning and Memory, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598–605. https://doi.org/10.3758/BF03193891

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual Structure and the Structure of Concepts: A Distributed Account of Category-Specific Deficits. *Brain and Language*, *75*(2), 195–231. https://doi.org/10.1006/brln.2000.2353

Urooj, U., Cornelissen, P. L., Simpson, M. I. G., Wheat, K. L., Woods, W., Barca, L., & Ellis, A. W. (2014). Interactions between visual and semantic processing during object recognition revealed by modulatory effects of age of acquisition. *NeuroImage*, *87*, 252–264. https://doi.org/10.1016/j.neuroimage.2013.10.058

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

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

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

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Chapter 3: How Sensorimotor and Linguistic Distributional Information Affect Performance in the Categorisation of Concept Labels

3.1. Chapter introduction

This Chapter builds on the findings from the previous Chapter, and aims to distinguish between the effects of sensorimotor and linguistic distributional information and taxonomic level on categorisation. In speeded picture verification, as described in Chapter 2, the effects of taxonomic level and sensorimotor-linguistic information may be difficult to discern, as their predictions may overlap. Furthermore, the study in Chapter 2 used a classic label → picture verification task, which may have inadvertently caused participants to rely more on sensorimotor (visual) than linguistic distributional information.

The study presented in this chapter therefore uses a forced-choice paradigm with only labels, where participants decide whether a basic- or superordinate-level label fits a subordinate-level cue best (e.g., *Labrador → dog* or *animal*). Vitally, sensorimotor and linguistic overlap is manipulated so that in half the trials, overlap is greatest between the exemplar cue and the basic-level label, and for the other half it is greatest between the exemplar clue and the superordinate-level label. As a consequence, for half the trials, sensorimotor-linguistic predictions match traditional taxonomic predictions (i.e., a basic-level advantage), whereas for the other half they pull in the opposite direction (i.e., a superordinate-level advantage).

How Sensorimotor and Linguistic Distributional Information Affect Performance in the Categorisation of Concept Labels

Rens van Hoef, Louise Connell, and Dermot Lynott

Department of Psychology, Lancaster University

Author Note

Rens van Hoef https://orcid.org/0000-0003-1355- 1541

Louise Connell https://orcid.org/0000-0002-5291-5267

Dermot Lynott https://orcid.org/0000-0001-7338-0567

This research has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 682848) to LC.

Correspondence concerning this article should be addressed to Rens van Hoef, Department of Psychology, Fylde College, Lancaster University, Bailrigg, Lancaster, LA1 4YF, UK. Email: r.vanhoef@lancaster.ac.uk

3.2. Abstract

Previous work has shown that a sensorimotor-linguistic account of categorisation, which is based on the overlap in sensorimotor and linguistic distributional information between category and member-concepts could predict the basic-level advantage in object picture categorization approximately as well as the traditional division into discrete taxonomic levels. In this study, we tested competing hypotheses of the sensorimotor-linguistic and taxonomic accounts via a forced-choice response task, in which participants decided between a basic- (e.g., *dog*) and superordinate-level label (e.g., *animal*) for a subordinate-level cue (e.g., *Labrador*). Crucially, we manipulated our stimuli so that for half the trials, sensorimotor and linguistic distributional information overlapped more strongly with the basic-level choice, whereas for the other half the overlapped more strongly with the superordinate-level choice. We hypothesised that basic-level labels would be chosen less often and responded to more slowly in pro-superordinate trials compared to pro-basic trials. While we found that choices were indeed less likely to be basic in pro-superordinate trials, we still found an overall basic-level advantage. Furthermore, in pro-superordinate trials, participants were slower to respond to both basic and superordinate-level labels.

3.3. Introduction

When seeing a robin in their garden, people tend to class it as a *bird,* rather than as a *robin* or *animal*. This intermediate level of abstraction is generally referred to as the basic level (e.g., *bird,* Rosch, Mervis, et al., 1976). Evidence for the preference for categorising objects at the basic level comes from decades of behavioural studies, which show that among other things, people are faster and more accurate to identify natural and artificial objects at the basic level (K. E. Johnson & Mervis, 1997; Jolicoeur et al., 1984; Murphy & Brownell, 1985; Murphy & Lassaline, 1997; Murphy & Smith, 1982; Rogers & Patterson, 2007; Rosch, Mervis, et al., 1976; Tanaka & Taylor, 1991). Moreover, the most frequent, earliest-acquired category names tend to be basic (Mervis & Crisafi, 1982b; Rosch, Mervis, et al., 1976). Together, these findings illustrate what is known as the *basic-level advantage* in categorisation.

Traditional explanations of the basic-level advantage are firmly rooted in what we will commonly refer to as feature-based theories of categorisation. These theories, while varied in their accounts of categorical representation, generally assume categories to be labelled classes of objects or entities that share discrete perceptual and/ or functional features to a sufficient degree to be similar to members from the same class, yet different from non-members (Hampton, 1981; Rosch, 1973; Rosch, Simpson, et al., 1976; B. Tversky & Hemenway, 1984). In one influential view, categories are represented in semantic memory through an abstracted summary representation (prototype) of the distribution of features within a given category (Posner & Keele, 1968; Rosch, 1975; Rosch & Mervis, 1975). In an alternative view, concepts and their features are stored in hierarchical networks, mirroring the inclusive relationships between concepts in a taxonomy (Collins & Loftus, 1975; Jolicoeur et al., 1984). Crucially, prototype and network accounts provide alternative explanations for the processing advantages observed for category labels of intermediate specificity. Hierarchical network accounts suggest that the basic-level is the preferred entry-point into the hierarchical organisation of concepts in semantic memory. On this account, categorising an object at more specific or abstract levels incurs a perceptual or inferential processing cost (Glass & Holyoak, 1974; Jolicoeur et al., 1984), that may be mediated by object typicality (i.e., atypical objects may be preferentially categorised at the subordinate level). Prototype accounts meanwhile argue that the basic level has an advantage due to its feature characteristics, which serve to maximise its within-category similarity and between-category differentiation (Cree & McRae, 2003; Markman & Wisniewski, 1997; Murphy, 2016; Murphy & Brownell, 1985; Murphy & Lassaline, 1997; Tyler et al., 2000). Murphy and Smith (1982) argue that differences in categorisation performance arise because verification of superordinate-level labels requires the activation of additional perceptual representations, and verification of subordinate-level items requires a different set of discriminatory criteria.

Over the past few decades, views on conceptual processing have shifted in a way that brings into question some of the explanations of processing advantages in categorisation as put forward by feature-based accounts. These accounts, which we will refer to as linguistic-simulation accounts of conceptual structure, argue that concepts comprise both sensorimotor (i.e., perception-action experience of the world) and linguistic distributional information (i.e., our knowledge about the distribution of words in language; (Barsalou et al., 2008; Connell, 2018; Connell & Lynott, 2014; Louwerse, 2011). Sensorimotor simulation refers to the notion that the same systems that are involved in online processing of perceptual experience may be re-activated to represent (simulate) this experience upon retrieval (Barsalou, 1999). Simulated representations may consist of perceptual, motor, affective and other information in direct or indirect experience, and do not need to be the same every time (e.g., the concept *dog* may be represented by the colour and touch of its fur or the sound of its bark). Evidence for simulated representations comes from a range of neuroimaging (Aziz-Zadeh et al., 2006; Carota et al., 2012; Goldberg et al., 2006; Hauk et al., 2004) and behavioural (Connell & Lynott, 2010;

Dils & Boroditsky, 2010; Zwaan & Taylor, 2006) studies, which have found strong links between perceptual experience and conceptual representation. Linguistic distributional information captures our inherent sensitivity to the statistical properties of language (Aslin & Newport, 2012; Landauer & Dumais, 1997; Lund & Burgess, 1996), which allows us to retrieve information from the specific patterns in the distribution of words relative to one another. For example, the fact that certain words (e.g., *bird* and *creature*) occur frequently in similar contexts may indicate that they are more similar than others (e.g., *bird* and *spaghetti*). Linguistic distributional information has been shown to predict conceptual processing in a range of tasks (Connell & Lynott, 2013; Goodhew et al., 2014; Louwerse & Connell, 2011; Louwerse & Jeuniaux, 2010), including categorisation processes such as typicality rating (Connell & Ramscar, 2001) and category production (Banks et al., 2021). Various sensorimotor-linguistic accounts may weigh the importance of both information types in conceptual representation differently. Barsalou et al. (2008), emphasise the importance of sensorimotor simulation over linguistic distributional information, whereas Louwerse (2011) argues language encodes a sufficient amount of sensorimotor experience to be a viable alternative to simulation most of the time. Connell and Lynott (2014, see also Connell, 2019) meanwhile argue that neither sensorimotor nor linguistic distributional information is more central to conceptual processing as a whole, but each type may become more important than the other depending on task demands and other factors.

If it is indeed the case that concepts comprise both sensorimotor experience and linguistic distributional knowledge, it follows that sensorimotor and or linguistic distributional information may also capture categorical relations between concepts. Concepts may be grouped together according to their similarity in terms of sensorimotor experience and or the distribution of their label in language. For example, the sensorimotor profile of our experience with *poodles* may be very similar to our sensorimotor experience with *dogs*, but quite different to the sensorimotor experience of *animals* in general. In language, we tend to talk about *poodles* and *dogs* in similar contexts (e.g., leads, kennels, fur, collar, walks etc.) whereas animals do not share many of those contexts (e.g., we may talk about *animals* in the context of zoo, farm, wild or endangered). Indeed, in a previous study (van Hoef et al., 2021), we found that overlap in sensorimotor and linguistic distributional information between a category- (e.g., dog) and member-concept (e.g., *poodle)* predicts performance in a classic picture categorisation task.

While (to our knowledge) no explicit account of the role of sensorimotor and or linguistic distributional information in categorisation exists, we believe that there are a number of ways in which such an account would be different from hierarchical network and prototype/differentiation theories of categorisation. Firstly, sensorimotor-linguistic distributional account of categorisation does not assume categories to be labelled classes that are represented through abstract feature summaries, or by their exemplars. Instead, all natural concepts may be represented directly through sensorimotor experience and distributional knowledge. That is, the relationship between a member-concept (e.g., *dog*) and a category-concept (e.g., *animal*) may be expressed directly as the degree of overlap in sensorimotor experience and or linguistic distributional information between *dog* and *animal*, without the need for comparison to an abstract prototype or exemplar. Secondly, sensorimotor-linguistic views deviate from feature-based accounts of categorisation in that they do not draw a distinction between features and concepts. Instead, they are viewed as related but not secondary to other concepts (e.g., concepts *tail* and *barks* are related to *dog*). Finally, similar to prototype-differentiation accounts but different to hierarchical network accounts, a sensorimotor-linguistic account does not assume an *a priori* division of concepts into taxonomic levels. Instead, task goals and available resources may determine which concept is activated first (Connell & Lynott, 2014).

In previous work (van Hoef et al., 2022a), we suggested that predictions from a sensorimotor-linguistic account with regards to object categorisation might be summarised in a process model that bears some similarity to the preparation model, which is a precursor of the differentiation model (Murphy & Smith, 1982), in that category labels may activate a representation that is verified upon seeing a potential member, and that additional processing may be required if the representation activated by the label does not match the object. In the preparation model, this processing involves the extraction of additional perceptual features for categorisation at the superordinate level, and the activation of additional discriminatory criteria for categorisation at the subordinate level. A sensorimotor-linguistic variant of the preparation model might be similar, in that a category label (e.g., *dog*) activates sensorimotor and linguistic distributional information about a category-concept (e.g., our sensorimotor experience with *dogs*, as well as our knowledge about the words the label *dog* occurs with). However, where the preparation/ differentiation models describe categorisation as a feature-matching process, the sensorimotor-linguistic model predicts that categorisation performance is improved when the activated sensorimotor-linguistic representation overlaps with that activated by the image to a greater extent. Consequently, the sensorimotor-linguistic model does not make predictions about processing differences at various taxonomic levels, but rather predicts that any processing differences arise because representations activated by a given label (e.g., *dog*) might overlap more closely with those activated by the presented image (e.g., *Labrador*) than others (e.g., *animal*).

Hierarchical network accounts meanwhile suggest that the basic-level is the preferred entry-level into a taxonomically organised semantic memory, and that processing objects at a more abstract or concrete taxonomic level incurs an additional processing cost (Collins & Loftus, 1975; Glass & Holyoak, 1974; Jolicoeur et al., 1984). By contrast, the sensorimotor-linguistic model simply holds that different taxonomic labels activate different sensorimotor-linguistic representations which overlap with the representation activated by the image to varying extents. If the representations overlap, less additional processing is required than if the representations do not overlap.

In summary, the hierarchical network, prototype/differentiation and sensorimotor-linguistic accounts differ in their views on the mechanisms underpinning categorisation. Nevertheless, all three accounts may produce a basic-level advantage *effect*. It may be the case that basic-level labels (e.g., *dog*) generally overlap more in sensorimotor and linguistic distributional information with member concepts (e.g., *Labrador*) than superordinate-level labels (e.g., *animal*). As a result, observing a basic-level advantage brings us no closer to determining whether a sensorimotor-linguistic distributional account of categorisation is more plausible than other accounts, particularly when the best-fitting account varies according to which measure of the basic-level advantage is being analysed (van Hoef et al., 2021 – see Chapter 2). In that sense, it may be more informative to look at where predictions would diverge. In other words, where traditional accounts might predict a basic-level advantage, where sensorimotor-linguistic distributional accounts would not. That is what we investigate in this study.

The current study

In the present study, we report a pre-registered forced-choice categorisation experiment, involving only object and category *labels*, where participants saw a cue (e.g., subordinate Labrador) and then had to decide between two competing category labels at different taxonomic levels (e.g., superordinate *animal* and basic *dog*). Crucially, we manipulated our set of stimuli using measures of sensorimotor and linguistic distributional information (see *critical predictors* below), so that for half the trials the basic-level label was closer in sensorimotor-linguistic experience to the cue (e.g., sensorimotor and linguistic distance between cue *trout* and basic-level label *fish* were .02 and .35 respectively, compared to .20 and .45 between cue *trout and* superordinate-level label *creature*), whereas for the other half the superordinate-level label was closer (e.g., sensorimotor and linguistic distance between cue *barracuda* and basic-level label *fish* were .13 and .06 respectively, compared to .06 and .52 between cue *barracuda and* superordinate-level label *creature*). We aimed to investigate whether participants' choices and decision-making speed could be predicted by sensorimotor (i.e., perception-action experience of the world) and linguistic distributional information (i.e., statistical distribution of words in language), and whether these effects would mediate the effects of taxonomic level, specifically when sensorimotor-linguistic information contrasted with the traditional basic-level advantage.

We hypothesised that the ease and likelihood of categorisation of a given object (e.g., *Labrador*) at either the basic (e.g., *dog*) or the superordinate (e.g., *animal*) would depend on the sensorimotor-linguistic representational overlap between the member- and category concepts. That is, that category concepts with more representational overlap would be preferentially chosen over category concepts with less representational overlap. Specifically, we predicted that superordinate categories (e.g., *animal*) would be chosen faster and more frequently than their basic-level counterpart (e.g., *dog*) if they were closer in sensorimotor and linguistic distributional experience to the object label, regardless of their taxonomic level. These predictions contrast what traditional accounts would predict, but follow predictions from sensorimotor-linguistic accounts. That is, traditional accounts would predict a basic-level advantage regardless of overlap in sensorimotor and/ or linguistic distributional information, whereas sensorimotor-linguistic accounts would predict no basic-level advantage for items where sensorimotor-linguistic experience overlaps more with the superordinate concept.

Of particular note, the present study used labels rather than images to represent category members. There are two reasons for this decision. Firstly, our test of sensorimotor-linguistic explanations of categorisation methodologically thus far relied on comparisons between the information activated by a category label, and the information activated by the implicit name of the pictured object (van Hoef et al., 2021 – see Chapter 2). This required us to make an assumption about what that implicit name would be. In the present study, the use of labels rather than images helped us to avoid this assumption altogether. Secondly, in a previous study, we observed that sensorimotor but not linguistic distributional information affected categorisation response times better than taxonomic information. Louwerse and Jeuniaux (2010) suggest that simulated and linguistic information may inform cognitive processing differently for pictorial and linguistic stimuli. As such, the use of pictorial stimuli may disproportionally highlight the effect of sensorimotor information over linguistic distributional information. That is, the lack of an effect of linguistic distributional information may be an artefact of using picture stimuli, and does not necessarily mean that linguistic distance does not affect categorisation RT. Therefore, we expect that using word stimuli may offer a greater opportunity to detect linguistic distance effects. Nevertheless, we expect that even when images are absent, an effect of sensorimotor and linguistic distributional information on categorisation may be observed.

3.4. Method

In this study (https://osf.io/vdka2/?view_only=ac6c78793a354a1f95da6c36b5ac6163) we examined the different categorisation performance advantages predicted by traditional accounts of categorisation (e.g., basic-level advantage) and sensorimotor-linguistic accounts (e.g., representational-overlap advantage) in a forced-choice label categorisation task. Participants first saw an object name at the subordinate level (e.g., *Labrador*). Then, they saw two labels: one at the basic level (e.g., *dog*) and one at the superordinate level (e.g., *animal*). Their task was to decide which of these labels best fit the object that preceded it. Crucially, half the trials were pro-basic (i.e., sensorimotor experience and linguistic distributional information of the cue was closer to that of the basic-level option), whereas half the trials were pro-superordinate (i.e., sensorimotor experience and linguistic distributional information of the cue was closer to that of the superordinate-level option). We expected that the basic-level label would be chosen faster and more frequently when it overlapped with the cue in sensorimotor and linguistic distributional information to a greater extent than the superordinate-level label, and vice versa.

Participants. We recruited thirty native speakers of English (24 female, M_{age}= 37 years old, SD_{age} = 14.78 years) through online platform Prolific.co, and received £1.75 in compensation. Participants were required through Prolific's screening criteria to have no reading impairments (e.g., dyslexia), and have a prolific approval rate of 95% based on their previous submissions. As pre-registered, participants were asked to respond to 5 attention checks at random intervals throughout testing, in which they completed simple arithmetic questions (e.g., 19-9 =?). We replaced 1 participant for answering incorrectly to more than 2 of these questions (3 incorrect answers). In addition to this, one participant was automatically replaced for exceeding the maximum duration of 44 minutes, as automatically calculated by Prolific.

We determined sample size via sequential hypothesis testing (Schönbrodt et al., 2017), which allows evidence for/against the hypothesis to accumulate until a pre-specified grade of evidence is reached. We stopped sampling at the minimum bound of Nmin = 30, when analysis 2, stage 2, step 2 (see Design and Analysis) of RT cleared the specified threshold of $BF_{10} \geq 3$ for both pro-basic ($BF_{10} = 1.62x10^9$) and pro-superordinate conditions ($BF_{10} = 328.17$). This showed that a categorical effect was evident for both pro-basic and pro-superordinate items.

Materials. We selected 80 object-category triads, consisting of a cue at the subordinate level (e.g., *Labrador*) and two category labels at the basic (e.g., *dog*) and superordinate (e.g., *animal*) level respectively. Cues corresponded to a total of 10 unique basic and 7 unique superordinate-level categories (see appendix for all test and filler triads). Crucially, during stimulus selection, we calculated sensorimotor and linguistic distances between potential cues and their basic and superordinate-level categories. To calculate linguistic distance. We used a subtitle corpus of 200 million words in British English (e.g., van Heuven et al., 2014) to calculate log co-occurrence frequencies around each word with a context radius of five. Each word in the corpus was represented as a vector of log-cooccurrence frequencies, allowing for the comparison of two words by calculating the cosine distance between their vectors. For example, our cue *Labrador* and category *dog* might occur in relatively similar contexts across language, whereas *Labrador* and the choice *animal* might occur in less similar contexts. As a result, the distance between the vectors for *Labrador* and *dog* is smaller than that between vectors for *Labrador* and *animal*. The resulting linguistic distances for each cue → category pair ranged from .25 to .91 (M = .50, SD =.13), with higher values denoting a greater distance to the cue (i.e., less overlap) and vice versa. To compare the extent to which a cue overlapped with a category in terms of sensorimotor experience, we followed the same approach as we did for a prior experiment (van Hoef et al., 2021 – see chapter 2), in calculating sensorimotor distance based on multidimensional ratings of sensorimotor strength. We used Lynott et al.'s (2020) sensorimotor norms for 40,000 concepts, which had people rate the extent to which they experienced a given concept through one of the six perceptual modalities (auditory, gustatory, haptic, interoceptive, olfactory and visual) and by performing an action with five action effectors (foot, hand, head, mouth and torso). Each dimension was rated separately on a scale from 0 to 5, allowing for the construction of an 11-dimensional vector of grounded sensorimotor experience for each concept. We retrieved sensorimotor vectors for all our cues and their associated choices, and calculated a cosine distance between them (i.e., 1 – cos (θ

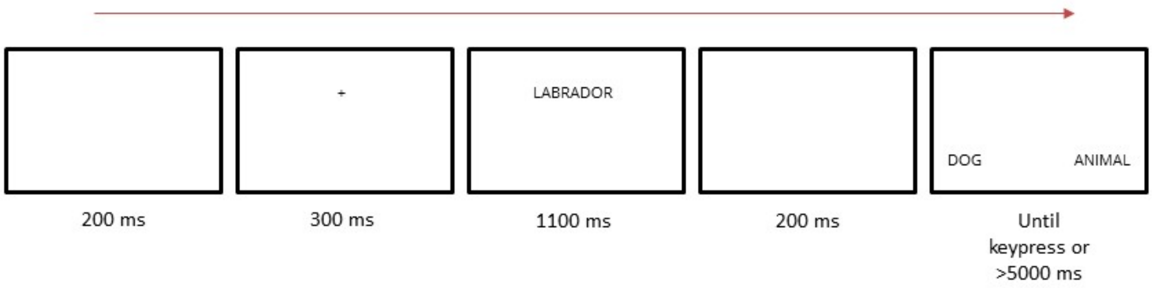
(u,v)). The resulting measure ranged from .01 to .56 (M = .1, SD = .09) with higher values reflecting greater distance to the cue (i.e., less overlap) and vice versa. The sensorimotor and linguistic distance measures formed the basis of our critical manipulation: we balanced the final set of 80 triads so that for half the triads sensorimotor and linguistic distance was biased towards the basic-level category (i.e., cue-basic distance was shorter than cue-superordinate distance: the pro-basic condition), and for the other half of the triads sensorimotor and linguistic distance was biased towards the superordinate-level category (i.e., cue-superordinate distance was shorter than cue-basic distance: the pro-superordinate condition. Finally, we used 36 fillers to balance the number of trials that represented a particular basic or superordinate category. That is, due to the limitations on our item selection (i.e., labels had to be present in the Lynott et al. sensorimotor norms, had to be have either a pro-superordinate or pro-basic sensorimotor-linguistic bias), some superordinate-labels only occurred in concurrence with one basic-level label (e.g., in test trials, the superordinate *animal* was only paired with *dog*), and some basic-level labels occurred only a limited number of times (e.g., only three *bottle-container* pairs fit our item requirements). The inclusion of fillers balanced this so that specific categories would not be subject to different practice effects in two ways. Firstly, we ensured that every superordinate-level label would occur with at least two different basic-level labels across the experiment (e.g., adding fillers containing the superordinate *animal* and the basic *insect*). Secondly, we added fillers to ensure that each basic-level label would occur in at least five trials across the experiment (e.g., by adding two *bottle-container* fillers). We ensured that for half of these filler triads sensorimotor but not linguistic distances were shortest between the cue and the basic-level, whereas for the other half this was reversed.

In addition to our measures of sensorimotor and linguistic distance, we retrieved Zipf log word frequencies for the basic and superordinate-level choice, as well as the presented cue from a large subtitle corpus (e.g., van Heuven et al., 2014). On average, basic-level labels were more frequent (*M*= 4.86, *SD*= .48) than superordinate-level labels (*M*=4.46, *SD*=.33), ($U = 5277.00$, $p < .001$). However, critically, log frequency averaged over both basic- and superordinate-level labels, was similar between the pro-basic (*M*=4.67, *SD*=.44), and pro-superordinate biased trials (*M*=4.65, *SD*=.48), ($U = 3186.00$, $p = .963$). Furthermore, average log word frequency for basic-level labels in pro-basic biased trials (*M*=4.81, *SD* =.52) was not significantly dissimilar to that of basic-level labels in pro-superordinate trials (*M*=4.52, *SD* = .29) ($U = 604.50$, $p = .06$), and average log word frequency for superordinate-level labels in pro-basic biased trials (*M*=4.90, *SD* =.45) was not significantly dissimilar that of superordinate-level labels in pro-superordinate trials (*M*=4.40, *SD* = .37) ($W = 958.50$, $p = .12$).

Procedure. We collected both informed consent and experimental data through web-based platform Gorilla.sc (Anwyl-Irvine et al., 2020). Trials were presented on a white background. Each trial began with a blank screen for 200 ms and was followed by a fixation cross for 300ms, the subordinate cue for 1100 ms presented centred in the top half of the screen (uppercase Open Sans), another blank screen for 200 ms, and finally a screen depicting both potential choices at the at the left and right bottom of the screen (uppercase, Open Sans), which remained on screen until a participant made a choice or a time-out threshold of 5000 ms was reached (see trial diagram, figure 1.).

Figure 1.

Trial design



The presentation order of superordinate and basic choices was counterbalanced across participants and trials, half the participants saw the basic-level choice on the left for half the trials, while they saw the superordinate-level choice on the left for the other half. For the other half of the participants, this presentation order was reversed. Participants were instructed to choose the option they believed best fit the cue by pressing the ‘z’ key on their keyboard to choose the left option, and the ‘m’ key to choose the right option, as quickly and accurately as they could. Fillers and test triads were displayed in random order. Five arithmetic questions (e.g., 19-9 =?) were displayed as attention checks at fixed intervals throughout the experiment, and halfway through participants were given a self-paced break. Testing took approximately 10 minutes, including informed consent and briefing.

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. Participants were informed that payment for their participation depended on passing at least two of the attention checks, and not timing out on more than 20% of trials. Consent included agreement to share publicly all alphanumeric data in anonymised form.

Design and analysis.

For analyses of choice and RT, our critical variables included taxonomic level of the choice (basic or superordinate) as well as log-transformed word frequencies for the cue and both choices, retrieved from a large subtitle corpus (e.g., van Heuven et al., 2014). Note that in the analysis of choice, taxonomic level was the dependent variable (i.e., the choice participants made was either a basic- or superordinate-level label), whereas in the analysis of RT, taxonomic level was a predictor. In addition to this, we included a binary predictor denoting whether a cue overlapped more in sensorimotor *and* linguistic distributional experience with the basic- or the superordinate-level cue (see Materials). Throughout our analyses we will refer to this predictor as *sensorimotor-linguistic bias*. During material selection and preparation, we used specific measures (outlined below) to capture the overlap in sensorimotor experience and linguistic distributional knowledge between a potential cue (e.g., *Labrador*) and its associated choices (e.g., basic-level *dog* and superordinate-level *animal*). We balanced our dataset so that for half the cue-choice triads, sensorimotor-linguistic overlap was greater between the cue and the basic-level choice, whereas for the other half, overlap was greater between the cue and the superordinate-level choice. To test our hypotheses, we planned two sets of analyses. Prior to each analysis we determined the best-fitting random effects structure (i.e., which random effects to include and whether to include random slopes) using Bayesian model comparisons, calculating Bayes Factors via BIC (Wagenmakers, 2007).

Analysis 1 tested whether the choice (basic or superordinate) participants selected was affected by sensorimotor-linguistic bias (pro-basic or pro-superordinate triad). We ran a hierarchical mixed effects logistic regression (binomial, logit link) of choice (basic = 0, superordinate = 1). Step 1 entered crossed random effects of participant and cue, and the covariate fixed effects of log word frequency of the chosen category, and of the rejected category (i.e., log word frequency of the category that was not chosen). Next, step 2 entered fixed effects of sensorimotor-linguistic bias (0 = pro-basic, 1 = pro-superordinate). Bayesian model comparisons between step 1 and 2 tested whether the data favoured a model containing sensorimotor bias over a null model containing random effects and covariate fixed effects of log word-frequency of both the chosen and rejected label. In addition to the pre-registered analyses of choice, we added three exploratory steps (3, 4, and 5) to establish whether high-frequency cues might yield different choices compared to low frequency cues, and to establish whether the effect of sensorimotor-linguistic bias was more likely to emerge for high- compared to average or low-frequency category labels. Therefore, step 3 added the fixed effect of word frequency of the cue, step 4 added the interaction between sensorimotor-linguistic bias and log word frequency of the chosen label, step 5 added the interaction between sensorimotor-linguistic bias and log word frequency of the rejected label. We used Bayesian model comparisons to compare each step to the model that preceded it, carrying over only those fixed effects that improved model fit.

Analysis 2 tested whether variance in RT could be explained by the taxonomic level of the choice participants selected (basic vs. superordinate) and or the sensorimotor-linguistic bias (pro-basic or pro-superordinate triad) in two separate stages. The first stage involved a hierarchical regression, where step 1 entered random effects of participant and cue, and covariate fixed effects of log word frequency of both the chosen and rejected category label. Step 2 entered a fixed effect of response choice (0 = basic, 1 = superordinate). Step 3 entered a fixed effect of sensorimotor-linguistic bias (0 = pro-basic, 1 = pro-superordinate), and step 4 entered the interaction between response choice and sensorimotor-linguistic bias. Bayesian model comparisons between step 3 and 4 tested whether response times for basic and superordinate categories varied by sensorimotor-linguistic bias, while Bayesian model comparisons between 2 and 4 tested whether the data favour the interaction effect over taxonomic information only (i.e., the classic basic-level advantage). We added a further seven exploratory steps, to establish whether high-frequency cues might elicit faster responses, and whether the effects of choice and sensorimotor-linguistic bias were more likely to emerge for high- compared to average and low cue and label frequencies. We used Bayesian model comparisons between each step and the previous best-fitting model to determine whether it improved model fit, and retained only the fixed effects for steps that did. Step 5 added the fixed effect of centred log word frequency of the cue. Step 6 added the interaction between word frequency of the chosen label and sensorimotor-linguistic bias. Step 7 added the interaction between word frequency of the rejected label and sensorimotor-linguistic bias. Step 8 added the interaction between word frequency of the cue and sensorimotor-linguistic bias. Step 9 added the interaction between word frequency of the chosen label and choice. Step 10 added the interaction between word frequency of the rejected label and choice, and step 11 added the interaction between word frequency of the cue and choice.

In the second stage, to test the direction of the sensorimotor-linguistic distance effects more detail, we split the dataset into pro-basic and pro-superordinate items, and ran a linear mixed effects regression analysis on each subset of the data. Step 1 entered random effects of participant and cue, and the fixed effects of log word frequency of the chosen and rejected labels. Next, step 2 entered an additional fixed effect of response choice (0 = basic, 1 = superordinate). Bayesian model comparisons between step 1 and 2 tested whether there was a difference in speed for categorising at basic versus superordinate level, and the coefficient of response choice showed the direction of the effect in each of the pro-basic and pro-superordinate subsets.

3.5. Results

We removed 65 trials (2.71 % of 2397 responses) as outliers from the dataset used in analyses 1 and 2, for having RTs more than 2.5SD away from the participant’s mean.

Table 1.

Model comparisons for confirmatory and exploratory logistic mixed effects regressions of choice (Analysis 1), showing change in R^2 and Bayes Factors for all comparisons.

Model	Variable entered	Comparison	ΔR^2	BF^{10}
Confirmatory				
Step 1 (null)	Random effects, log word frequency of chosen category, log word frequency of rejected category	-	.495*	-
Step 2	Sensorimotor-linguistic bias	Vs. step 1	.043	12.81
Exploratory				
Step 3	Log word frequency of the cue	Vs. step 2	.002	0.02
Step 4	Interaction sensorimotor-linguistic bias and log word frequency of chosen category	Vs. step 2	.040	8.28×10^{14}
Step 5	Interaction sensorimotor-linguistic bias and log word frequency of rejected category	Vs. step 4	.047	8040490.00

*reflects marginal R2 for random effects and fixed effects of word frequency and word length

Analysis 1: choice. Table 1 shows all model comparisons for analysis 1 of the choice participants made (basic or superordinate). Out of all 2332 non-outlier responses, participants chose a basic-level choice 1856 times (79.59 %), and a superordinate-level choice 476 times (20.41%). Bayesian model comparisons showed that the data favoured the step 2 model containing sensorimotor-linguistic bias over the null model containing random effects of participant and item, and log word frequency of the chosen and rejected category (BF_{10} = 13.06). This model reiterated that participants were on average most likely to choose a basic-level label, but also showed that they were up to 4.85 times more likely to choose a superordinate-level label when sensorimotor-linguistic bias was pro-superordinate compared to pro-basic [β =1.58, 95% CI = ±.84, z = 3.69, p <.001]. Furthermore, participants were also up to 12.38 times[9] more likely to choose a basic-level label when it was of high compared to average frequency [β =-3.70, 95% CI = ±.68, z = -10.71, p <.001]. By contrast, participants were up to 103.30 times[10] more likely to choose a superordinate-level label when the

---

[9] Odds ratio calculated with the highest value for centred log word frequency of the chosen word (.68).
[10] Odds ratio calculated with the highest value for centred log word frequency of the rejected word (.88).

alternative label was of high compared to average frequency [β =5.27, 95% CI = ±.84, z = -12.36, p <.001].

We carried out a set of exploratory analyses to investigate whether the choice participants made varied with the frequency of the cue, and whether the effect of sensorimotor-linguistic bias might vary with word frequency of the chosen and rejected category labels. Bayesian model comparisons between step 2 and the exploratory (i.e., not pre-registered) step 3 revealed evidence against inclusion of log word frequency of the cue over the step 2 (BF_{10}= 0.02). As a consequence, cue frequency was not carried forward to subsequent models. Further model comparisons between steps 2 and 4, and 4 and 5 revealed strong evidence for the inclusion of both the interaction between sensorimotor-linguistic bias and log word frequency of the chosen category and the interaction between sensorimotor-linguistic bias and word frequency of the rejected label see Table 2 for step 5 model coefficients. This model suggests that sensorimotor-linguistic bias by itself had little effect, but rather interacted with both the word frequency of the chosen and the rejected category in predicting which choice participants made (see Figure 1).

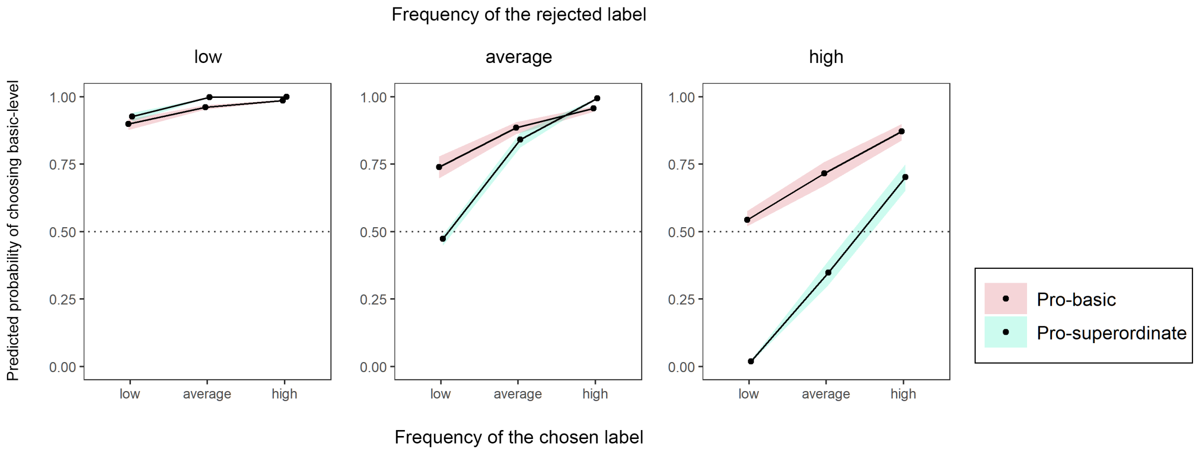
Table 2.

Estimates and standard errors of choice for sensorimotor-linguistic bias (0 = pro-basic), Log word frequency of the chosen and rejected label, and the interaction between sensorimotor-linguistic bias and word frequencies. Derived from the exploratory step 5 model.

	β	SE
(Intercept)	-3.10	.36
Sensorimotor-linguistic bias (0 = pro-basic)	.38	.54
Log word frequency of the chosen label	-2.09	.53
Log word frequency of the rejected label	2.36	.65
Sensorimotor-linguistic bias x Log word frequency of the chosen label	-4.80	.83
Sensorimotor-linguistic bias x Log word frequency of the rejected label	6.26	1.05

Figure 2.

Predicted probabilities of choosing basic-level categories by sensorimotor-linguistic bias for high (.5) average (0) and low (-.5) values of centred log word frequency of the chosen and rejected label (with 95% confidence intervals). Dotted line indicates chance threshold.



As shown in Figure 2, Panel 1, our exploratory analyses suggest that the basic-level advantage was stronger when the rejected superordinate-label was low-frequency, with little effect of the sensorimotor-linguistic bias or the frequency of the basic-level label itself (see panel 1). As shown in Figure 2, Panel 2, the basic-level advantage was present, but weaker when the rejected superordinate label was of average frequency with a relatively small effect of sensorimotor-linguistic bias. That is, in the pro-basic condition we observed a weaker basic-level advantage, when a low-frequency basic-level label was chosen over an average-frequency superordinate-level label. However, this advantage disappeared in the pro-superordinate condition, where sensorimotor-linguistic bias pushed participants to choose a low-frequency superordinate-level label over an average-frequency basic-label at least half the time (i.e., the predicted-probability of choosing basic of was .47). Finally, as shown in Figure 2, Panel 3, we observed strong effects of sensorimotor-linguistic bias when the rejected label was high-frequency. When sensorimotor-linguistic information was biased towards the basic-level label, the basic-level advantage appeared (albeit in weakened form) when a basic-label was of average to high frequency was chosen over a high-frequency superordinate alternative. The basic-level advantage also appeared in the pro-superordinate condition when both the basic and superordinate-level labels were high-frequency (i.e., predicted probability of choosing basic = .70). However, the basic-level advantage virtually disappeared in the pro-basic condition, where participants were reluctant to choose a low-frequency basic-level label over a high-frequency superordinate label (i.e., predicted probability of choosing basic = .54), and critically the basic-level advantage was *reversed* into a superordinate-level advantage in the pro-superordinate condition, where sensorimotor-linguistic bias pushed participants to choose a low- to average-frequency superordinate label over a high-frequency basic-label.

In summary, our confirmatory analysis confirmed our hypothesis that sensorimotor-linguistic overlap between the cue and the label affects categorisation performance. On average, participants were most likely to choose a basic-level label, showing a classic basic-level advantage. However, the likelihood that participants chose a basic-level label was lower in the pro-superordinate compared to the pro-basic condition. In addition to this, word frequency of both the chosen and rejected label affected participants' choices. Higher frequency of the chosen label resulted in more basic-level choices. This is perhaps unsurprising, as basic-level labels are generally of higher frequency. By contrast, when the rejected label was of higher frequency, the choice was more likely to be superordinate. This is the inverse of the effect we observed for frequency of the chosen label (i.e., because high-frequency words are more likely to be basic, a rejection of a high-frequency word is more likely to result in a superordinate choice). These results illustrate that sensorimotor and linguistic distributional information affect categorisation separately from taxonomic information.

Our additional analysis revealed an interesting pattern, where we noted that a pro-superordinate sensorimotor-linguistic bias was enough to make participants select a low- (or average-) frequency superordinate-level label, even when the basic-level option was high-frequency. However, given the exploratory nature of these analyses must be interpreted with care, and may require further research to determine their robustness.

Analysis 2: Response times (full dataset). Table 3 shows all model comparisons for stage one of analysis 2. On average, participants took 811.90 ms (*SD* = 423.37 ms) from the onset of the screen displaying both category labels to make a decision.

Table 3.

Model comparisons for confirmatory and exploratory linear mixed effects regressions of RT (full dataset), showing change in R^2 and Bayes Factors for all comparisons.

Model	Variable entered	Comparison	ΔR_2	BF_{10}
Confirmatory				
Step 1	Random effects, word frequency of the chosen label and word frequency of the rejected label	-	.356*	-
Step 2	Choice	Vs. step 1	.062	4.80x10^{14}
Step 3	Sensorimotor-linguistic bias	Vs. step 2	.023	0.52
Step 4	Interaction choice and sensorimotor-linguistic bias	Vs. step 2	.024	0.07
		Vs. step 3	.001	0.04
Exploratory				
Step 9	Interaction log word frequency of chosen category and choice	Vs. step 3	.009	7.55

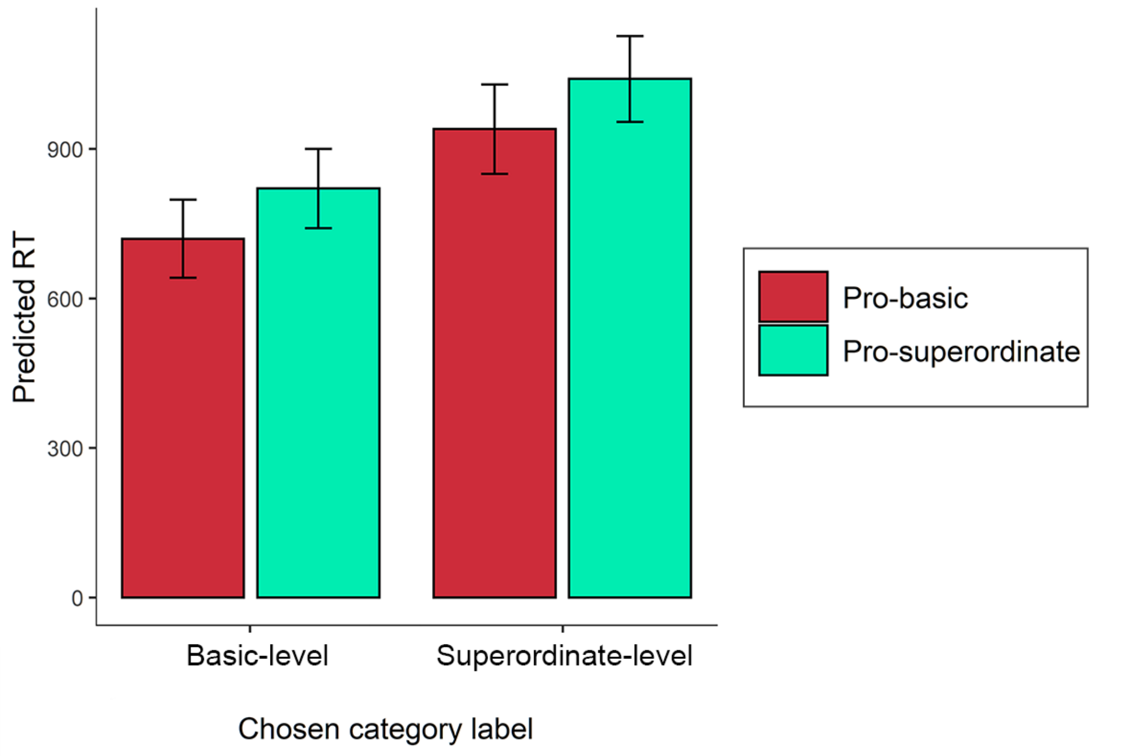
* Denotes conditional R^2, variance explained by both fixed and random effects.
Note. Only step 9 outperformed the best-performing confirmatory model (step 3), full model comparisons in supplementals on OSF.

Bayesian model comparisons between step 2 and 4 revealed that the data did not support the inclusion of the interaction between taxonomic level of the choice (basic or superordinate) and sensorimotor-linguistic bias (pro-basic, pro-superordinate) over a model including only log word frequency of the chosen and rejected label and taxonomic level of the choice (step 2). Evidence suggested that the data were BF_{01} = 13.66 times more likely under the step 2 model. Similarly, Bayesian model comparisons showed that the data did not support the step 4 interaction model over the step 3 model including all predictors from step 2 plus a fixed effect of sensorimotor-linguistic bias. Here, evidence suggested that the data were $BF_{10} = 26.03$ times more likely under the step 3 model. Taken together, these findings suggest that the effects of taxonomic level were robust across sensorimotor-linguistic bias conditions, thus not support our hypotheses regarding the effect of sensorimotor-linguistic bias.

An exploratory (not-preregistered) comparison between step 2 and 3 provided equivocal evidence for the inclusion of sensorimotor-linguistic bias over a model that included only word frequency and taxonomic level ($BF_{10} = 1.90$). Inspection of the coefficients for the step 3 model revealed that participants were faster in choosing high-frequency labels compared to labels of average and low frequency [β =-61.17, 95% CI = ±55.38, $t(136.69) = -2.16$, p =.032] as well as faster to make a response when the rejected label was high- compared to average and low-frequency labels [β =-116.94, 95% CI = ±64.45, $t(224.23) = -3.56$, p <.001]. Furthermore, participants were overall slower to choose superordinate-level labels [β =219.87, 95% CI = ±50.27, $t(2204.79) = 8.57$, p <.001], but also slower to choose both basic and superordinate-level labels in pro-superordinate biased trials compared to pro-basic trials [β =100.75, 95% CI = ±64.03, $t(78.19) = 3.08$, p =.002], see figure 3.

Figure 3.

Predicted response times (step 3 model) by taxonomic level and sensorimotor-linguistic bias with 95% confidence intervals.



As before in the analysis of choice, we carried out a set of exploratory analyses to investigate whether response times varied as a function of the frequency of the cue, as well as to see whether the effects of choice and sensorimotor-linguistic bias interacted with each of the various word frequencies. Bayesian model comparisons revealed that step 5 (adding word frequency of the cue) did not improved model fit over the confirmatory step 3 (BF_{10}= 0.02). As a consequence, cue frequency was not carried forward to subsequent steps. Only step 9 (adding the interaction between centred log word frequency of the chosen label and the taxonomic level of the choice) further improved model fit over step 3 (BF_{10}= 7.55), making it the best-fitting model overall. This model suggests that participants were faster to choose high-frequency over low- and average-frequency labels. However, the significant interaction between frequency and taxonomic level of the chosen label suggests that the speed-difference between basic and superordinate-level labels was higher for high-frequency compared to average and low frequency level labels (see Figure 4, and Table 4). That is, the basic-level advantage in RT was much larger when high-frequency labels were chosen compared to low-frequency labels. As before, this model suggests that participants were also faster in pro-basic compared to pro-superordinate biased trials, and when the alternative was a high-frequency compared to a low-frequency label.

Figure 4.

Predicted RT (step 9 model), for high (.5) average (0) and low (-.5) values (values chosen for illustrative purposes) of centred log word frequency by choice for pro-basic trials with a non-chosen label of average frequency, with 95% confidence intervals.

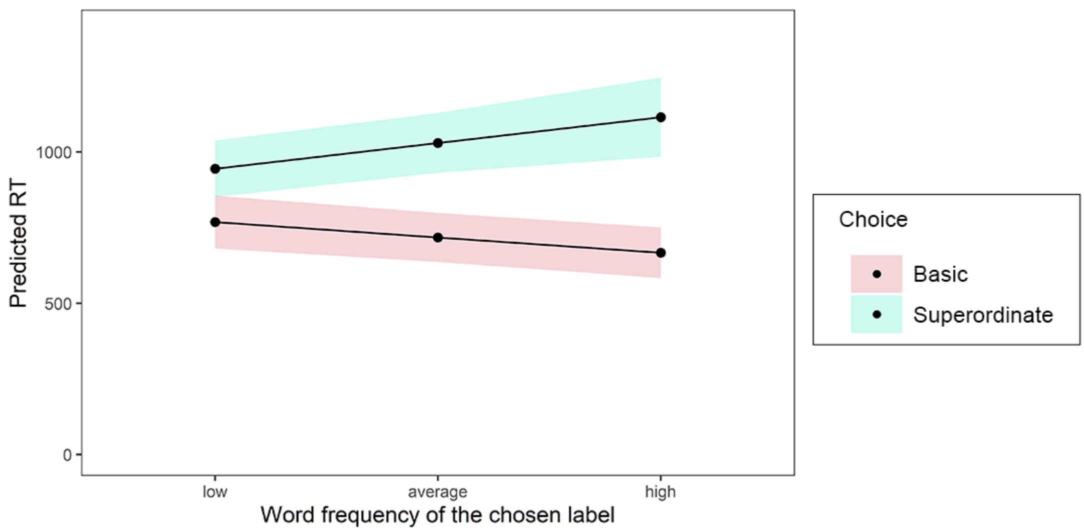


Table 4.

Estimates and standard errors of RT for choice (0 = basic, 1 = superordinate), sensorimotor-linguistic bias (0 = pro-basic), log word frequency of the chosen and rejected label, and the interaction between choice and word frequencies. Derived from the exploratory step 9 model.

	β	SE
(Intercept)	721.13	39.44
Choice (1 = superordinate)	295.56	30.58
Sensorimotor-linguistic bias (1 = pro-superordinate)	107.59	31.71
Log word frequency of the chosen label	-100.73	28.98
Log word frequency of the rejected label	-109.42	32.29
Choice x word frequency of the chosen label.	256.47	57.89

Table 5.

Model comparisons for confirmatory and exploratory linear mixed effects regressions of RT (dataset split by sensorimotor-linguistic bias condition), showing change in R^2 and Bayes Factors for all comparisons.

Pro-basic

Model	Variable entered	Comparison	ΔR^2	BF^{10}
Confirmatory				
Step 1 (null)	Random effects + log word frequency of chosen category + log word frequency of rejected category	-	.334*	-
Step 2	Choice (basic or superordinate)	Vs. step 1	.086	$1.62\text{x}10^9$

Pro-superordinate

Model	Variable entered	Comparison	ΔR^2	BF_{10}
Confirmatory				
Step 1 (null)	Random effects + log word frequency of chosen category + log word frequency of rejected category	-	.335*	-
Step 2	Choice (basic or superordinate)	Vs. step 1	.014	328.17
Exploratory				
Step 5	Interaction between choice and word frequency of the chosen category	Vs. step 3	.022	76.25
Step 6	Interaction between choice and word frequency of the rejected category	Vs. step 5	<.001	0.03

** Denotes conditional R^2, variance explained by both fixed and random effects.*

Note: Only exploratory models that improved model fit over the previous best-fitting model were included in this table. None of the exploratory models outperformed the step 2 model for a pro-basic dataset (full comparisons in supplementals on OSF).

Response times (split by sensorimotor-linguistic condition). In the second stage of analysis 2, we split the RT dataset by sensorimotor-linguistic bias (pro-basic and pro-superordinate). In the pro-basic dataset, Bayesian model comparisons (see Table 5) provided strong evidence for the step 2 model including taxonomic level of the chosen category over the step 1 null model including random effects and covariate fixed effects of log word frequency of the chosen and rejected label (BF_{10}= 1.62x10^9), where participants were faster to choose basic-level categories than they were to choose superordinate-level categories [β =240.09, 95% CI = ±65.21, t(1152.80) = 7.22, p <.001]. As for the full dataset, higher word frequency for the chosen [β =-98.47, 95% CI = ±70.59, t(56.84) = -2.73, p =.008] and rejected word [β =-153.63, 95% CI = ±85.78, t(108.25) = -3.51, p <.001] reduced response times.

In the pro-superordinate dataset, Bayesian model comparisons also provided strong evidence for the step 2 model including taxonomic level over the step 1 null model (BF_{10}= 328.65). Again, participants were faster to choose basic-level categories than they were to choose superordinate-level categories, although the difference was smaller [β =190.84, 95% CI = ±85.80, t(972.14) = 4.36, p <.001], but neither word frequency of the chosen and rejected label affected RT significantly.

We explored whether RT was affected by the word frequency of the cue, as well as the interaction between choice and centred log word frequency of the chosen label, rejected label and cue in 4 additional exploratory steps. In step 3, we added the fixed effect of centred log word frequency of the cue to the step 2 model. In step 4, we added a fixed effect of the interaction between choice and cue word frequency. In step 5, we added the fixed effect of the interaction between choice and chosen label word frequency, and in step 6 we added the interaction between choice and rejected label word frequency.

Bayesian model comparisons between successive steps 2, 3, 4 5 and 6 for a pro-basic dataset revealed that the original step 2 model, which included only centred log word frequency for the chosen and rejected label and taxonomic level of the choice fit the data best (see full details in supplementals), showing evidence against the inclusion of word-frequency of the cue (BF_{01}=34.06), and the interaction between choice and word frequency of the cue (BF_{01}=83.98), chosen label (BF_{01}=2620.73) or rejected label (BF_{01}=85146.03).

Bayesian model comparisons between steps 2, 3, 4 5 and 6 for a pro-superordinate dataset revealed evidence against step 3 (inclusion of word frequency of the cue) and step 4 (inclusion of the interaction between word frequency of the cue and choice), but for the inclusion of step 5 (including the interaction of word frequency of the chosen label with choice) over step 2. Step 6 did not improve model fit over step 5, providing evidence against the inclusion of the interaction between word frequency of the rejected label and choice.

Inspection of the step 5 model coefficients (see Table 6) showed that the taxonomic level of the chosen concept affected RT as previously observed, where superordinate-level choices were slower than basic-level choices. Increasing word frequency of the chosen and rejected labels resulted in faster RT. However, word frequency of the chosen concept differentially affected how quickly participants chose basic- versus superordinate category labels. That is, when people chose the basic-level label, higher word-frequency resulted in faster responses, but when people chose the superordinate-level label, higher word frequency resulted in *slower* responses (see Figure 5).

Figure 5.

Predicted response times for pro-superordinate and pro-basic datasets, by taxonomic level chosen, and non-chosen word frequency. With 95% confidence intervals.

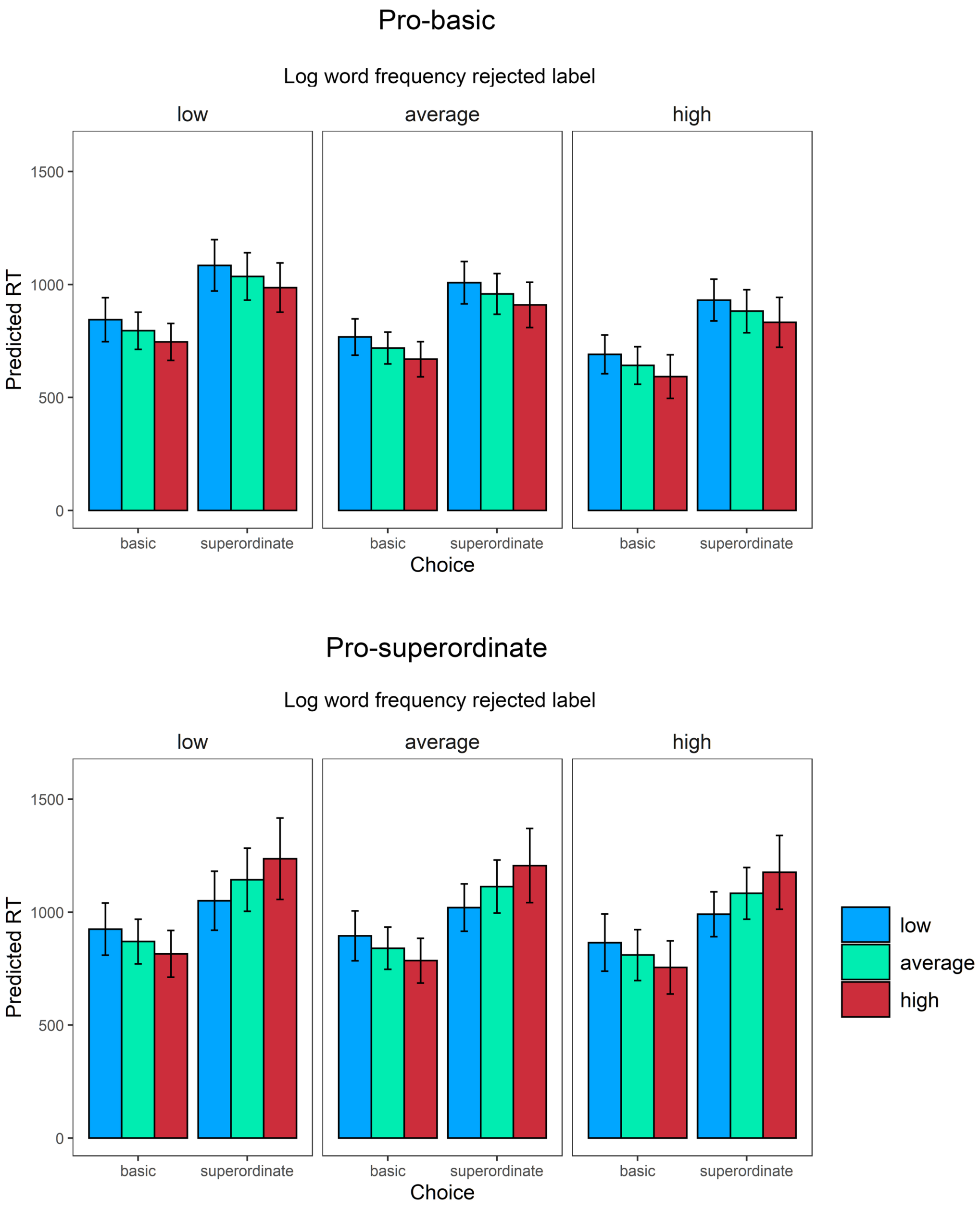


Table 6.

Model coefficients for the exploratory step 5 model of RT in a pro-superordinate dataset.

	B	SE
(Intercept)	839.87	45.97
Choice (1 = superordinate)	273.24	47.43
Centred log word frequency of the chosen label	-109.79	50.33
Centred log word frequency of the rejected label	-59.81	48.01
Centred word frequency of the chosen label x Choice	295.55	74.17

In summary, our confirmatory analyses suggested equivocal evidence for the effect of sensorimotor-linguistic bias, and no evidence of a variable effect of sensorimotor-linguistic bias with taxonomic level on RT. Our simple effects analysis of the data split by sensorimotor-linguistic bias condition revealed no additional information for pro-basic trials, but an exploratory analysis revealed that word frequency of the chosen and rejected label only affected RT in the pro-superordinate dataset for basic compared to superordinate items.

Additional exploratory analyses: psycholinguistic characteristics. In the confirmatory section of our study, we limited our investigation of the effects of psycholinguistic characteristics of our stimuli to word frequency. However, research has shown that other psycholinguistic variables may also affect performance in tasks using linguistic stimuli. For example, it has been shown that the age at which a word is typically learned (age of acquisition; AoA) affects how quickly participants recognise and classify words (). Similar, albeit less consistent, effects have been found for familiarity (i.e., how well participants know a given object) and word length. It is therefore possible that these variables explain variance in RT and choice in our study. In this exploratory analysis, we aimed to investigate whether AoA, familiarity ratings and word length of both the cue and choice labels indeed affected the choice participants made and the speed with which they did so.

In this exploratory analysis, we ran logistic mixed effects models of choice (0 = basic-level label, 1 = superordinate-level label) as per confirmatory analysis 1 and linear mixed effects models of RT as per confirmatory analysis 2. In the analysis of accuracy, we followed the procedure outlined in confirmatory analysis 1, but included additional psycholinguistic characteristics of age of acquisition, familiarity and word length of chosen and not chosen label to steps 1 and 2. As a consequence, in step 1, we entered all new psycholinguistic variables as well as word frequency of the chosen and not chosen label. In step 2, we entered sensorimotor-linguistic bias.

In the analysis of RT, we followed the procedure outlined for analysis 2, stage one. The only change we made to this procedure was the inclusion of fixed effects of age of acquisition, familiarity and word length of the chosen and not chosen category label to the null model at step 1, and to all subsequent models (steps 2,3 and 4, as outlined in confirmatory analysis 2). We used Bayesian model comparisons between steps 1 and 2 to test whether the inclusion of response choice (0 = basic, 1 = superordinate) improved model fit over a psycholinguistic-only model. Similarly, we used Bayesian model comparisons to test whether the inclusion of sensorimotor-linguistic bias in step 3 improved model fit over step 2, and whether the inclusion of an interaction between sensorimotor-linguistic bias and choice in step 4 improved model fit over step 3. Because of missing values in the familiarity dataset, we restricted our analyses to complete cases only, which resulted in a final dataset of 2167 trials (removing 7.07% from 2332 trials).

Choice: Non-convergence. Attempting to fit the step 1 model as described resulted in severe convergence issues. We first assured all variables were centred around their mean, and scaled them so that they all had a standard deviation of 1. Next, an inspection of the variance inflation factor revealed severe multicollinearity. As a consequence, we removed all fixed effects with a variance inflation factor (*VIF*) over 10: word length of the chosen label (*VIF* =

28.13), AoA of the label not chosen (*VIF* = 10.47) and word frequency of the chosen label (*VIF* = 13.56). While this reduced the collinearity issues, we still observed that our model fit was singular, with random effects estimates and standard errors dropping to near zero with the inclusion of psycholinguistic characteristics of the chosen and not chosen labels. One potential reason for this is that for each cue, only two potential values of age of acquisition, familiarity, word length and word frequency existed; one for each potential choice. For 13 of the 74 triads for which we had psycholinguistic data, participants always chose the basic-level option. As a result, for all these triads, there was no variance in any of the psycholinguistic characteristics of the label that was chosen and the label that was not chosen. Limiting the data to include only triads for which more than one choice was made (n = 1781 observations) did not resolve the issue. A further exploration of the remaining data revealed that the median number of times a superordinate-level label was 5, compared to a median of 25 times for basic-level labels. As such, it is possible that there was simply not enough data available to include psycholinguistic characteristics of the label that was not chosen. Therefore, we ran the model again, this time retaining only the two remaining psycholinguistic characteristics of the chosen label that were left (age of acquisition and familiarity), which improved matters, but nonetheless did not result in a converging model. Of particular interest was that the variance explained by the random effect of participant was near-zero. However, on theoretical grounds, we decided not to attempt dropping any random effects, and concluded that in its current form, the data is too sparse to support a complex psycholinguistic model. We note here that our study was designed to balance items based on their sensorimotor-linguistic bias, and may as such not have provided the optimal grounds for testing the effects of highly related variables. A future study might be designed to disentangle the interrelated effects of psycholinguistic characteristics such as frequency and age of acquisition.

RT. In step 1, our initial psycholinguistic model included all psycholinguistic characteristics of chosen and not chosen labels. However, inspection of the variance inflation factors for this model again revealed severe multicollinearity. As a consequence, we opted to remove the fixed effect with the highest variance inflation factor (word length of the chosen label, *VIF* = 23.13). An inspection of the updated model still showed evidence of multicollinearity, leading us to drop a second fixed effect (AoA of the label not chosen, *VIF* = 10.91). The final psycholinguistic model thus contained random effects of cue and participant, as well as fixed effects of word frequency and average familiarity ratings of the chosen and not chosen label, AoA of the chosen label and word length of the not chosen label, we carried these fixed and random effects forwards to steps 2 through 4.

Bayesian model comparisons showed evidence that the data favoured the step 2 model including the taxonomic level of the choice over the final step 1 model (BF_{10} = 16.82). Moreover, model comparisons showed strong evidence that the data favoured a step 3 model including sensorimotor-linguistic bias over the step 2 model (BF_{10} = 41.90). However, as we observed in our confirmatory analysis 2, we found strong evidence against the interaction between choice and sensorimotor-linguistic bias over a non-interaction step 3 model (BF_{01} = 33.17).

The coefficients of the best-fitting model (step 3) suggest that participants were slower to choose a label with higher-than-average age of acquisition [β =56.06, 95% CI = ±39.55, $t(395.90) = -2.78$, p =.006]. Furthermore, the degree of familiarity of either the chosen [β =-16.20, 95% CI = ±125.04, $t(196.33) = 0.25$, p =.254] or not chosen label had little effect [β =-23.73, 95% CI = ±97.13 $t(122.91) = -0.48$, p =.632]. Interestingly, participants were faster to make a decision when the word length of the label they did not choose was longer than average [β =-55.17, 95% CI = ±37.43, $t(84.53) = -2.89$, p =.005]. As in our confirmatory analyses, participants were faster to make a decision when the chosen label was of high compared to average frequency [β =-80.10, 95% CI = ±65.38, $t(76.83) = -2.40$, p =.019], as well as when the not chosen label was of high compared to average frequency [β =-176.28, 95% CI = ±87.62, $t(226.77) = -3.94$, $p < .001$]. Finally, as before, participants were overall slower to choose superordinate compared to basic level labels [β =161.68, 95% CI = ±90.75, $t(1349.38) = -3.49$, $p <.001$], and were slower to choose either basic or superordinate in a pro-superordinate compared to a pro-basic condition [β = 130.68, 95% CI = ±62.04, $t(68.50) = -4.13$, $p <.001$]. We note here that a model of the truncated data that did not include any psycholinguistic fixed effects fit the data better than the step 3 psycholinguistic model (BF_{10} = 150196.00).

In summary, our efforts to test the effects of various psycholinguistic variables on choice and RT in a forced-choice categorisation task proved only partially successful. In RT, results deepened our understanding, by showing that in particular age of acquisition and word length affected the time participants took to make a decision. Nevertheless, we still observed strong effects of taxonomic level (i.e., the basic-level advantage) as well as of our critical manipulation (sensorimotor-linguistic bias), suggesting that psycholinguistic characteristic and sensorimotor-linguistic bias explain distinct variance in our data.

The convergence issues in our analysis of choice do not allow us to draw any meaningful conclusions about the effects of age of acquisition, familiarity and word length on the decision participants made. One thing that may be taken away from this analysis is that future studies investigating psycholinguistic effects in this particular paradigm will have to take special care to balance their stimuli on the psycholinguistic characteristics of interest, as well as to ensure sufficient observations are available for all levels of analysis.

3.6. General discussion

In this experiment, we aimed to investigate the effects of sensorimotor and linguistic distributional information on the categorisation of labels referring to natural objects. Specifically, we aimed to distinguish between previously observed effects of sensorimotor and linguistic information and taxonomic level in object categorisation, by selecting items for which traditional and sensorimotor-linguistic accounts would predict contrasting outcomes. Traditional accounts would predict a basic-level advantage over the superordinate level in most cases, whereas sensorimotor-linguistic accounts might predict that a greater overlap in sensorimotor experience and linguistic distributional knowledge would yield faster and more accurate categorisation regardless of taxonomic level. We anticipated that word frequency would also be likely to affect speed and accuracy of processing the category labels, so they were included as covariates in the null model. We consequently included log frequencies of the chosen and rejected labels as covariate fixed effects in all our analyses.

In our analysis of the choice participants made (analysis 1), we found that on average, participants were more likely to choose a basic-level label. As predicted, the likelihood that participants chose a basic-level label was affected by sensorimotor-linguistic bias, such that participants were less likely to choose a basic-level label when the overlap between the cue and superordinate level category label was greater than that between the cue and the basic-level label. Moreover, we found that word frequency, as anticipated, also affected the likelihood of a basic-level choice. Exploratory analyses revealed furthermore that the effect of word frequency on choice was more extreme when sensorimotor-linguistic bias was pro-superordinate. This suggests that sensorimotor-linguistic bias may induce a superordinate-level advantage (i.e., make it more likely to choose a superordinate over a basic-level label), but only in those specific circumstances where the chosen superordinate label was of low- or average frequency and the rejected basic-level label was of high frequency (see Figure 1, panel 3). As this is an exploratory result, future studies which specifically aim to replicate these effects are necessary before any strong conclusions may be drawn.

In our analysis of the full RT dataset (analysis 2, stage one), we found that the time it took for participants to make a decision depended on the taxonomic level of the label, the frequency of the label *and* its overlap in sensorimotor and linguistic distributional information with the cue. Overall, we found a classic basic-level advantage. Participants were faster to choose a basic-level label compared to a superordinate-level label, and faster to choose any label in the pro-basic compared to the pro-superordinate condition. That is, while participants were slower to choose a basic-level label in the pro-superordinate sensorimotor-linguistic condition relative to the pro-basic condition, they were also slower to choose a superordinate-level label.

To gain a clearer insight into the relative effect of sensorimotor-linguistic bias on the basic-level advantage, we split our dataset of RT into two sets by sensorimotor-linguistic bias. As expected, for the dataset that had a sensorimotor-linguistic bias towards the basic level, we observed a clear basic-level advantage. However, in the dataset that was biased towards the superordinate level, we saw that participants were slightly less slow when selecting superordinate-level labels, but still faster at the basic-level overall.

Contrary to our hypotheses, sensorimotor-linguistic distance condition did not decisively predict categorisation speed or choice better than taxonomic level. Instead, our data showed a strong basic-level advantage in speed and choice. While categorisation was overall faster when sensorimotor-linguistic and taxonomic information pulled in the *same* direction (i.e., in the pro-basic condition) and slower when they pulled in *different* directions (i.e., pro-superordinate), basic-level categorisation was faster than superordinate-level categorisation in both conditions. Furthermore, while sensorimotor-linguistic bias predicted choice over word frequency for the chosen and rejected labels, the basic-level choice was overall still more likely.

At a glance, these results might indicate support for hierarchical accounts of categorisation, which argue the basic-level is the preferred entry-level into a taxonomically organised semantic memory (Jolicoeur et al., 1984). However, this would not explain directly why basic-level categorisation was slower in the pro-superordinate condition, as this suggests that participants were in fact sensitive to the experimental manipulation of overlap in sensorimotor-linguistic experience. That is, when the cue appeared to be biased towards the superordinate category label, it delayed participants in choosing the basic-level category. A limitation of our study is that we did not control for typicality of the cues in both conditions. It is therefore theoretically possible that our pro-superordinate condition contained more items that were atypical of their respective basic and superordinate categories than the pro-basic condition. However, while this could explain why basic-level labels were chosen more slowly in the pro-superordinate condition, it would not explain why superordinate-level labels were chosen more slowly, as typicality effects do not seem to extend to the superordinate level (Murphy & Brownell, 1985). A follow-up study might investigate whether typicality for the cue of its basic and superordinate-level label would predict performance over sensorimotor-linguistic information.

Since our study involved only category *labels*, it is possible that participants simply chose the more frequent label, and were faster to do so than choosing infrequent labels. Since basic-level labels tend to be more frequent (Rosch, Mervis, et al., 1976), it is not always clear whether what we observe is the basic-level advantage, or a simple frequency effect. In our analysis of choice, we observed an interesting effect of choice. When the chosen word was of high frequency, it was more likely to be basic. Conversely, when the rejected word was of high frequency, it was more likely to be superordinate. Given the fact that basic-level labels are generally of higher frequency, this effect is not entirely unsurprising, as both measures are two sides of the same coin. Our exploratory analysis of RT, including the effects of a range of psycholinguistic characteristics of the chosen and rejected labels, showed distinct effects of age of acquisition and word length, while confirming the previously observed effects of taxonomic level of the choice, sensorimotor-linguistic bias and word frequency.

In summary, our study shows that sensorimotor and linguistic information do affect both the decision making and decision-making speed in categorisation. While they do not decisively overturn the basic-level advantage, when overlap with the superordinate alternative is greater, categorisation is nevertheless slowed down when taxonomic and sensorimotor-linguistic information pull in opposite directions. Furthermore, our findings suggest that word frequency may mediate the effects of sensorimotor-linguistic overlap between category and member concepts, opening up potential avenues for future research.

3.7. References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning: From Acquiring Specific Items to Forming General Rules. *Current Directions in Psychological Science*, *21*(3), 170–176. https://doi.org/10.1177/0963721412436806

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

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

Barsalou, L. W., Santos, A., Simmons, K. W., & Wilson, C. D. (2008). Language and Simulation in Conceptual Processing. In M. de Vega, A. M. Glenberg, & A. Graesser (Eds.), *Symbols, Embodiment and meaning* (pp. 245–283). Oxford University Press.

Carota, F., Moseley, R., & Pulvermüller, F. (2012). Body-part-specific Representations of Semantic Noun Categories. *Journal of Cognitive Neuroscience*, *24*(6), 1492–1509. https://doi.org/10.1162/jocn_a_00219

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

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. (2010). Look but don't touch: Tactile disadvantage in processing modality-specific words. *Cognition*, *115*(1), 1–9. https://doi.org/10.1016/j.cognition.2009.10.005

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & 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., & Ramscar, M. (2001). Using Distributional Measures to Model Typicality in Categorization. In *Proceedings of the Twenty-Third Annual Meeting of the Cognitive Science Society* (pp. 226–231). Lawrence Erlbaum Associates.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*(2), 163–201. https://doi.org/10.1037/0096-3445.132.2.163

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

Glass, A., & Holyoak, K. (1974). Alternative conceptions of semantic theory. *Cognition*, *3*(4), 313–339. https://doi.org/10.1016/0010-0277(74)90002-X

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

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

Hampton, J. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition*, *9*(2), 149–156. https://doi.org/10.3758/BF03202329

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

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*(3), 248–277. https://doi.org/10.1037/0096-3445.126.3.248

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243–275. https://doi.org/10.1016/0010-0285(84)90009-4

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

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

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 54–70. https://doi.org/10.1037/0278-7393.23.1.54

Mervis, C. B., & Crisafi, M. A. (1982). Order of Acquisition of Subordinate-, Basic-, and Superordinate-Level Categories. *Child Development*, *53*(1), 258–266. https://doi.org/10.2307/1129660

Murphy, G. L. (2016). Is there an exemplar theory of concepts? *Psychonomic Bulletin & Review*, *23*(4), 1035–1042. https://doi.org/10.3758/s13423-015-0834-3

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 70–84. https://doi.org/10.1037/0278-7393.11.1.70

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical Structure in Concepts and the Basic Level of Categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts and Categories* (pp. 93–131). Psychology Press - Taylor & Francis.

Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 1–20. https://doi.org/10.1016/S0022-5371(82)90412-1

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363. https://doi.org/10.1037/h0025953

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502. https://doi.org/10.1037/0096-1523.2.4.491

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

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, *113*(2), 169–193. https://doi.org/10.1037/0096-3445.113.2.169

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual Structure and the Structure of Concepts: A Distributed Account of Category-Specific Deficits. *Brain and Language*, *75*(2), 195–231. https://doi.org/10.1006/brln.2000.2353

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

van Hoef, R., Lynott, D., & Connell, L. (Manuscript in Preparation). *The Role of Sensorimotor and Linguistic Information in the Basic-level Advantage*.

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

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Chapter 4: Timed Picture Naming Norms for 800 Photographs of 200 Objects in English

4.1. Chapter introduction

This Chapter presents a set of picture naming norms. These norms serve as the basis for the study described in Chapter 5, but also stand on their own as a resource for future research. Research that uses visual stimuli largely relies either on ad-hoc collected materials, or normed picture sets. While there are clear advantages to using normed sets (e.g., they allow for controlling stimuli for multiple psycholinguistic variables) the dominant picture naming norms rely on abstracted, black and white line drawings. Moreover, these norms typically feature low-resolution images which limits their utility on modern high-resolution screens, and/or comprise relatively small item sets (which limits their use to researchers running large-scale experiments). Given the enormous availability of photographic material, and the fact that research suggests photographs offer additional layers of information compared to line drawings, it is surprising that named picture sets containing high-quality photographs are still a rarity. The aim of this paper therefore was to create a comprehensive set of picture naming norms containing high resolution photographs for dissemination to other researchers, and for use in subsequent experiments, as shown by their use in Chapter 5.

Timed Picture Naming Norms for 800 Photographs of 200 Objects in English

Rens van Hoef[1], Dermot Lynott[1] and Louise Connell[1]

[1] Department of Psychology, Lancaster University

Author Note

Rens van Hoef https://orcid.org/0000-0003-1355- 1541

Louise Connell https://orcid.org/0000-0002-5291-5267

Dermot Lynott https://orcid.org/0000-0001-7338-0567

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 682848) to LC.

Correspondence concerning this article should be addressed to Rens van Hoef, Department of Psychology, Fylde College, Lancaster University, Bailrigg, Lancaster, LA1 4YF, UK. Email: r.vanhoef@lancaster.ac.uk

4.2. Abstract

The present study presents a large set of 800 high-quality photographs of 200 natural objects and artefacts spanning a range of categories, with four unique images per object. Participants provided names for each image. We provided recognition latencies for each image, and several normed variables for the provided names: name agreement, h-statistic (i.e., level of naming uncertainty), Zipf log-word frequency and word length. Rather than simply focusing on a single name per image (i.e., the modal, most common, name), analysis of recognition latencies showed that it is important to take into account the diversity of labels that participants may ascribe to each image. The norms therefore provide weighted measures of word length and frequency per image, that incorporate all provided names, as well as modal measures based on the most common name only.

4.3. Introduction

Pictures and photographs of objects are widely used as stimuli in many fields of research, such as perception, memory, cognition and language processing. However, pictures may vary in terms of visual characteristics (e.g., colour, texture), as well as (e.g., concept familiarity) characteristics and of the lexical characteristics associated with the labels they elicit (e.g., word frequency, name agreement (see Alario et al., 2004; Perret & Bonin, 2019 for a review). This variability has prompted researchers to create standardised sets of images, which describe their visual and semantic characteristics, as well as the lexical characteristics of their associated names. Picture naming norms have been used in psycholinguistic (Ostarek & Vigliocco, 2016; Vinson et al., 2015), object recognition (Bramão et al., 2011; Catling et al., 2008; Rossion & Pourtois, 2004) and neuroimaging research (Gerlach, 2009; Thompson-Schill et al., 1997).

To date, the majority of picture naming norms comprise line drawings. Arguably. the most influential set of picture naming norms is the one compiled by Snodgrass and Vanderwart (1980). This set consists of a standardised database of 260 black and white line drawings depicting natural objects and artefacts from a range of categories, and their associated values for image agreement, familiarity and complexity as well as name agreement for their given names in English. This set has since been extended by other researchers, who have included more images and/or collected their names across multiple languages (e.g., Bates et al., 2003; Sanfeliu & Fernandez, 1996; Severens et al., 2005), investigated further psycholinguistic variables (e.g., age of acquisition, Barry et al., 1997) and added naming times (Snodgrass & Yuditsky, 1996). The pervasiveness of black and white line drawings rests on the assumption that they are processed similarly to more realistic depictions (Salmon et al., 2014), while being easier to control. However, as Brodie et al. (1991) note, the adequacy of this approach is rarely addressed in norming studies. Early support for the use of use of black and white over colourised drawings comes from edge-based accounts of object recognition, which argue that object recognition relies primarily on object shape (Biederman, 1987), and that colourised images hold no advantage over grayscale images (Biederman & Ju, 1988). More recently however, research suggests that colour information may in fact improve object recognition for objects for which colour is diagnostic (Bramão et al., 2011; Tanaka et al., 2001; for review and meta-analysis), showing improved recognition accuracy for objects that are associated with a specific colour when they are depicted in that colour (e.g., a banana in *yellow*, a fire engine in *red*; Martínez-Cuitiño & Vivas, 2019; Tanaka & Presnell, 1999). Moreover, Rossion and Pourtois (2004) found that colourised and texturized versions of the Snodgrass and Vanderwart (1980) items were processed faster and more accurately than their non-shaded black and white counterparts, a finding that was replicated by Bonin et al. (2019).

If the addition of colour and texture to black-and-white line drawings improves how quickly and accurately they are processed, it may be the case that photographs, which typically contain colour and texture information as well as luminance and 3D visual cues (e.g., shade), offer a further processing advantage over black-and-white line drawings (Heuer, 2016). Edge-based accounts (e.g., Biederman & Ju, 1988; Ostergaard & Davidoff, 1985), argue objects colour is merely a cue that points towards the more relevant diagnostic aspect of object shape. Indeed, Humphrey et al. (1994) found that participants were faster at naming objects when they were depicted in full-colour rather than grayscale (see also Brodie et al., 1991), and found that this advantage was particularly strong for natural objects, which have stronger shape-colour correlations (e.g., a *daffodil* is always yellow, whereas a *cup* may have any number of colours; see also Therriault et al., 2009), although Humphrey et al. note that colour may still facilitate recognition of familiar objects (e.g., my favourite *coffee cup* is *red*). However, other research suggest that colour is more than a simple cue pointing towards shape, and has shown that object recognition is significantly worse for images that contain only edges compared to full-colour images (Heuer, 2016; Sanocki et al., 1998). Further evidence for the advantage of full-colour photographs over line drawings comes from eye-tracking work (Heuer, 2016) which shows that language-normal participants fixate significantly more on photographs compared to line drawings, suggesting enhanced recognition for photographic stimuli.

Further evidence for the usefulness of photographs compared to line drawings comes from simulated (or grounded, embodied) accounts of conceptual processing. On these accounts, concepts are represented in semantic memory as partial simulations of sensorimotor, affective and other experience with our environment (Barsalou, 1999; Wilson, 2002). Indeed, there is evidence that photographs tap into such simulated information. (Salmon et al., 2014) show that grayscale photographs of manipulable objects are processed faster than black and white line drawings of the same object, although the effect size was small. Taken together, these results suggest that the additional detail present in photographs may indeed facilitate object recognition. While the evidence in favour of photographs does not discount the use of line drawings completely (e.g., line drawings may abstract over features of multiple exemplars whereas photographs can by definition only show one exemplar, Heuer, 2016) it shows that the present skew in picture norms towards line drawings may need to be corrected, especially given the ever-increasing amount of available photographic material.

Previous work has found strong evidence for the effects of uncertainty (as expressed by the h-statistic; Lachman, 1973) and name agreement or codability on response times in picture naming and recognition (e.g., Székely et al., 2003). Other variables typically included in picture naming norms are word frequency (written, spoken) and word length of the modal (most commonly given) name. However, evidence for the effect of these variables on naming latencies is not consistent. For example, while researchers previously found evidence for the effects of written (Snodgrass & Yuditsky, 1996) and spoken word frequency (Ellis & Morrison, 1998) and of word length (D'Amico et al., 2001; Székely et al., 2003) on response latencies in object naming, a recent meta-study (Perret & Bonin, 2019) of 18 normative studies containing black-and-white line drawings found that evidence for the effects of word frequency and length was inconclusive.

One explanation for the lack of a clear effect of word frequency is the fact that the methods used to collect word frequency vary considerably across different studies. Brysbaert et al. (2018) suggest that corpus selection tailored to the language participants are most frequency exposed to allows for better predictions. Of the studies used in Perret and Bonin (2019), five tested undergraduate populations using frequency ratings based on written texts (e.g., novels, essays, poems, dramatic works, non-fiction books, newspaper articles and magazines), which were collected years (e.g., Nishimoto et al., 2005; Pind & Tryggvadóttir, 2002) and sometimes even multiple decades (e.g., Bonin et al., 2002; Perret & Laganaro, 2013) prior to testing, and may not have reflected the language their participants were exposed to. As a consequence, the present study, retrieved names in English from British participants, and used frequencies obtained from a large corpus of British English subtitles (see van Heuven et al., 2014).

Another possibility is that restricting analyses to the effects of the word frequency of only the modal name is too narrow. Since the effects of name agreement and uncertainty indicate that the number of competing names strongly affects performance, it is possible that including the frequency and length of competing names may bolster the predictive power of word frequency and length. As a consequence, the present study includes variables for average word frequency and length per image, weighted by the frequency of each name given to each image.

The present study aims to create a large set of photographic picture naming norms for images divided across a range of artefactual and natural categories at the basic level (see OSF for norms, images and attribution: https://osf.io/vdka2/?view_only=ac6c78793a354a1f95da6c36b5ac6163). In this article, we report the h-statistic (measure of uncertainty in labelling an object; Lachman, 1973), name agreement (percentage of participants that gave the modal name), word length in characters and Zipf log word frequency for every image our set. These variables are consistent with previous picture naming norms, and are linked to the name and response activation stages of picture naming as outlined by Johnson et al., (1996) and Alario et al. (2004). We compare these norms against previous picture naming norms. Furthermore, we explore whether the average word frequency and length of *all* responses given to an image explain response latencies better than word frequency and length of the modal name only.

4.4. Norming Method

Participants. 31 participants (20 female; Mage = 33.61 years old, SD = 11.77) were recruited trough online recruiting platform Prolific (www.prolific.co). Using Prolific's custom pre-screening, we selected participants to be native speakers of English, have normal or corrected-to-normal vision, and no reading impairments (e.g., dyslexia). Participants were paid per completed list (see procedure), starting at £2.50 for submitting one list and £1.00 for every additional submitted list, with submitting the maximum of 16 lists yielding a payment of £17.50. Participants could complete as many lists as they wanted. On average, participants completed 10.7 lists (*SD* = 5.2). Data collection was approved by the Lancaster University Faculty of Science and Technology Research Ethics Committee. All participants read information detailing the purpose of and expectations of the study, and gave informed consent which included the acknowledgement they would be paid for each completed (but not partially completed) list, and the permission to share all anonymised alphanumeric data publicly.

Materials. The stimulus set consisted of 800 full colour images of 200 objects (100 natural objects, 100 artefacts, see supplementals on OSF for the full set of photographs), which depicted only the target object on a white background. The 100 natural objects belonged to 23 basic-level categories (e.g., dog, cat, bird, lizard, fish, insect, tree, vegetable, fruit etc.). The 100 artefact objects belonged to 26 basic-level categories (e.g., boat, box, car, cup, aircraft, watercraft, snowcraft, case, bag, ball, truck, tool, etc.).

We intended these norms to be compatible with Lynott et al.'s (2020) norms of sensorimotor strength. Therefore, we only selected objects whose target name was present in the Lynott et al. set. For each object, we sourced 4 photographs through Google image search. We selected photographs to be free for use with modification (please see norms files on OSF for attribution information for all images), to depict the target object clearly and without obstructions that rendered it unidentifiable, and to have a minimal size of 1024x768 pixels.

We edited the photographs with Adobe Photoshop 2020 (version 21.2.3). Specifically, we cut each target object from the original photograph and placed it central on a white background sized 1920x1080px, with a minimal margin of 200 pixels on every side (see Figure 1, and supplemental materials on OSF). In addition to this, we removed any visible text and distracting objects (e.g., humans on a steamboat).

Figure 1.

Example of processed images for the category dog

Procedure. We designed and hosted the experiment on Gorilla.sc (Anwyl-Irvine et al., 2020). We randomly divided all photographs over 16 lists of 50 trials each. Lists were rotated across participants, where participants were randomly assigned to one of four starting lists (list 1, 5, 9 and 13). If a participant opted to complete more than one list, they were presented with the next list in sequence (e.g., participants who started with list one saw list 2, 3, 4 and so on), until a participant decided they did not want to complete more lists or they had completed all 16 lists.

Participants were instructed that they would be shown a series of photographs, and that each photograph depicted one object. They were asked to press *spacebar* as soon as they recognised the depicted object, and to enter the most appropriate name they could think of in the text box that followed each photograph, and to enter *DK* for 'don't know' if they did not know what an object was. Trials were presented in a randomised order. Each trial was presented centred on a white background in the participant's browser window, and started with a fixation cross for 200ms, followed by the photograph, which remained on screen until participants indicated they recognised and could name it by pressing the spacebar. Once a participant pressed the spacebar, the photograph disappeared and was replaced by a textbox in which they could enter a name.

Our choice to use keyboard responses was largely a consequence of the fact that, at the time of testing, recording of vocal responses – typically used in object naming- was not yet available on the online testing platform we used. In one of the few other picture naming studies that used keyboard responses, Torrance et al., (2018) recorded RT from the onset of the image to the first keypress, as well as mean interkeypress intervals. However, they also noted they had to remove a significant number of trials (15% on average) for being *non-fluent* (e.g., participants used backspace to edit their submissions) which Torrance et al. argued made it hard to associate initial keypress RT with the preparation of the final response. At the time of testing, the online testing platform we used did not allow for the recording of fine-grained keypress responses, including whether participants pressed backspace to edit their responses, which meant we could not test this. As a consequence, we opted to specifically ask participants to press spacebar once they had recognised and could name it. We recorded recognition RT, measured from the onset of the photograph until the participant pressed the spacebar. Furthermore, we recorded the name participants entered.

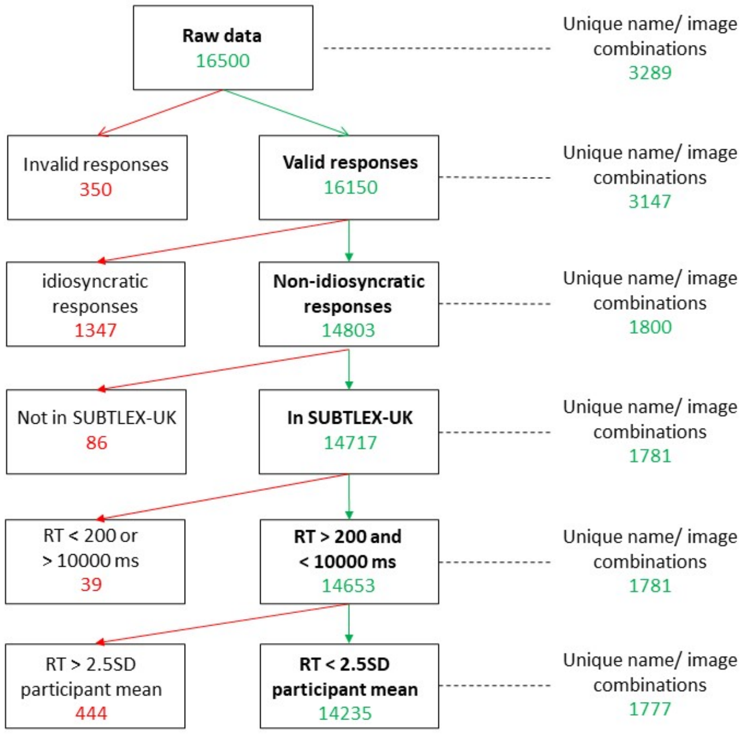
A brief practice session with 4 items (*monkey wrench*, *screwdriver*, *skirt* and *armadillo*) that were not present in the experimental stimuli set nor belonged to the same basic-level category as any of the experimental stimuli, familiarised participants with the experiment procedure. In this section, participants received reminder instructions at each step of the trial (e.g., “Press spacebar as soon as you know what an object is” and “enter the best and/or shortest name that describes the object in the previous picture and press *enter* to submit”). These instructions were absent in the main testing session. At the end of each main testing session, participants were presented with the number of lists they had completed up to that point, and had the option to stop or to complete another list. Testing took approximately 10 minutes including participant information, informed consent and debrief for completing one list, and took an approximate 3 further minutes for each additional list completed.

Coding. We coded responses in three steps. Firstly, we coded *response validity*, coding all non-name responses as *invalid*. This included all instances where participants entered ‘dk’ or a variant thereof (e.g., *na*, *d/k*, *unsure?* etc.), as well as missing or non-word responses (e.g., lkjlk). Out of 16500 responses, we coded 350 responses (i.e., 2.1%) as invalid. Following the procedure outlined in Snodgrass and Vanderwart (1980) these responses were included in the total count of responses for the calculation of % naming agreement (see measures), but excluded for the calculation of the h-statistic. Secondly, following Torrance et al., (2018) we grouped together different spellings of a valid name, and coded them as instances of a correctly spelled group name (e.g., for an image of a chihuahua, responses *chiwawa, chiuaua* and *chihuahua* were all grouped and coded as instances of the name *chihuahua*). This resulted in a total of 3147 unique picture/name combinations.

Next, we coded responses as idiosyncratic if they occurred only once per item. In contrast to previous studies, which sought to validate pre-existing picture sets (e.g., Bates et al., 2003; Rossion & Pourtois, 2004), our study included only images for which (to our knowledge) no previous naming distributions had been determined. As such, we could not discount responses for not matching a predetermined target answer (Snodgrass & Vanderwart, 1980). Instead, we adopted a strategy more akin to the lenient correctness scoring variant in (Snodgrass & Yuditsky, 1996) and included all non-idiosyncratic responses, even if they were clearly erroneous (e.g., *helmet* in response to an image of a *computer mouse*). Out of 16150 valid responses, 14803 responses occurred more than once, corresponding to 1800 unique image/name combinations. 86 of these responses were not present in the frequency database. These responses were included in the calculation of name agreement and the h-statistic, but not in the calculation of weighted average word frequency, weighted average word length and the analysis of response latencies. For the analysis of recognition RT only, we removed outliers for the calculation of average recognition and naming RT, and the analysis of response latencies. Out of 14717 valid, non-idiosyncratic responses that were present in the frequency database, we removed 28 responses with recognition RTs longer than 10 seconds outright. Furthermore, we removed 11 responses with recognition RTs below 200ms (motor error). All participants had a mean recognition RT within 2.5SD of the overall mean, so it was not necessary to remove anyone on that criterion. Finally, we removed 443 recognition RTs for being further than 2.5SD away from the participant mean. The final trial-level dataset consisted of 14235 responses, corresponding to 1777 unique image/name combinations (see figure 2).

Figure 2.

Breakdown of removal of trial-level response data and corresponding image/name combinations.



4.4.5. Variables in the norms. The norms contain the following variables for each image:

Names. We included the most common name produced for each image (i.e., the modal name), as well as all alternative names 1-k (i.e., non-modal names listed in descending order of production frequency).

Agreement. We calculated the percentage of participants that gave the modal name for every image, based on all responses. In the norms, we reported one name agreement percentage per image.

H-statistic. We calculated the h-statistic for every image, based on only valid responses. The h-statistic is a measure of uncertainty in labelling an object (Lachman, 1973). This measure is frequently used in image naming studies to reflect naming agreement across participants, and shown to be a stronger predictor of naming latencies (Bates et al., 2003; Severens et al., 2005; Snodgrass & Vanderwart, 1980; Székely et al., 2003; Torrance et al., 2018). Low values of h represent low uncertainty (i.e., strong agreement) between participants, whereas high scores of h represent high uncertainty (e.g., a h-statistic of 0 means only one name was given for an image). In the norms, we reported one h-value per image.

Latencies. We recorded Recognition RT, measured from the onset of the image to the onset of a valid keypress indicating a participant had recognised the object. In the image-level norms, we included the overall Recognition RT per image, averaged over all responses after outlier removal (see coding, above).

Word frequencies. We retrieved Zipf log word frequency for all spelling-corrected, valid non-idiosyncratic unigram names from a large database of word frequencies derived from a corpus of subtitles in British English (e.g., van Heuven et al., 2014). The Zipf scale is a logarithmic scale, ranging from 1 to 7, with 1 corresponding to a frequency of 1 per 100 million words (very low frequency, e.g., *antifungal*), and 7 corresponding to a frequency of 1 million per 100 million words (very high frequency, e.g., *and*). For bigram names, we calculated Zipf log frequencies from the raw frequency counts for two-word phrases in the bigram-frequency database derived from the same large subtitle corpus. The average Zipf Log frequency across all unique, non-idiosyncratic responses was 3.92 (SD = .94), with Zipf log frequencies ranging from 1.00 (*money tin*) to 6.56 (*drum*). In the norms, we reported Zipf log word frequency for the modal name (Modal word frequency), as well as the average Zipf log word frequency of all names produced for an image, weighted by their relative production frequency (weighted word frequency). In addition to this, we included Zipf log word frequencies for each non-idiosyncratic non-modal name (alternative name word frequency).

Word lengths. We calculated word length as the number of letters a participant entered, excluding the single space in the middle of bigram responses (e.g., *palm tree* had a length of 8 letters). In the norms, we reported response length for all non-idiosyncratic modal and non-modal names for every image, as well as the average name length of all names produced for an image, weighted by their relative production frequency (Weighted word length). In addition to this, we included word lengths for each non-idiosyncratic non-modal name.

4.4.6. Summary of variable characteristics.

The median number of valid, non-idiosyncratic responses per image was 19 (min = 3, max = 22). The median number of unique names per image was 2, (min = 1, max = 6). We calculated summary descriptive statistics for all variables, per image and per name (see supplementals). Table 1 shows examples of images with high, average and low agreement, divided by natural items and artefacts. Table 2 shows overall descriptive statistics for each %variable (naming agreement, h-statistic, word length, word frequency and response time) for all responses, and for modal responses in particular.

Table 1.

Examples of the listed names for artefacts and natural objects, ranging from low to high name agreement and h-statistic.

	Image name	Listed names [a]	Name agreement	H-statistic [b]
	Bobsled_4	Ski, bobsled, jet ski, sled, sledge, *bobsleigh, ski car, skiddo, snow ski, snowmobile*	18.20 %	3.21
	Sponge_2	Coral, plant, urchin, *anemone, fruit, sea animal, sea urchin, vegetable, sea plant*	38.09 %	3.08
	Mantis_3	Insect, cricket, grasshopper, mantis, *bug, fly, praying mantis, stick insect*	33.33%	2.58
	Minivan_2	Car, van, *minibus, MPV, SUV, campervan, minivan, people carrier, people wagon*	35.00%	2.56
	Chicken_4	Chicken, hen, *bird*	52.38%	1.22
	Briefcase_3	Briefcase, case, suitcase	61.90	1.21
	Gecko_4	Lizard, gecko, reptile, *iguana*	75.00 %	1.19
	Wineglass_4	Glass, wine glass, *cup*	65.00 %	1.14
	Stapler_2	Stapler, *staple*	85.71	0.28
	Spaniel_3	Dog, *spaniel*	95.45	0.27
	Clock_1	Clock	95.23	0.00
	Sycamore_1	Tree	100.00	0.00

[a] Names listed in order of production frequency; first listed name is the modal name; idiosyncratic responses are italicised.
[b] Note that h-statistic is calculated from only *valid* responses, whereas name agreement is calculated from all responses.

Table 2. Image-level summary statistics: average name agreement, h-statistic, modal word frequency, modal word length, weighted word frequency, weighted word length and modal production frequency.

Variable	Mean	Median	SD	Min	Max
Name agreement	65.20	65.00	22.17	15	100
H-statistic	1.22	1.19	0.77	0.00	3.62
Recognition RT	1074.38	1020.47	252.32	690.23	2403.34
Naming RT	95.37	94.90	7.41	75.98	198.68
Modal name word frequency	4.19	4.27	0.80	1.60	5.44
Weighted average word frequency	4.10	4.073	0.63	2.03	5.44
Modal name word length	5.50	5.00	2.18	2.00	12.00
Weighted average word length	5.62	5.38	1.75	2.94	12.27
Modal name production frequency	13.82	14.00	4.61	2.00	22.00

4.5. Comparison with previous norms

Table 3 contains a comparison between recent and relevant other norming studies and the present study for all variables present in this work. We chose to compare our study to six recent norms. Of these norms, four used photographic stimuli (Adlington et al., 2009; Brodeur et al., 2010; Moreno-Martínez & Montoro, 2012; Navarrete et al., 2019), and two used line drawings (Székely et al., 2003; Torrance et al., 2018). Of the studies using photographic stimuli, two were English-language norms, and two were other languages (Spanish, Italian). Both studies using line-drawings were multi-language projects, however for the sake of comparison we only report values for their English-language components here.

The picture naming agreement in the present norms is generally on par with existing photographic picture naming norms with regards to % agreement and the h-statistic. At 65.20 % name agreement and a h-statistic of 1.14, the present study has slightly better name agreement than the BOSS set of normed stimuli by Brodeur et al., (2010, 2014), and the Italian normed set by Navarrete et al. (2019). By contrast, naming agreement in the present study was slightly lower (i.e., lower name agreement and higher h-statistic), compared to a few other normative studies using photographs (e.g., (Adlington et al., 2009; Moreno-

Martínez & Montoro, 2012). This might be due to the larger number of images in the present study, as well as the inclusion of hard to identify objects such as *sea urchin* and *snowmobile*. As for all studies using photographs, the present study had lower naming agreement than studies using line drawings.

A number of variables recorded in the present study have no counterpart in previous norms, and thus cannot be compared. For example, other studies did not consistently report word frequency of the modal (or non-modal) names. Moreover, where they did, the methodologies for retrieving frequencies varied. For example, Adlington et al., (2009) and Moreno-Martínez and Montoro (2012), used the log-transformed number of hits for their names (in English and Spanish respectively) in a popular search engine as their estimate, whereas and Navarrete et al. (2019) report a traditional corpus-derived on the natural log scale. By contrast, the present study used Zipf log word frequencies, derived from a large corpus of British subtitles (see van Heuven et al., 2014). Although word length of the modal name can be easily extracted post-hoc, the present study also reports word length of the competing non-modal names, as well as the average length weighted over all responses per image, which are not consistently reported in other studies. Furthermore, in contrast to other studies, which use voice recordings or a combination of first keypress and interkey intervals (e.g., Torrance et al., 2018), the present study recorded RT from the moment a picture appeared on screen to the moment they pressed a key to indicate they had recognised it and could start naming it (recognition RT).

Table 3.

Comparison of summary statistics with other norming studies

	N	Name agreement	H-statistic	Language	Stimulus type
Present study	800	65.20(22.17)	1.22(.77)	English	Photographs
Adlington et al., (2009)	147	67.61(26.99)	1.11(.89)	English	Photographs
Brodeur et al. (2010, 2014)	1469	59.00(25.00)	1.86(1.08)	English	Photographs
Moreno-Martínez & Montoro (2012)	360	72.00(28.00)	.94(.87)	Spanish	Photographs
Navarrete et al. (2019)	357	56.20(35.45)	1.49(1.01)	Italian	Photographs
Torrance et al. (2018)	260	85.00(19.00)	.66(.75)[a]	English [b]	Line drawings
Szekely et al. (2003)	520	89.00	.67(.61)	English [b]	Line drawings

Notes: Table reports average values and standard deviations (in parentheses).
[a] Torrance et al. (2018) report median H-statistic rather than mean.
[b] Results from English language components in multi-language picture naming studies.

4.6. Recognition latencies

In this study, we recorded the time it took participants to indicate they had recognised an image and could name it (recognition RT). While this measure deviated from previous work (e.g., Barry et al., 1997; Snodgrass & Yuditsky, 1996; Székely et al., 2003), which relied on voice recordings, it is comparable to the first-keypress measure used in Torrance et al. (2018). Torrance et al., (2018) found that first-keypress RT decreased as uncertainty (h-statistic) decreased. This finding mirrored previous work done using voice recordings (e.g., Barry et al., 1997; Székely et al., 2003). Here, we investigated whether uncertainty similarly affected the recognition RT collected in our study; with a higher h-statistic resulting in slower recognition RT.

In addition to this, previous work has found variable evidence for the effects of word frequency and length. This may partially be explained by the variability in the measures used to record word frequency. However, it may also be the case that the usual practice of predicting picture naming latencies through psycholinguistic properties of the modal name *only* can be improved by using weighted measures of all names produced in response to a given image. As a consequence, we also investigated whether recognition RT was better predicted by the Zipf log word frequency and word length of the modal response (modal word frequency/ modal word length) or by the weighted average word frequency and length of *all* responses (weighted word frequency / weighted word length).

4.6.1. Method

Materials. For this analysis, we used the image-level norms. Recognition RT was retrieved for every image, by averaging over the recognition RT for all its associated responses (M = 1074.38, SD = 252.32), Previously, we retrieved Zipf log word frequency and word length in letters for both modal and nonmodal names for each image (see norming method). For three images (flytrap_1, flytrap_3 and flytrap_4), frequency information was not available for the modal name. As a consequence, these images were excluded from this analysis. Non-modal responses for which frequency information was not available were also excluded from this analysis.

In this analysis, we included h-statistic (M = 1.22, SD = .77), modal word frequency (M = 4.19, SD = .80) and modal word length (M = 5.47, SD = 2.15). Furthermore, we included the weighted average word frequency (M = 4.11, SD = .62) and weighted average word length (M = 5.61, SD =1.74), calculated across all non-idiosyncratic responses given to each image. The final image-level dataset used in this analysis thus consisted of 797 images and their associated variables. All independent variables except name agreement and h-statistic were centred around their mean.

Analyses.

In order to determine whether increasing name agreement and decreasing h-statistic reduced recognition RT, we ran a linear regression with recognition RT as the dependent variable, and h-statistic as the independent variable.

Furthermore, in order to determine whether production frequency weighted word frequency and length were better predictors of recognition RT than modal word frequency and length, we ran two separate linear regressions, both with recognition RT as dependent variable. In the first regression, we included independent variables of modal word frequency (range = -2.59, 1.25), and modal word length (range = -3.50, 5.50). We inspected the coefficients for the direction and strength of the effects, to determine whether greater word length and lower word frequency resulted in longer recognition RT. In the second regression, we included the independent variables of weighted word frequency (range = -1.90, 1.34), and weighted word length (range = -2.68, 6.64), here too, we inspected the coefficients for direction and strength of the effects. Finally, we used Bayesian model comparisons with Bayes Factors calculated from the BIC (see Wagenmakers, 2007) to test whether the data was predicted best by the best-fitting model from set one, or the best-fitting model from set two.

4.6.2. Results

Average recognition RTs were calculated per image. Outlier RTs were removed as detailed in above (see norming method).

H-statistic. A regression model of Recognition RT including the independent variable of h-statistic was significant (F (1, 795) = 372.60, p <.001), with an adjusted R^2 of .32. Participants were up to 667.32 ms slower to press spacebar in response to images with the highest h-statistic (3.62), see Table 4 for coefficients.

Modal versus weighted word frequency and length. A multiple regression of recognition RT including modal word frequency and word length was significant (F (2, 794) =38.13, p <.001), with an adjusted R^2 of .085. Participants were up to 118.87 ms faster to press spacebar in response to images that for which the modal name was a high-frequency compared to average-frequency word, but also up to 17.95 ms faster to recognise objects in images when the modal name was longer than average (see Table 4 for model coefficients.

A second multiple regression of recognition RT, including weighted mean word frequency and length was also significant (F (2, 794) =57.52, p <.001), with an adjusted R^2 of .124. Participants were up to 167.39 ms faster to press spacebar in response to images that on average received higher-frequency compared to average frequency Furthermore, participants were again also slightly faster to recognise objects in images that on average received longer compared to average names

Table 4.

Linear regression coefficients of recognition RT for modal and weighted word frequency and length.

	Unstandardized β	SE	t	p
H-statistic	184.50	9.56	I19.30	<.001
Modal word frequency	-118.87	14.46	-8.22	<.001
Modal word length	17.95	5.37	-3.34	<.001
Weighted word frequency	-168.54	17.62	-9.56	<.001
Weighted word length	-16.07	6.25	-2.57	.010

Bayesian model comparisons showed that the model based on the weighted means of all picture names per image was over one million times better (BF_{10} = 10799400) than the model based only in the modal (i.e., most common) picture name.

In summary, word frequency and length of the modal response predicted how quickly participants recognise an object, but not as well as the weighted average of word frequency and length of all responses given to an image. Moreover, this suggests that the weighted variables provided in our norms may be more useful to researchers than merely focusing on the modal name.

4.7. Conclusions

The description of the norming methods and our analyses of response latencies in this work illustrate the potential value of our picture-naming norms as stimuli in behavioural experiments. In contrast to other influential norms sets such as the Snodgrass and Vanderwart set, as well as the extended set presented by Szekely et al. (2003), the present study contains high-resolution photographs rather than line drawings. Furthermore, in contrast to other norms which have used photographic stimuli (Brodeur et al., 2010, 2014) the present norms contain measures of word frequency and length for the modal name, as well as their average per image weighted by the frequency with which a name was given. Summary statistics indicate that our study is comparable to similar naming studies, with the added benefit that our set contains norms for multiple images of the same object. Our analysis of the effects of word frequency and length of the modal name, as well as the average word frequency and length of all names per image, shows that word frequency and length affect the speed with which participants recognise an image. Furthermore, our analysis shows that incorporating the average word frequency and length of *all* names given to an image weighted by their naming frequency improves their predictive power, suggesting that restricting analyses to the modal names may be too limited.

The present work is compatible with a number of pre-existing norms on word characteristics as well as further conceptual processing, and may thus be used in concurrence with other datasets. Firstly, our work is largely compatible with the frequency database from which we retrieved Zipf log word frequencies, and which contains a range of other word characteristics that may be used in concurrence with the stimuli presented here. Furthermore, out of 164 unique modal responses, 158 responses (96.34 %) were present in the Lancaster sensorimotor strength norms (Lynott et al., 2020), as well as the concreteness norms by Brysbaert et al., (2014). As such, our set may be a valuable resource for studies aiming to investigate aspects of sensorimotor grounding using photographical stimuli in concurrence with norms of sensorimotor and or linguistic distributional experience.

We believe the norms presented here are a useful resource for researchers interested any aspect of object recognition and naming, and will allow researchers more choice and control over the selection of their stimuli.

4.8. References

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of Clinical and Experimental Neuropsychology*, *31*(6), 731–753. https://doi.org/10.1080/13803390802488103

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140–155. https://doi.org/10.3758/BF03195559

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart Pictures: Effects of Age of Acquisition, Frequency, and Name Agreement. *Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 560–585. https://doi.org/10.1080/027249897392026

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380. https://doi.org/10.3758/BF03196494

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS One*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLOS ONE*, *9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, *120*(1), 113–139. https://doi.org/10.1037/0033-2909.120.1.113

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, *99*(2), 199–208. https://doi.org/10.1037/h0034633

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An Ecological Alternative to Snodgrass & Vanderwart: 360 High Quality Colour Images with Norms for Seven Psycholinguistic Variables. *PLOS ONE*, *7*(5), e37527. https://doi.org/10.1371/journal.pone.0037527

Navarrete, E., Arcara, G., Mondini, S., & Penolazzi, B. (2019). Italian norms and naming latencies for 357 high quality color images. *PLOS ONE*, *14*(2), e0209524. https://doi.org/10.1371/journal.pone.0209524

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's Object Pictorial Set: The Role of Surface Detail in Basic-Level Object Recognition. *Perception*, *33*(2), 217–236. https://doi.org/10.1068/p5117

Severens, E., Lommel, S. V., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*(2), 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Snodgrass, J., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Snodgrass, J., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 516–536. https://doi.org/10.3758/BF03200540

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 621–633. https://doi.org/10.3758/BF03195542

Torrance, M., Nottbusch, G., Alves, R. A., Arfé, B., Chanquoy, L., Chukharev-Hudilainen, E., Dimakos, I., Fidalgo, R., Hyönä, J., Jóhannesson, Ó. I., Madjarov, G., Pauly, D. N., Uppstad, P. H., van Waes, L., Vernon, M., & Wengelin, Å. (2018). Timed written picture naming in 14 European languages. *Behavior Research Methods*, *50*(2), 744–758. https://doi.org/10.3758/s13428-017-0902-x

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

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

Chapter 5: The Role of Sensorimotor and Linguistic Distributional Information in Ultra-Rapid Categorisation

5.1. Chapter introduction

In previous Chapters, I explored the effects of sensorimotor and linguistic distributional information on categorisation performance. In Chapter 2, I found that overlap in sensorimotor and linguistic distributional information between category and member concepts explains processing advantages in speeded picture categorisation. Nevertheless, sensorimotor information was the strongest predictor out of the two.

Of note is that sensorimotor-linguistic accounts argue that the conceptual system dynamically adapts to task demands (i.e., in its reliance on sensorimotor vs. linguistic information). A possible explanation for the findings of Chapter 1 lies in the notion that speeded picture categorisation may have allowed participants to rely on visual aspects of simulations activated by the category label. The task allowed enough time for a sensorimotor simulation to form in full, and so participants had little need to rely on a linguistic shortcut of "good-enough" distributional representations. In the study reported in Chapter 5, I therefore identify what happens when the time for visual processing is cut short, by testing this using backwards masking following ultra-rapid image display.

The Role of Sensorimotor and Linguistic Distributional Information in Ultra-Rapid Categorisation

Rens van Hoef[1], Dermot Lynott and Louise Connell[1]

[1] Department of Psychology, Lancaster University

Author Note

Rens van Hoef https://orcid.org/0000-0003-1355- 1541

Louise Connell https://orcid.org/0000-0002-5291-5267

Dermot Lynott https://orcid.org/0000-0001-7338-0567

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 682848) to LC.

Correspondence concerning this article should be addressed to Rens van Hoef, Department of Psychology, Fylde College, Lancaster University, Bailrigg, Lancaster, LA1 4YF, UK. Email: r.vanhoef@lancaster.ac.uk

5.2. Abstract

When given unrestricted time to process an image, people are faster and more accurate at making categorical decisions about a depicted object (e.g., *Labrador*) if it is close in sensorimotor and linguistic distributional experience to its target category concept (e.g., *dog*). In this preregistered study, we examined whether sensorimotor and linguistic distributional information affect object categorisation differently as a function of time available for perceptual processing. Using an ultrarapid categorisation paradigm with backwards masking, we systematically varied onset timing (SOA) of a post-stimulus mask (17 – 133ms) following a briefly displayed (17ms) object. Results suggest that linguistic distributional distance between concept and category (e.g., *Labrador* → *dog*), affects categorisation accuracy and RT even in rapid categorization. While sensorimotor distance also affected accuracy, it did not do so above and beyond linguistic distance. Finally, these effects do not vary systematically by SOA. These findings support the role of a linguistic shortcut (i.e., using linguistic distributional instead of sensorimotor information) in rapid object categorisation.

5.3. Introduction

Imagine yourself walking in a forest. From the corner of your eye, you suddenly spot something: a scurry of movement, a glimpse of speckled brown. Was that fur? Were there feathers? Your mind races through different options to determine what is hiding in the bushes, matching perceptual cues to conceptual representations. But which representation is activated first?

Traditional accounts of object categorisation predict that on average, people are most likely to class objects at an intermediate level of abstraction, i.e., *bird*, *dog*, or *rabbit*, known as the basic-level advantage (Rosch et al., 1976). The basic-level advantage is one of the most reliable effects in object categorisation research, supported by evidence from decades of behavioural research showing that people are faster and more accurate at categorising objects at this intermediate level of abstraction (e.g., *bird*) than they are at categorising objects at more abstract (e.g., *animal*) or more specific (e.g., *robin*: K. E. Johnson & Mervis, 1997; Jolicoeur et al., 1984; Murphy & Lassaline, 1997, 1997; Murphy & Smith, 1982; Rogers & Patterson, 2007; Rosch, Mervis, et al., 1976; Tanaka & Taylor, 1991).

A considerable portion of the empirical evidence for the basic-level advantage comes from speeded category verification tasks, in which participants are given relatively long or unlimited time to verify whether an object photograph matches to a previously displayed category name. However, there is some evidence that the classic advantage the basic level has over categorisation at a more abstract (i.e., superordinate) level may be reversed as a function of the time available for perceptual processing. When perceptual processing is limited, superordinate-level categorisation is faster and more accurate than categorisation at the basic level (e.g., participants are faster and more accurate to identify a specific object as an *animal* than they are to identify it as a *dog*). Researchers observed that participants are able to correctly distinguish between rapidly flashed photographs (20ms) of animals and non-animals as fast as 250ms (manual response: VanRullen & Thorpe, 2001), which is close to the fastest possible manual response. At that rate, Mack and Palmeri (2011) argue, it is unlikely a basic-level response would be significantly faster. However, it appears that the classic basic-level advantage not only disappears but may even give way to a superordinate-level processing advantage when objects are presented briefly, followed by a perceptual mask, or when given limited time to respond (Macé et al., 2009; Mack et al., 2008; Mack & Palmeri, 2011, 2015).

The superordinate-level advantage observed in ultra-rapid categorisation is not readily explained by traditional, feature-based accounts of categorisation, such as network (e.g., Collins & Loftus, 1975; Jolicoeur et al., 1984) and differentiation accounts (Cree & McRae, 2003; Markman & Wisniewski, 1997; Murphy & Brownell, 1985; Murphy & Lassaline, 1997; Murphy & Smith, 1982). For example, hierarchical accounts argue that the basic-level may be viewed as the most common 'entry-point' into semantic memory. While this account supports the notion that the entry-level may shift downwards (i.e., to the more specific, subordinate level) depending on object characteristics (e.g., typicality: Jolicoeur et al., 1984) or subject expertise (K. E. Johnson & Mervis, 1997; Tanaka & Taylor, 1991), the possibility of an upwards shift (i.e., to the more abstract superordinate level) is dismissed. Rather, it is argued, basic-level identification is a necessary part of the two-step process required to identify an object at the superordinate level (i.e., you first identify an object as *dog,* after which you infer that it must also be an *animal*: Anderson & Reder, 1974), making a superordinate-level advantage all but impossible. differentiation accounts of categorisation meanwhile argue that people are unlikely to have a single, holistic perceptual representation or prototype of a superordinate concept, such as *animal* (Murphy & Brownell, 1985; Murphy & Smith, 1982). Instead, according to this account, superordinate concepts are represented in semantic memory through multiple perceptual representations. For example, the superordinate-level category *animal* contains *whales* and *mice*. As a result, Murphy and

Smith (1982) argue, the category label *animal* cannot activate a unitary representation that prepares participants for object categorisation. Finally, in contrast to hierarchical accounts, differentiation accounts do not require the activation of a basic-level prior to a superordinate-level concept, but rather require the activation of more perceptual representations, since people are unlikely to have a single perceptual representation for more abstract categories (e.g., *animal*; Murphy & Smith, 1982).

Previously (van Hoef et al., 2022a - see Chapter 2) we have proposed an account of categorisation that contrasts with hierarchical and prototype accounts, based on linguistic-simulated accounts of conceptual representation. These accounts argue that concepts generally comprise both sensorimotor (perception-action experience of the world) and linguistic distributional information (i.e., our knowledge of the relative distribution of words in language: Barsalou, 2008; Connell, 2019; Connell & Lynott, 2014; Louwerse, 2011). Sensorimotor simulation may be defined as (partial) reactivation of systems involved in online processing of perceptual and motor experience during offline retrieval (Barsalou, 1999). As such, a sensorimotor representation for a given concept (e.g., *dog*) may consist of perceptual (e.g., colour of the dog's fur, sound of its bark), motor (e.g., throwing a ball for the dog to fetch), affective (e.g., fondness of dogs) or other information. Evidence for simulated representations spans neuroimaging (Aziz-Zadeh et al., 2006; Carota et al., 2012; Goldberg et al., 2006; Hauk et al., 2004) and behavioural studies (Connell & Lynott, 2010; Dils & Boroditsky, 2010; Zwaan & Taylor, 2006), citing strong links between perceptual experience and conceptual representations. Linguistic distributional information meanwhile stems from people's inherent sensitivity to the statistical properties of language, which allows them to automatically infer meaning from the distributional characteristics of a word, for example by repeated exposure to its context. Linguistic distributional information has been shown to reliably predict conceptual processing (Connell & Lynott, 2013; Goodhew et al., 2014;

Louwerse, 2011; Louwerse & Jeuniaux, 2010). While linguistic-simulated accounts agree on the importance of both linguistic and sensorimotor information to conceptual processing, they vary in the relative importance they attribute to either information type. Broadly speaking, Barsalou (2008) emphasises sensorimotor simulation over linguistic distributional information, whereas Louwerse (2011) argues that in many cases, only minimal sensorimotor grounding is required, with linguistic distributional information doing most of the heavy lifting in conceptual processing. Finally, Connell and Lynott (2014) argue that neither sensorimotor nor linguistic distributional information is absolutely more important to conceptual processing than the other. Rather, they argue that both information types are part of an adaptive conceptual system, and that the relative importance of sensorimotor and linguistic distributional information may change depending on task constraints.

In a previous study (van Hoef et al., 2022a - see Chapter 2), we found evidence for the hypothesis that sensorimotor and linguistic distributional information affect categorisation performance in a speeded category verification task. We found that when given unrestricted time to process an object photograph (e.g., a photograph of a *Labrador*), participants are faster and more accurate to categorise it when it overlaps in sensorimotor and linguistic distributional information with its category label (e.g., *dog* or *animal*) to a greater extent. As a result, we proposed a sensorimotor-linguistic model of object categorisation that is reminiscent of the feature-based preparation model from prototype theory (Murphy & Smith, 1982), in that category labels may activate semantic representations, which are verified upon seeing a potential category member. When the representation activated by a category label does not match the displayed object, additional processing is required. Crucially, where the preparation model proposes different strategies for categorising objects at the superordinate (i.e., requiring the extraction of additional perceptual features) or subordinate level (i.e., activating additional discriminatory criteria), the sensorimotor-linguistic model does not make a priori assumptions about the processing of concepts at various taxonomic levels. Rather, the sensorimotor-linguistic model assumes that superordinate concepts like *animal* are representations in their own right, and holds that different taxonomic category labels (e.g., *dog*, *animal*, or *Labrador*) may activate different sensorimotor and linguistic distributional representations which overlap with representations activated by a member photograph or label (e.g., *Labrador*). The more the representations overlap, the less additional representational processing is required.

The sensorimotor-linguistic model of object categorisation does not necessarily predict that one taxonomic level is preferred over the other (although it may be the case that a latent taxonomic structure may arise spontaneously from sensorimotor information (Connell et al., 2021). The sensorimotor-linguistic model does not directly predict a basic-level advantage in speeded category verification, nor a superordinate-level advantage in ultra-rapid categorisation. It simply predicts an advantage for concepts that are more similar in sensorimotor experience and linguistic distributional knowledge to their categories. However, as Connell and Lynott argue (2014), the relative importance of either information type may depend on task demands, such as the amount of time available for perceptual processing. As such, a sensorimotor-linguistic model of rapid object categorisation might predict that as more perceptual information becomes available (e.g., through longer display times), the verification of the sensorimotor representation activated by a category label becomes more effective. This effect would be mirrored for linguistic distributional information, as more perceptual information would make an object easier to recognise and therefore to retrieve a label for it. However, since previous work on speeded category verification (i.e., with unlimited time to recognise and respond to an image) suggests that linguistic distributional information is not as strong a predictor of categorisation performance as sensorimotor information, linguistic distance effect may also be reduced here.

The current study

The present study (https://osf.io/vdka2/?view_only=ac6c78793a354a1f95da6c36b5ac6163) aims to determine the role of sensorimotor and linguistic distributional information in the categorisation of briefly displayed objects with backwards masking, where every stimulus is followed by a visual mask after a predetermined amount of time (stimulus onset asynchrony, SOA, which has been shown to reduce the amount of information available from neurons in the visual system; (Rolls et al., 1999). Based on the results from similar ultra-rapid categorisation studies (e.g., Bacon-Macé et al., 2005, 2007; Mack & Palmeri, 2011, 2015), we expect that categorisation performance (accuracy; response times) will improve and then level off as the time for accumulating perceptual information increases. That is, we expect accuracy to increase, and response times to decrease, as more time is allowed to pass between the presentation of a photograph and the onset of a perceptual mask (i.e., by increasing SOA). Furthermore, in line with simulated-linguistic accounts of conceptual processing (Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse, 2011) and our previous findings (van Hoef et al., 2022a), we expect that sensorimotor and linguistic distributional information will affect categorisation performance, such that a greater overlap in both information types between a category and member-concept will result in lower RT, and higher accuracy. Crucially, we expect that the relative impact of sensorimotor compared to linguistic distributional information will vary with the amount of time available for perceptual processing. We expect that the effect of both information types will increase as more time for perceptual processing becomes available, but that the effect of linguistic distributional information will require longer SOA to manifest than the effect of sensorimotor experience, as implicit activation of an object label may be less likely until there is sufficient time to process an object visually. Specifically, we expect to observe an effect on accuracy and RT of sensorimotor but not linguistic distributional information at shorter SOA, but an effect of both information types at longer SOA.

5.4. Method

Participants. We recruited 128 participants (65 female, *M*age = 37 years, *SD* = 11.76) through online recruitment platform Prolific (www.prolific.co). Participants received £4.25 in compensation for their time. Using Prolific's screening tool, we selected participants who were native speakers of UK English, had normal, or corrected-to-normal vision, had not completed any other of our studies, and had a Prolific approval rate of over 95%. In addition to this, due to the nature of our experiment requiring rapid visual displays, participants with a history of photosensitive epilepsy or a related neurological condition which might cause seizures in reaction to rapidly flashing imagery were not allowed to take part. Participants were explicitly asked to confirm they did have no such condition as part of the informed consent form, and were informed of the potential risks in the study advert, a warning screen which loaded once the experiment commenced, as well as the participant information sheet.

As pre-registered, we established sample size through sequential hypothesis testing using Bayes Factors (Schönbrodt et al., 2017). We set the lower boundary at 48 participants, and continued testing until a maximal upper boundary of 128 participants was reached or the critical BF10 of >3 was achieved. We calculated Bayes factors for analysis B steps 4 through 10 (see analysis) explaining variance in RT and accuracy from the interaction between stimulus onset asynchrony and sensorimotor and/or linguistic distance. At 48 participants, we found strong evidence against the inclusion of sensorimotor and linguistic distance, however a number of the interaction models did not converge. As a result, we deviated from our pre-registration, and tested up to our upper boundary to see if increasing sample size would counter these issues. At 128 participants, some models still had convergence issues, which we will highlight in the results section below. None of the best-fitting models for which we reported coefficients failed to converge.

Prior to each sequential hypothesis test we had to replace a number of participants, for various reasons. Firstly, we replaced 17 participants due to technical malfunctions (e.g., issues with trial display, extreme delays in presentation etc.) Secondly, we pre-registered several replacement criteria. To ensure participants would be able to see the rapidly flashed photographs, we included a (pre-registered) screening task prior to the main experiment. We presented participants with ten geometric shapes in various colours, paired with a binary choice of shape and colour (e.g., a red circle followed by the question 'did you see a CIRCLE or a SQUARE'). Each shape was displayed for 17 ms, which was equal to the display duration in the main experiment, and corresponded to one screen refresh at 60 Hz, the most common LCD refresh rate. We replaced 12 participants for answering more than the pre-registered two questions incorrectly. Third, we pre-registered that we would replace all participants who responded incorrectly to more than 30 % of the trials, or reached trial timeout (> 1500 ms) on more than 20% of trials. We replaced 50 participants on this basis.

Ethics and consent. The study received ethical approval from the Lancaster University Faculty of Science and Technology Research Ethics Committee. All participants received information detailing the purpose and expectations and potential risks of the study before giving informed consent to take part. Consent included agreement to share publicly all alphanumeric data in anonymised form, as well as the explicit confirmation that the participant did not have a history of photosensitive epilepsy or any related neurological conditions that could cause seizures in response to flashing imagery. Participants were informed that they could leave the experiment at any point without consequences.

Materials. Test items consisted of 320 full-colour photographs of 80 specific object concepts, where each object concept was represented by four different instances (e.g., photographs of four different Labradors represented the object concept Labrador). Half of these objects were natural objects, half were artefacts. Object concepts were evenly divided over 8 basic-level categories (*dog*, *bird, flower, tree, car, boat, box* and *cup*) and 4 superordinate-level categories (*animal, plant, vehicle* and *container*). Photographs for each object concept were retrieved from a set of picture naming norms generated in a previous study (van Hoef et al., 2022b - see Chapter 4). Each photograph depicted the target object central on a white background, without any visible labels, sized 1920x1080 pixels, with a minimal margin of 200 pixels on each side. Pilot testing revealed that high resolution paired with the large number of photographs caused display issues (e.g., photographs did not display consistently), so we reduced the resolution to 800x450 pixels, in order to reduce file size (see supplemental materials for all stimuli used in this study).

For each basic and superordinate category, we selected an additional 40 photographs of 10 objects – four per object, for a total of 480 distractor photographs retrieved from the same picture naming norms set (van Hoef et al., 2022b – Chapter 4). We selected basic-level distractors to be part of the same superordinate (e.g., *animal*) but different basic-level category, and to match some of the visual features of the test objects (e.g., we selected *gecko* as a distractor for *bird* because it belongs to the superordinate category *animal* and is small and brightly coloured). We selected superordinate-level distractors to be part of the same global (e.g., *artefact*) but different superordinate category (e.g., household appliances as distractors for *vehicle*), but to match visual features of the test objects (e.g., we selected *vacuum cleaner* as a distractor for *vehicle* because it has wheels and has an asymmetric oblong shape and or metallic colour). Crucially, because our critical predictors included a measure of sensorimotor experience (see critical predictors) we selected both target- and distractor-objects to have names specified in the picture naming norms (van Hoef et al., 2022b – Chapter 4) which were present in the Lancaster Sensorimotor norms (Lynott et al., 2020).

We created 16 perceptual masks from 6 additional images, none of which were present in either the test or filler set. We used Adobe Photoshop's (version 21.1.2) cut-out filter and liquify forward warp tool to distort these images until no recognisable shapes were visible (see Figure 1). All mask images can be found in the supplementals on OSF.

Figure 1.

Mask creation process, showing an example of an original image, an image to which a cut-out filter was applied, and the final mask after using the forward warp tool.

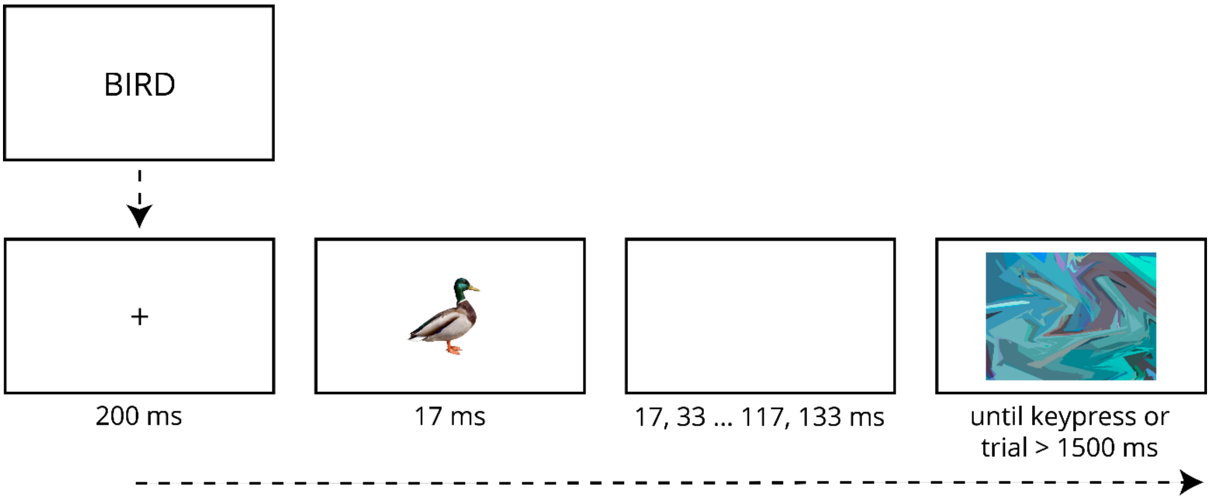
Blocked design. We divided 320 target images into 8 blocks according to their basic-level category (e.g., 40 images of dogs constituted the *dog* block, and 40 distractor images specially selected for *dog*). These basic-level blocks were *dog, bird, flower, tree, car, boat, box* and *cup*. We then paired these 8 basic-level blocks to form 4 superordinate categories (*animal*, *plant*, *vehicle* and *container*), and split each category in half to form 8 superordinate-level blocks that were the same size as the basic-level blocks (i.e., 40 target items). There were therefore two different blocks for each superordinate category, featuring the same object concepts via different images (e.g., both *animal* blocks comprised 20 images of *dogs* and 20 images of *birds*, but different sets of images in each block.

We used eight different durations for the stimulus onset asynchrony between picture and mask (SOA, see procedure). Within each block, we varied SOA from trial to trial, so that each SOA was featured equally across all target images (i.e., 5 trials at 8 SOA per block), and presented in a random order. To ensure that all block-image combinations (e.g., *animal* – *Labrador* [image 1]) would be seen paired all SOA throughout the experiment, we created 16 lists of 12 blocks (i.e., all 16 lists contained all 800 items divided over all eight basic-level blocks, each half of the lists contained a different version of the superordinate-level blocks), which we counterbalanced so that across all lists, every target image appeared paired with every SOA three times, twice in a basic-level block and once in a superordinate-level block.

Each participant only saw 12 blocks (i.e., all basic blocks and one version of each superordinate block). Every participant saw every target image, with some images repeated across blocks (e.g., they saw *Labrador* [image 1] once in the *dog* and once in the *animal* block). Finally, every participant was randomly assigned to one of the 16 lists, so that across every 8 participants, every image was seen at least once at every SOA following a basic-level label (e.g., in the *dog* block, participant one saw *Labrador* [image 1] followed by a 17ms SOA, whereas participant two saw the same image followed by a 33ms SOA), and across every 16 lists, all images were seen at least once at every SOA following a superordinate-level label.

Figure 2.

Trial diagram, showing the variable SOA (blank screen) between the rapidly-displayed target image and the perceptual mask. The block label (e.g., bird) was only displayed once per block.



Procedure. We used online testing platform Gorilla.sc (Anwyl-Irvine et al., 2020) to run the experiment. Each block of trials started with a brief reiteration of the task instructions, in which we informed participants that they would see a category label followed by a series of rapidly displayed photographs of objects that might be members of that category. We instructed them to press ‘z’ on their keyboard if they believed the object was a category member, or ‘m’ if they believed that it was not. These instructions were followed by a fixation cross for 250 ms and the block label for 3000 ms, and then by all trials. Each trial began with a fixation cross (250ms), directly followed by the stimuli, centred on a white background for 17 ms, followed by a blank screen with a variable duration (SOA) and a mask which was presented until a participant made a decision or the until the trial duration excluding fixation cross exceeded 1500 ms (see trial diagram in Figure 2).

Crucially, we systematically varied the onset timing of a perceptual mask (stimulus onset asynchrony: SOA) following an object photograph (i.e., the duration of the blank screen between photograph and mask) to manipulate the amount of time available for perceptual processing. Because our study relied on web-based testing, we selected SOA based on the assumption that most people would have their monitor set to 60Hz, a common refresh rate on LCD screens. As a result, we selected eight SOA corresponding to screen refreshes at 60Hz, starting at 17 ms (the approximate duration of one screen refresh at 60Hz) and incrementally increasing up to 133 ms in steps of 17 ms (17, 33, 50, 67, 83, 100, 117, 133).

We measured accuracy, as well as response times from the onset of the mask to the onset of a valid keypress. Target and distractor items appeared in random order, with a self-paced break after each block of 80 trials. The order in which blocks were presented was automatically randomised within Gorilla. On average, testing took 35 minutes, including participant information and consent, screening, practice and debriefing.

Critical predictors. In addition to the taxonomic level of the block category label, within each block, each photograph was assigned to a specific mask SOA condition. Furthermore, each photograph had a corresponding value in two critical predictors which represent the overlap in sensorimotor experience and linguistic distributional knowledge between the category name and its potential members.

Helmert-coded SOA. Previous research showed a performance curve that rapidly improved and then levelled off with increasing SOA (Bacon-Macé et al., 2005; Mack & Palmeri, 2015). Therefore, we used Helmert-coding of SOA (i.e., comparing a given SOA to the mean of subsequent, longer SOAs) in order to determine the point where overall performance levelled off and, via interactions, the points where effects of linguistic and/or sensorimotor distance first appeared. This resulted in seven Helmert-coded SOA (e.g., SOA 1 = 17ms vs. the average of all subsequent SOA, SOA 2 = 33ms vs. the average of all subsequent SOA etc.). Note that Helmert coding was chosen over simple adjacent comparisons because it is better suited to capture non-linear monotonic relationships such as the rise-and-plateau relationship we expected here.

Linguistic distance between block label and photograph name(s). To compare the degree to which category and member concepts overlapped in linguistic distributional knowledge, we used a large subtitle corpus (e.g., van Heuven et al., 2014) which consists of 200 million words in British English, to calculate log co-occurrences frequencies for the block label and the names given to an image in a context radius of five. Each word in the corpus was represented as a vector of log co-occurrence frequencies, allowing us to compare two words by calculating the cosine distance between their vectors (i.e., $1 - \cos(\theta(u,v))$). For example, the words *dog* and *animal* generally appear in relatively similar contexts across language. As a result, the distance between their vectors in linguistic space is smaller (cosine distance = .23) than between words that appear in different contexts, such as *jet ski* and *animal* (cosine distance = .82).

In a previous study (van Hoef et al., 2022a – Chapter 2, Experiment 1), we calculated cosine distances between a category label (e.g., *dog*) and the basic-level name of the picture. However, while research has shown that images of objects are indeed most frequently named at the basic level (Rosch, Mervis, et al., 1976), evidence from picture-naming norms suggest that images tend to be named by a variety of labels (e.g., Snodgrass & Vanderwart, 1980). In van Hoef et al. (2022b – Chapter 4), we found that measures of word frequency and length that are weighted by production frequency are better at predicting object recognition times than measures based on the most common object. As a result, we incorporated all non-idiosyncratic names for each photograph from van Hoef et al., weighted by production frequency, into the calculation of linguistic distance. For each photograph of each object concept within a given block (e.g., the photograph of a *Labrador* in the block *dog*) we calculated cosine distances between the co-occurrence vectors of the category label (e.g., *dog*), and that of all non-idiosyncratic names for that particular photograph (e.g., *Labrador*, *dog, animal*). We then calculated the weighted average by production frequency. For example, if the naming proportions for the *Labrador* image were *70% dog*, *20% Labrador* and 10% *animal*, the weighted linguistic distance associated with that image in the *dog* block would comprise: distance *dog* → dog*0.7 + distance *dog* → *Labrador*0.2* + distance *dog* → *animal* *0.1. The final measure of weighted linguistic distance for each block label → photograph label pair theoretically ranged from -1 to +1 (actual range = [0, .77], M = .26, SD = .14). In all models, weighted linguistic distance was centred.

Sensorimotor distance between block label and photograph name. To compare how category and member concepts overlapped in terms of sensorimotor experience, we followed our previous approach (van Hoef et al., 2022a – Chapter 2) in calculating sensorimotor distance based on multidimensional ratings of sensorimotor strength. For both the category and all photograph names, we retrieved 11-dimension rating vectors from Lynott et al.'s (2020) sensorimotor norms for 40,000 concepts, which express the degree to which people experience a given concept via six perceptual modalities (auditory, gustatory, haptic, interoceptive, olfactory and visual), and by performing actions with five action effectors (foot, hand, head, mouth and torso). We compared vectors for category names and photograph names for their potential members by calculating their weighted average cosine distances, based on the assumption that participants will retrieve a different conceptual representation of an object depending on how they have identified the object in the picture. That is, participants who identify a given object as *Labrador* will have a slightly different representation to those that identify it merely as a *dog*, and so on. The calculation of sensorimotor distance therefore followed the same approach as for linguistic distance, incorporating all possible identifications of an object weighted by their production frequency in the van Hoef et al. (2022b - See Chapter 4) picture naming norms. The final measure weighted sensorimotor distance measure for each block label → photograph pair ranged in theory from -1 to +1 (actual range = [0, .15], M = .03, SD = .03). In all models, weighted sensorimotor distance was centred. We note here that weighted sensorimotor and linguistic distance were moderately correlated (Pearson's r = .57, $BF_{10} = 9.55 \times 10^{53}$).

Relationships between dependent variables

Bayesian correlation analyses (JASP, v.0.14.0.0, with default beta prior width =1) of the full dataset revealed evidence ($BF_{10} = 3.20$) for a weak negative ($r = -.230$) relationship between participant accuracy and average RT. This suggested that overall, accuracy went down as RT increased. A by SOA analysis revealed that evidence for a correlation was strongest at SOA of 17, 33 and 67 ms, but disappeared at longer SOA (see table 1).

Table 1.

Bayesian correlations between participant accuracy and average RT

SOA	Pearson's *r*	BF_{10}
Overall	-.230	3.202
17	-.239	4.286
33	-.259	8.286
50	-.190	1.073
67	-.268	11.239
83	-.130	0.317
100	-.212	1.901
117	-.166	0.620
133	-.197	1.274

Data analysis. We pre-registered three sets of analyses. In analysis A, we investigated whether mask SOA affected categorisation performance (response time, accuracy), and tested the hypothesis (1) that performance would improve and then level off as SOA increased. That is, accuracy would increase, and RT would decrease, and then level off. To this end, we ran a series of hierarchical mixed effects regression models of RT (linear models, on correct trials) and accuracy (logistic models), with crossed random effects of participant and nested concept and photograph, and fixed effects of Helmert-coded SOA. In addition to this, we ran a series of mixed effects logistic regression model of accuracy (incorrect = 0, correct = 1), with crossed random effects of participant and nested concept and image and fixed effects of

Helmert-coded SOA. For both RT and accuracy analyses, step 0 (null model) entered the random effects. Step 1 entered SOA 1 (i.e., 17 ms vs. the mean of 33 to 133 ms), step 2 entered SOA 2 (i.e., 33 ms vs. the mean of 50-133 ms) and so on up to step 7, which entered SOA 7 (i.e., 117 vs the mean of 133 ms). We used Bayesian model comparisons (calculating Bayes Factors based on BIC; Wagenmakers, 2007) between each step and the previous one to determine at which step performance no longer improved with longer SOA. Furthermore, we examined coefficients for each Helmert-coded SOA to ascertain their direction was as expected. Specifically, in the analysis of RT, we expected that coefficients would be positive, reflecting longer RT at each mask SOA compared to subsequent ones. By contrast, in the analysis of accuracy, we expected that coefficients would be negative, reflecting worse accuracy at each mask SOA compared to subsequent ones.

In analysis B, we investigated whether sensorimotor and linguistic distance affected RT and accuracy, and whether they interacted with SOA. To test our hypothesis (2) that RT increases and accuracy decreases as weighted average sensorimotor and linguistic distance between the category label and image name(s) increase, we ran a set of hierarchical mixed effects regressions as per analysis A. Step 0 (null model) entered random effects of participant and concept/photograph. Step 1 entered all Helmert-coded SOA; step 2 entered weighted sensorimotor distance; and step 3 entered weighted linguistic distance. We used Bayesian model comparisons between steps 1, 2, and 3 to determine whether sensorimotor and/or linguistic distance affected categorisation performance, and inspected model coefficients to ascertain that the effects had the hypothesised direction (greater distance = longer RT, lower accuracy).

To test our hypothesis (3) that the effect of sensorimotor and linguistic distance on accuracy and RT increases with SOA, we tested for interactions at each SOA to determine the point at which effect size started to change. Step 4 tested candidate models for interactions with SOA1: Step 4a comprised the interaction between SOA 1 and sensorimotor distance; Step 4b comprised the interaction between SOA1 and linguistic distance; and Step 4c comprised both interactions together. We used Bayesian model comparisons to select the best-fitting model between the candidate interactions and the preceding model (i.e., step 3), and carried it forward to the subsequent step. Next, Steps 5a-c examined candidate interactions with SOA2, Steps 6a-c examined candidate interactions with SOA3, and so on until Step 10 (SOA7), carrying forward the best-fitting model at each step. In the event that the best model was only equivocally better (i.e., BF < 3) than the next-best, we carried forward the model with fewer parameters. Where we observed an interaction effect, we examined whether the direction of the coefficients was as expected.

Finally, in analysis C, we tested the hypothesis (4) that linguistic distance would require longer SOA to take effect than sensorimotor distance, and investigated at which mask SOA the respective effects of sensorimotor and linguistic distributional information first appeared. To this end, we split the dataset by all eight SOA durations (e.g., 17, 33, 50, 67, 83, 100, 117 and 133 ms), and ran hierarchical mixed effects regressions of RT (linear models) and accuracy (logistic models). For both RT and accuracy at each SOA duration, step 0 entered random effects of participants and concept/image; step 1 entered a fixed effect of sensorimotor distance; and step 3 entered a fixed effect of linguistic distance.[11]

5.5. Results

For the analysis of accuracy, we removed 625 (1.02%) out of 61444 correct and incorrect responses to target trials, for having RT below 200ms (motor error). The resulting dataset contained 60819 responses. For the analysis of RT, we selected 49061 correct target trials, and removed 412 trials (0.84%) for having RT below 200 ms. We replaced no

[11] To address convergence issues in models including sensorimotor and linguistic distance, we explored a model with an alternative linguistic distance measure: PPMI-ngram, which did not meaningfully improve performance. Full analyses and results are available on OSF.

participants for having an average RT further than 3 SD away from the overall mean. Finally, we removed as outliers 1298 trials for having an RT further than 2.5 SD away from the participant's mean RT (2.67%), bringing the final dataset used for the analysis of RT to 47351 responses. Full dataset and code available at https://osf.io/vdka2/?view_only=ac6c78793a354a1f95da6c36b5ac6163.

Analysis A: SOA. In RT, Bayesian model comparisons for analysis A showed that every additional Helmert-coded SOA improved model fit over the previous one (see Table 1 for model comparisons). The best-fitting model included all seven Helmert-coded SOA (BF_{10} compared to next-best model = 9.49×10^{10}), and showed that RTs were faster at every mask SOA compared to the average of all subsequent SOA. These results show that, as predicted, RT became faster as SOA increased. However, against our predictions, performance did not level off: rather, participants responded more quickly at each mask SOA from 17 to 133 ms (see Figure 3).

In the analysis of accuracy, Bayesian model comparisons provided moderate evidence for steps one and two over their predecessors (i.e., favouring the inclusion of Helmert-coded SOA 1 and 2), meaning that accuracy was worse at a mask SOA of 17 [unstandardized β = -.32, 95% CI = ±0.06, z = -9.96, p < .001] and 33 ms [β = -.20, 95% CI = ±0.06, z = -5.94, p < .001] compared to all longer SOA. No subsequent step (3 to 7) was better than its predecessor, indicating consistent accuracy performance across SOA 50, 83, 100, 117 and 133 ms (see Figure 2). Overall, following predictions, accuracy improved steadily as SOA increased from 17 to 50 ms, and started to level off from that point onwards.

Figure 3.

Coefficient plots for analysis A step 7 models of accuracy and RT, displaying estimated change in log odds of answering correct and RT in ms per Helmert-coded SOA (with error bars showing ±1SE).

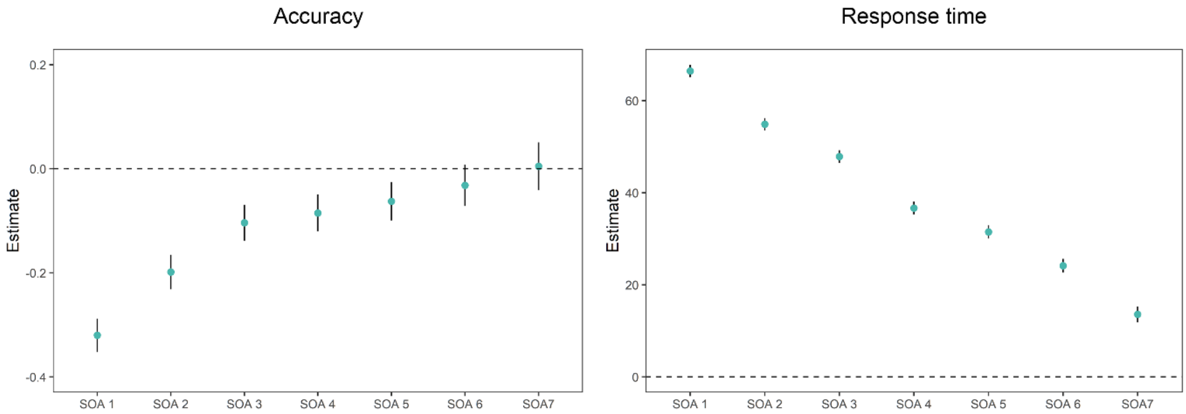


Table 2.

Analysis A: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for the comparison of each step against the preceding one.

Step	Model comparison	RT		Accuracy	
		$\%\Delta R^2$	BF_{10}	$\%\Delta R^2$	BF_{10}
0	Null model (random effects)	32.610	-	27.770	-
1	SOA1 vs. null	2.992	6.38×10^{463}	.242	1.64×10^{18}
2	SOA2 vs. step 1	2.060	2.56×10^{323}	.093	115266.26
3	SOA3 vs. step 2	1.547	1.36×10^{257}	.025	0.33*
4	SOA4 vs. step 3	.880	4.65×10^{148}	.016	0.07
5	SOA5 vs. step 4	.610	2.28×10^{103}	.008	0.01
6	SOA6 vs. step 5	.316	3.28×10^{53}	.002	<0.01
7	SOA7 vs. step 6	.076	9.49×10^{10}	<.001	<0.01

Note: The data was BF_{01} = 3.02 times more likely under the step 2 model than under the step 3 model, which narrowly passes the Bayes Factor threshold for inferencing (BF_{10} > 3).

Analysis B: Sensorimotor and linguistic distributional information. In steps 1 – 3 of analysis B, we tested the hypothesis that categorisation performance (RT and accuracy) would be affected by the weighted average sensorimotor and/or linguistic distance between the block label (e.g., DOG) and the names given to a particular image (e.g., *Labrador* [image 1]), see Table 2 for all model comparisons.

In RT, Bayesian model comparisons between step 1 (which entered all mask SOA) and step 2 (which entered sensorimotor distance) revealed evidence against the effect of sensorimotor distance over that of the time available for perceptual processing, showing that the data was BF_{01}=6.38 times more likely under the step 1 model. Furthermore, model comparisons between step 2 and 3(which entered linguistic distance) revealed strong evidence against the effect of linguistic over sensorimotor distance, showing that the data was

BF_{01}= 12.86 times more likely under the step 2 compared to the step 3 model. See Table 2 for all model comparisons.

In accuracy, Bayesian model comparisons between step 1 and step 2 revealed very strong evidence for the effect of sensorimotor distance over that of the time available for perceptual processing, showing that the data was $BF_{10}=1.35\text{x}10^{55}$ times more likely under the step 2 model. Furthermore, model comparisons between 2 and 3 once again revealed strong evidence for the effect of linguistic distance over the time available for perceptual processing and sensorimotor distance, showing that the data was $BF_{10}= 5.86\text{x}10^{17}$ times more likely under the step 3 compared to the step 2 model. According to the step 3 model, participants were up to 1.70 times[12] more likely to respond incorrectly to trials when the average centred sensorimotor distance between the block label and the image names was high compared to average [$\beta = -4.43$, 95% CI = ±1.74, $z = -5.024$, $p < .001$], and up to 2.66 times[11] more likely to answer incorrectly to trials when the average centred linguistic distance was high compared to average [$\beta = -1.26$, 95% CI = ±.32, $z = -7.74$, $p < .001$].

These findings partially confirmed our hypotheses that sensorimotor and linguistic distance would affect categorisation performance, showing that accuracy decreased as the average sensorimotor and linguistic distance between the block label and the image names increased. However, in contrast to our hypothesis, neither sensorimotor nor linguistic distance affected RT above and beyond SOA.

Interactions between sensorimotor and linguistic distance and SOA. In steps 4 through 10 of analysis B, we tested the hypothesis that the effect of sensorimotor and linguistic distance on accuracy and RT would increase with SOA.

In RT, Bayesian model comparisons (see Table 5 for all model comparisons) failed to favour any of the interactions with SOA in steps 4(a-c) through 10(a-c), with one exception:

[12] Calculated with the largest value for centred sensorimotor (.12) and linguistic distance (0.54).

there was moderate evidence ($BF_{10} = 9.33$) for including the interaction at Step 6b, between linguistic distance and SOA3 (i.e., 50 vs. 67-133 ms). That is, the effect size of linguistic distance on RT shifted in size as SOA increased, from no effect between SOA 67-133 ms. Nonetheless, the fit of the Step 6b interaction model was still worse than the original SOA-only model, suggesting that overall neither sensorimotor nor linguistic distance affected RT more than the time available for perceptual processing.

In accuracy, Bayesian model comparisons did not favour any of the interactions with SOA. The best model was always the step 3 model, with linear effects of sensorimotor and linguistic distance, but without any of the candidate interactions of SOA with sensorimotor and/ or linguistic distance.

Overall, these results were partially consistent with our hypotheses. Larger sensorimotor distance did result in lower accuracy, but we found no evidence that this effect changed with increasing SOA. Larger sensorimotor distance did not affect response latencies, and we found no evidence that this effect changed with increasing SOA. Larger linguistic distance also resulted in lower accuracy, but we found no change in effect size with increasing SOA. In RT, we found some evidence that the linguistic distance effect changed with SOA, but the data overall still favoured the SOA-only model more strongly. As a consequence, we interpreted the effect of linguistic distance on RT to be potentially quite weak. We examine the linguistic effect in detail per SOA in analysis C (see below).

Table 3.

Analysis B, model comparisons showing change in R-squared and Bayes Factors for the comparison for each step and the previous one (steps 1, 3) and for all interaction models compared to the non-interaction model (Step 3 vs. 4,5,6,7,9,10a-c).

Step	Model comparison	RT		Accuracy	
		%ΔR^2	BF_{10}	%ΔR^2	BF_{10}
0	Null model (random effects)	32.61	-	27.770	-
1	SOA (1-7) vs. null	8.481	7.47×10^{1369}	1.269	1.58×10^{15}
2	Sensorimotor distance vs. step 1	.002	<0.01	.317	1.00×10^{44}
3	Sensorimotor distance + Linguistic distance vs. step 2	.040	0.08	.049	4029950249.00
4a	Sensorimotor distance*SOA1 vs. step 3	<.001	<0.01	.077	0.22
4b	Linguistic distance*SOA1 vs. step 3	.007	0.09	.084	1.63
4c	Sensorimotor distance*SOA1 + Linguistic distance vs. step 3	.011	<0.01	.013	0.01
5a	Sensorimotor distance*SOA2 vs. step 3	<.001	<0.01	.025	0.01
5b	Linguistic distance*SOA2 vs. step 3	.003	0.02	.027	0.02
5c	Sensorimotor distance*SOA2 + Linguistic distance vs. step 3	.003	<0.01	<.001	<0.01
6a	Sensorimotor distance*SOA3 vs. step 3	<.001	<0.01	.013	<0.01
6b	Linguistic distance*SOA3 vs. step 3	.018	9.33	.013	0.01
6c	Sensorimotor distance*SOA3 vs. step 3	.006	0.49	<.001	<0.01
7a	Sensorimotor distance*SOA4 vs. step 3	<.001	0.05	<.001	<0.01
7b	Linguistic distance*SOA4 vs. step 3	<.001	0.05	<.001	<0.01
7c	Sensorimotor distance*SOA4 + Linguistic distance vs. step 3	.001	<0.01	<.001	<0.01
8a	Sensorimotor distance*SOA5 vs. step 3	<.001	0.04	<.001	<0.01
8b	Linguistic distance*SOA5 vs. step 3	.006	<0.01	<.001	<0.01
8c	Sensorimotor distance*SOA5 + Linguistic distance vs. step 3	<.001	<0.01	<.001	<0.01

9a	Sensorimotor distance*SOA6 vs. step 3	.005	0.29	<.001	<0.01
9b	Linguistic distance*SOA6 vs. step 3	.008	1.28	.003	0.01
9c	Sensorimotor distance*SOA6 + Linguistic distance vs. step 3	.009	<0.01	.003	<0.01
10a	Sensorimotor distance*SOA7vs. step 3	<.001	<0.01	.001	<0.01
10b	Linguistic distance*SOA7 vs. step 3	<.001	<0.01	.001	<0.01
10c	Sensorimotor distance*SOA7 + Linguistic distance vs. step 3	<.001	<0.01	.002	<0.01

Note: best-fitting interaction model printed in bold, interaction carried forward to all subsequent steps

Analysis C: Sensorimotor and linguistic effects per SOA. In analysis C, we examined at what SOA the effects of sensorimotor and linguistic distance first appeared, and hypothesised that linguistic distance would require longer SOA to affect categorisation performance than sensorimotor distance. See Table 3 for all model comparisons.

In RT, Bayesian model comparisons between step 1 (which entered sensorimotor distance) and 0 (which entered random effects) at every SOA showed strong evidence against the effect of sensorimotor distance over random effects. By contrast, model comparisons between steps 2 (which entered linguistic distance) and 1 showed moderate to strong evidence for the effect of linguistic distance over sensorimotor distance at mask SOA of 83 and 100 ms. At a mask SOA of 83 ms, participants were up to 32.99 ms[1] slower to respond for words with high compared to average linguistic distance ($\beta = 61.09$, 95% CI = ±32.30, $t(1943.50) = -3.71$, $p < .001$) At a mask SOA of 100 ms, participants were up to 35.26 ms[1] slower to respond for words with high compared to average linguistic distance ($\beta = 65.29$, 95% CI = ±31.91, $t(1842.58) = 4.01$, $p < .001$). At all other SOA, model comparisons showed evidence against the added effect of linguistic distance.

In accuracy meanwhile, Bayesian model comparisons between step 1 and 0 showed strong evidence for the effect of sensorimotor distance over random effects at all SOA. Furthermore, model comparisons between step 2 and 1 at each SOA showed mixed evidence for the effect of linguistic distance over sensorimotor distance. At a SOA of 17ms, we found moderate evidence that linguistic distance predicted accuracy above and beyond sensorimotor distance, however at SOA of 33 and 50ms, evidence was in the equivocal zone. At SOA of 67, 83, 100 and 177 ms, linguistic distance again predicted accuracy above and beyond sensorimotor distance, but at the longest SOA (133 ms), evidence was once again in the equivocal zone. Inspection of the coefficients revealed that at where the effect appeared, greater linguistic distance decreased the likelihood of a correct answer (see Figure 4). Greater sensorimotor distance also negatively affected accuracy, but not significantly so in any of the step 2 models for all SOA.

In summary, contrary to our expectations, linguistic distance did not require longer SOA to take effect than sensorimotor distance. As observed in analysis B, there was a relatively late effect of linguistic distributional information at 83 to a 100 ms, but no effect of sensorimotor distance at any SOA. In accuracy, analyses showed a mixed effect of sensorimotor and linguistic distance, with linguistic distance predicting accuracy above and beyond sensorimotor distance alone, although not consistently so. While sensorimotor distance affected accuracy at all SOA, linguistic distance had a weaker, inconsistent effect that appeared intermittently from 17 ms onwards.

Figure 4.

Predicted response times and probability of a correct answers derived from the step 2 models at each SOA, for low, average and high sensorimotor and linguistic distance (corresponding to the highest, average and lowest values for weighted centred distance measures). With error bars showing ±1SE).

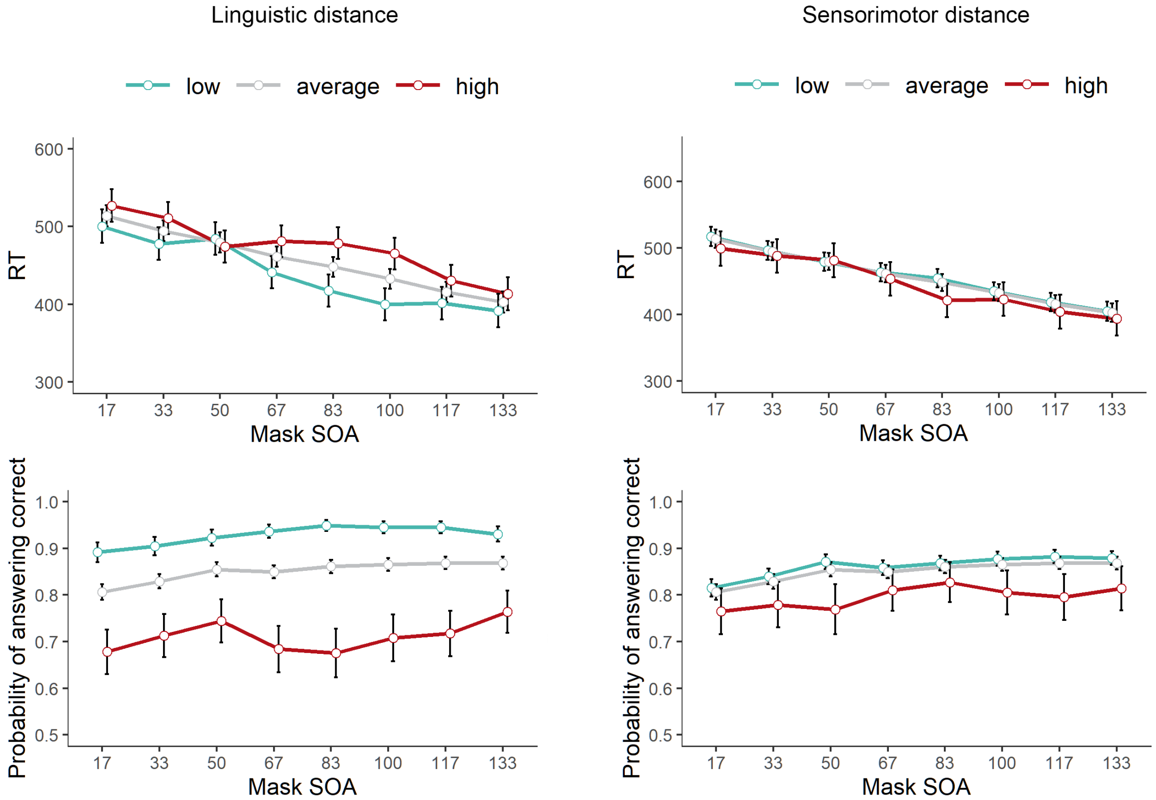


Table 4.

Analysis C model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for the comparison of each step against the preceding one.

Mask SOA (ms)	Step	Model comparison	RT		Accuracy	
			%ΔR^2	BF_{10}	%ΔR^2	BF_{10}
17	0	Null model (random effects)	39.308	-	26.112	-
	1	Sensorimotor distance vs. step 0	.003	0.01	.750	490.05
	2	Linguistic distance vs. step 1	.064	0.05	1.236	5.30
33	0	Null model (random effects)	38.834	-	26.69	-
	1	Sensorimotor distance vs. step 0	.022	0.02	.850	823.03
	2	Linguistic distance vs. step 1	.091	0.10	1.331	2.77
50	0	Null model (random effects)	36.551	-	32.412	-
	1	Sensorimotor distance vs. step 0	.002	0.01	1.370	425066.11
	2	Linguistic distance vs. step 1	.009	0.02	.431	2.87
67	0	Null model (random effects)	36.370	-	27.099	-
	1	Sensorimotor distance vs. step 0	.030	0.03	1.280	124616.96
	2	Linguistic distance vs. step 1	.135	0.20	1.018	430.74
83	0	Null model (random effects)	36.959	-	30.355	-
	1	Sensorimotor distance vs. step 0	<.001	0.01	1.437	541176.09
	2	Linguistic distance vs. step 1	.350	10.59	1.239	4094.91
100	0	Null model (random effects)	35.139	-	30.602	-
	1	Sensorimotor distance vs. step 0	.129	0.27	1.649	4694949.29
	2	Linguistic distance vs. step 1	.358	28.65	.982	254.68
117	0	Null model (random effects)	35.104	-	29.400	-
	1	Sensorimotor distance vs. step 0	<.001	0.01	1.887	122362682.50
	2	Linguistic distance vs. step 1	.075	0.05	.868	194.22
133	0	Null model (random effects)	38.208	-	30.255	-
	1	Sensorimotor distance vs. step 0	.001	0.01	1.058	1717.28
	2	Linguistic distance vs. step 1	.041	0.03	.488	1.78

Exploratory analyses. We pre-registered two exploratory analyses. Firstly, we investigated the effect of taxonomic level on categorisation performance at all mask SOA. Second, we intended to explore two alternative measures of accuracy, *d'* and *c*, in a signal detection paradigm. However, the signal detection analysis proved unworkable in the current design, and we only report the results of the analysis of taxonomic level here.

In addition to the pre-registered exploratory analyses, we ran an additional set of analyses on the effects of psycholinguistic characteristics associated with the image and category labels on accuracy and RT.

Taxonomic level. Although it was not the main focus of the present study, previous work (Macé et al., 2009; Mack & Palmeri, 2011, 2015) suggests that the classic performance advantage for categorisation at the basic (e.g., *dog*) compared to the superordinate level (e.g., *animal*) is reduced or even reversed when participants have limited time to recognise an object. Consequently, we pre-registered and carried out an additional hierarchical regression analysis of accuracy and RT for a dataset split by mask SOA (i.e., as per analysis C). For both RT and accuracy, at each SOA duration, step 0 entered crossed random effects of participant and nested concept/image, and step 1 entered taxonomic level (0 = basic, 1 = superordinate). We used Bayesian model comparisons between step 1 and 0 at each SOA duration to see whether a model including taxonomic level outperformed a null model including only random effects of participant and concept/photograph. Furthermore, we inspected the taxonomic coefficient to determine the direction of any effect (i.e., better performance at the basic or superordinate level). Finally, at every SOA duration, we compared the taxonomic step 1 model to the sensorimotor-linguistic model from analysis C, step 2 (see above), in order to examine whether the data were better explained by a taxonomic or sensorimotor-linguistic model.

In RT, Bayesian model comparisons (see Table 4 for full model comparisons), showed that evidence was against the inclusion of taxonomic level at all SOA except 50 ms (BF_{10}= 33.95). Inspection of the coefficients revealed that at 50 ms, superordinate-level categorisation was slightly faster than basic-level categorisation (β = -10.46, 95% CI = ±5.16, $t(5569.12) = -3.97$, $p < .001$). The overall pattern in the data suggests that there was no clear basic-level advantage in RT: categorisation was equally fast at the basic compared to the superordinate level. Comparison to the step 2 model from analysis C revealed that the taxonomic model outperformed the sensorimotor-linguistic distance at all SOA except 83 and 100 ms. Nonetheless, it is important to stress that few of these models were favoured over the null, and as such the best-fitting model was highly variable across SOA. That is, neither taxonomic nor sensorimotor-linguistic distance predicted RT at a SOA of 17 and 33 ms, taxonomic level was the better predictor at 50 ms, neither predicted RT at 67 ms, sensorimotor-linguistic distance was the best model at 83 and a 100 ms, and neither model predicted RT at a SOA of 133 ms.

In accuracy, Bayesian model comparisons between step 0 and 1 revealed strong evidence for the effect of taxonomic level at all SOA except at 33 and 133 ms (see Table 4 for all model comparisons). Inspection of the coefficients showed that where an effect was present, categorisation was more accurate at the basic compared to the superordinate level (i.e., the classic basic-level advantage). Here it is important to stress that sensorimotor-linguistic models strongly outperformed the taxonomic model and the null-model at every SOA. Combined with the results from analysis C, this means that the best-fitting model always included sensorimotor and/ or linguistic distance. A sensorimotor-linguistic model predicted accuracy best at SOA of 17, 67, 83, 100 and 117 ms, and a sensorimotor-only and sensorimotor-linguistic model predicted accuracy about equally well at 33, 50 and 133 ms.

Table 5:

Exploratory taxonomic analyses: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for Taxonomic levels vs. null, and sensorimotor-linguistic vs. taxonomic levels.

Mask SOA	Step	Model comparison	RT		Accuracy	
			$\%\Delta R^2$	BF_{10}	$\%\Delta R^2$	BF_{10}
17	0	Null model (random effects)	39.308	-	26.112	-
	1	Taxonomic level vs. step 0	.024	0.04	.343	99.51
	-	Step 2 (analysis C) vs. step 1	-	0.02	-	26.11
33	0	Null model (random effects)	38.834	-	26.693	-
	1	Taxonomic level vs. step 0	.008	0.02	.078	0.08
	-	Step 2 (analysis C) vs. step 1	-	0.11	-	29100.19
50	0	Null model (random effects)	36.551	-	32.412	-
	1	Taxonomic level vs. step 0	.173	33.95	.414	334.411
	-	Step 2 (analysis C) vs. step 1	-	<0.01	-	3651.89
67	0	Null model (random effects)	36.370	-	27.099	-
	1	Taxonomic level vs. step 0	.029	0.05	.359	50.86
	-	Step 2 (analysis C) vs. step 1	-	0.11	-	1055473.00
83	0	Null model (random effects)	36.959	-	30.355	-
	1	Taxonomic level vs. step 0	.006	0.02	.628	32032.29
	-	Step 2 (analysis C) vs. step 1	-	8.00	-	69182.28
100	0	Null model (random effects)	35.139	-	30.603	-
	1	Taxonomic level vs. step 0	.010	0.02	.643	30576.90
	-	Step 2 (analysis C) vs. step 1	-	387.61	-	39104.78
117	0	Null model (random effects)	35.104	-	29.400	-
	1	Taxonomic level vs. step 0	.36	0.22	.689	397819.60
	-	Step 2 (analysis C) vs. step 1	-	<0.01	-	333534.00
133	0	Null model (random effects)	38.208	-	30.255	-

1	Taxonomic level vs. step 0	.064	.27	.125	0.61
-	Step 2 (analysis C) vs. step 1	-	<0.01	-	108.64

In summary, taxonomic level did not systematically affect categorisation speed, showing neither the classic basic-level advantage nor the clear superordinate-level advantage previously observed in studies examining rapid categorisation with a blocked design (e.g., Macé et al., 2009). Furthermore, where a difference appeared, categorisation was less accurate at the superordinate-level compared to the basic-level, consistent with the classic basic-level advantage literature but inconsistent with previous work that suggests that the basic-level advantage may disappear in rapid categorisation with a blocked design (e.g., Mack & Palmeri, 2015). Crucially, these results suggest that neither a taxonomic level nor sensorimotor-linguistic model predicted RT systematically better across all SOA. However, in accuracy we found that sensorimotor-linguistic models were consistently better at predicting categorisation accuracy at every SOA.

Psycholinguistic variables. We explored the effects of psycholinguistic characteristics of subjective age of acquisition rating (AoA, retrieved from: Kuperman et al., 2012) Zipf log word frequency (retrieved from: van Heuven et al., 2014) subjective familiarity rating (retrieved from Scott et al., 2019; Stadthagen-Gonzales & Davis, 2006 and the MRC Psycholinguistic Database), and word length of the category block label, as well as the weighted average AoA rating, Zipf log word frequency, familiarity rating and word length per photograph.

For both accuracy and RT, we carried out two exploratory analyses. Analysis D consisted of a set of mixed effects regressions, which we carried out in two stages. Stage one consisted of 9 steps, where each step added a psycholinguistic variable. We used Bayesian model comparisons between each step and the previous one to determine whether the variable included improved model fit, and carried only those variables forward that did. In this stage, step 0 entered random effects (as per analysis A). Next, steps 1 through 4 entered category label AoA, Zipf log word frequency, familiarity and word length. Then, steps 5 through 9 entered weighted average image name AoA, Zipf log word frequency, familiarity and word length. In the second stage, we added all SOA to the best-fitting model from stage one, following the same procedure as outlined in confirmatory analysis A. Here, we again used Bayesian model comparisons between each step and the previous one to determine at which step performance no longer improved. We were particularly interested to explore whether controlling for psycholinguistic characteristics changed the SOA-patterns we observed in our confirmatory analysis A.

Our exploratory analysis E of RT and accuracy consisted of a set of hierarchical mixed effects regressions as per exploratory analysis A. Step 0 (null model) consisted of the best-performing psycholinguistic model from exploratory analysis D. Step 1 entered all Helmert-coded SOA; step 2 entered weighted sensorimotor distance and step 3 entered weighted linguistic distance. We used Bayesian model comparisons between steps 0, 1, 2, and 3 to determine whether sensorimotor and/or linguistic distributional information affected categorisation performance beyond psycholinguistic characteristics and the time available for perceptual processing (SOA). Finally, steps 4 through 10 (a-c) explored whether interactions between sensorimotor and/or linguistic distributional distance and any of the SOA improve model fit over a non-interaction model (step 3). Again, we were particularly interested to explore whether the inclusion of psycholinguistic characteristics altered the patterns we observed in our confirmatory analysis B.

Analysis D: SOA. In the analysis of RT, Bayesian model comparisons for stage one of analysis D showed that a model containing label AoA, frequency length and familiarity as well as weighted average image name AoA fit the data best (BF_{10} compared to the next-best model = 6.15×10^{141}). Upon further inspection, we found that category label word frequency and length were strongly correlated. However, we did not find evidence of multicollinearity (label word frequency, Tolerance = 2.97, *VIF* = 8.83; label word length, Tolerance = 2.89, *VIF* = 8.38). Since the primary aim of this analysis was to determine whether the previously established effects of SOA, and sensorimotor and linguistic distributional information could be observed in models including psycholinguistic characteristics, we opted to continue with the best-fitting psycholinguistic model. The final psycholinguistic model included category label AoA, frequency, familiarity, word length and weighted average image name AoA (see table 6 for full model comparisons)

We carried this model forward to stage 2, where Bayesian model comparisons revealed a similar pattern to our confirmatory analyses. We found strong evidence that every additional Helmert-coded SOA improved model fit over the previous one (see Table 7 for model comparisons). The best-fitting model included all seven Helmert-coded SOA (BF_{10} compared to next-best model = 1.14×10^{11}), and showed that RTs were faster at every mask SOA compared to the average of all subsequent SOA. Inspection of the coefficients showed that RT became faster as SOA increased, even with the inclusion of relevant psycholinguistic characteristics of the label and image name. Furthermore, we found that participants were slower to categorise images that followed labels with high compared to average AoA ratings [β = 23.97, 95% CI = ±2.42, t(10609.25) = 19.42, p < .001], high compared to average familiarity ratings[β = 27.87, 95% CI = ±6.50, t(30323.68) = 8.40, p < .001], as well as to images as well slightly slower when categorising photographs that were given names with higher compared to average age of acquisition [β = 7.55, 95% CI = ±3.66, t(232.03) = 4.04, p < .001]. Labels with high compared to average word frequency meanwhile resulted in faster response times [β = -20.03 95% CI = ±7.59, t(32997.13) = -5.17, p < .001], as did longer category names [β = -13.35, 95% CI = ±1.50, t(43583.84) = -17.43, p < .001].

In the analysis of accuracy, Bayesian model comparisons in stage one of analysis D showed that a model containing category label AoA, frequency, familiarity, as well as weighted average image name AoA fit the data best (BF_{10} compared to the next-best model = 2.28×10^{53}; see Table 6 and 7 for full model comparisons). We carried this model forward to stage 2, where Bayesian model comparisons revealed a similar pattern to our confirmatory analyses. That is, we found strong evidence for steps one and two over their predecessors (i.e., favouring the inclusion of Helmert-coded SOA 1 and 2) but found no evidence that any other step improved performance over the previous one. In addition to the previously observed lower accuracy at SOA1 [β = -.32, 95% CI = ±0.06, z = -10.03, p < .001], and SOA2 [β = -.20, 95% CI = ±0.06, z = -5.92, p < .001], compared to all subsequent SOA, inspection of the coefficients for the best-fitting model revealed that participants were less accurate at categorising photographs of objects following labels of high compared to average AoA [β = -.28, 95% CI = ±0.06, z = -9.46, p < .001], high compared to average frequency [β = -.62, 95% CI = ±0.11, z = -11.03, p < .001], as well as photographs that were assigned names with higher compared to average age of acquisition [β = -.25, 95% CI = ±0.10, z = -4.72, p < .001] . Finally, accuracy was higher for photographs following labels that had higher than average familiarity ratings [β = .66, 95% CI = ±.16, z = 8.16, p < .001].

Table 6.

Exploratory psycholinguistic analysis D stage one: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for the comparison of each step against the previous best-fitting one.

Step	Model comparison	RT		Accuracy	
		$\%\Delta R^2$	BF_{10}	$\%\Delta R^2$	BF_{10}
0	Null model (random effects)	32.611	-	28.890	-
1	Label_AoA vs. Null	-.232	91.07	.719	8.72×10^{33}
2	Label frequency vs. step 1	.796	1.58×10^{35}	2.258	1.30×10^{32}
3	Label familiarity vs. step 2	.190	1.13×10^{18}	-.278	6.11×10^{11}
4	Label word length vs. step 3	-.377	1.25×10^{53}	-.026	<0.01
5	Weighted AoA vs. step 4*	.314	1.76×10^{139}	2.313	2.27×10^{53}
6	Weighted frequency vs. step 5	-.460	0.98	-.091	0.37
7	Weighted familiarity vs. step 5	-.057	0.51	-.141	0.03
8	Weighted word length vs. step 5	.334	<0.01	-.120	0.08

*In accuracy, step 5 was compared to step 3, as step 4 did not improve model fit over step 3.

Table 7.

Exploratory psycholinguistic analysis D stage two: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for the comparison of each step against the previous one.

Step	Model comparison	RT		Accuracy	
		%ΔR^2	BF_{10}	%ΔR^2	BF_{10}
9	SOA1 vs. psycholinguistic	3.099	$5.96x10^{470}$	.234	$3.03x10^{18}$
10	SOA2 vs. step 9	2.186	$6.88x10^{337}$	.085	103048.50
11	SOA3 vs. step 10	1.634	$3.74x10^{260}$	.024	0.36*
12	SOA4 vs. step 11	.906	$1.09x10^{150}$	.014	0.06
13	SOA5 vs. step 12	.636	$1.11x10^{105}$	.007	0.02
14	SOA6 vs. step 13	.357	$5.09x10^{54}$	.002	<0.01
15	SOA7 vs. step 14	.077	$1.14x10^{11}$	<.001	<0.01

Note: The data was BF_{01} = 2.79 times more likely under the step 10 model than under the step 11 model, which does not exceed the Bayes Factor threshold for inferencing (BF_{10} > 3) and thus constitutes equivocal evidence. Bayes Factors for models in step 9 and 10 of RT are approximate values.

Analysis E: Sensorimotor and linguistic distributional information.

In RT, Bayesian model comparisons between step 1 (which entered the best-performing psycholinguistic model + all SOA) and step 2 (which entered sensorimotor distance) revealed evidence against the effect of sensorimotor distance over that of the time available for perceptual processing, showing that the data was BF_{01}=18.21times more likely under the step 1 model. Furthermore, model comparisons between step 2 and 3(which entered linguistic distance) revealed strong evidence against the effect of linguistic over sensorimotor distance, showing that the data was BF_{01}= 134.85 times more likely under the step 2 compared to the step 3 model. This finding matched confirmatory analyses. See Table 2 for all model comparisons.

In accuracy, Bayesian model comparisons between step 1 and step 2 revealed very strong evidence for the effect of sensorimotor distance over that of psycholinguistic characteristics of the label and image name and time for perceptual processing, showing that the data was $BF_{10}=1.35 \times 10^{55}$ times more likely under the step 2 model. However, contrasting our confirmatory analyses, the inclusion of linguistic distributional information did not improve model fit; we found evidence that the data was *less* likely under the sensorimotor-linguistic step 3 model than under the sensorimotor step 2 model ($BF_{01} = 21.62$). Inspecting the coefficients of the step 2 model showed that, in addition to the previously observed effects of SOA and psycholinguistic characteristics of the label and image name (see analysis D, and full model statistics in supplemental materials on OSF), participants were more likely to respond incorrectly to trials with increasing weighted average sensorimotor distance between the category label and image name(s) [$\beta = -4.43$, 95% CI = ±1.74, $z = -5.024$, $p < .001$].

Interactions between sensorimotor and linguistic distance and SOA.

In RT, Bayesian model comparisons (see Table 8 for all model comparisons) favoured the step 6b model (adding the interaction between an SOA of 50ms and linguistic distributional information over a non-interaction SOA-psycholinguistic-sensorimotor-linguistic model (step 3; $BF_{10} = 9.97$). However, this was the only interaction that improved model fit. Moreover, this model did not outperform a psycholinguistic-SOA model, suggesting that overall neither sensorimotor nor linguistic distance affected RT more as the time available for perceptual processing increased, when controlling for psycholinguistic variables.

In accuracy, Bayesian model comparisons favoured the step 4b model (adding the interaction between the shortest SOA and linguistic distributional information) over a non-interaction SOA-psycholinguistic-sensorimotor-linguistic model (step 3; $BF_{10} = 3.35$). We did not observe this effect in a non-psycholinguistic model of the truncated data. However, the data were overall more likely under the simpler SOA-psycholinguistic-sensorimotor-only model (step 2) than under any of the interaction models.

Table 8.

Exploratory psycholinguistic analysis E, model comparisons showing change in R-squared and Bayes Factors for the comparison for each step and the previous one (steps 1, 3) and for all interaction models compared to the non-interaction model (Step 3 vs. 4,5,6,7,9,10a-c).

Step	Model comparison	RT		Accuracy	
		%ΔR^2	BF_{10}	%ΔR^2	BF_{10}
0	Null model (random effects)	32.61	-	27.770	-
1	SOA (1-7) + psycholinguistic vs. null	11.84	$3.71\text{x}10^{1640}$	5.347	$1.21\text{x}10^{139}$
2	Sensorimotor distance vs. step 1	-.176	0.05	-.240	8387430000.00
3	Sensorimotor distance + Linguistic distance vs. step 2	-.141	0.01	-.274	0.05
4a	Sensorimotor distance*SOA1 vs. step 3	<.001	<0.01	-.225	0.28
4b	Linguistic distance*SOA1 vs. step 3	.010	0.11	-.188	3.35
4c	Sensorimotor distance*SOA1 + Linguistic distance vs. step 3	.005	<0.01	-.182	0.02
5a	Sensorimotor distance*SOA2 vs. step 3	-.013	<0.01	-.260	<0.01
5b	Linguistic distance*SOA2 vs. step 3	.004	0.02	-.244	0.10
5c	Sensorimotor distance*SOA2 + Linguistic distance vs. step 3	<.001	<0.01	-.243	<0.01
6a	Sensorimotor distance*SOA3 vs. step 3	-.007	<0.01	-.274	0.01
6b	Linguistic distance*SOA3 vs. step 3	.017	9.97	-.265	0.02
6c	Sensorimotor distance*SOA3 vs. step 3	.008	0.07	-.265	<0.01
7a	Sensorimotor distance*SOA4 vs. step 3	<.001	<0.01	-.274	0.01
7b	Linguistic distance*SOA4 vs. step 3	<.001	<0.01	-.274	0.01
7c	Sensorimotor distance*SOA4 + Linguistic distance vs. step 3*	<.001	<0.01	-.274	<0.01

8a	Sensorimotor distance*SOA5 vs. step 3	<.001	<0.01	-.275	0.01
8b	Linguistic distance*SOA5 vs. step 3	<.001	<0.01	-.275	0.01
8c	Sensorimotor distance*SOA5 + Linguistic distance vs. step 3*	<.001	<0.01	-.275	<0.01
9a	Sensorimotor distance*SOA6 vs. step 3	.004	0.03	-.273	0.01
9b	Linguistic distance*SOA6 vs. step 3	.004	0.06	-.271	0.03
9c	Sensorimotor distance*SOA6 + Linguistic distance vs. step 3	<.001	<0.01	-.271	<0.01
10a	Sensorimotor distance*SOA7vs. step 3	<.001	<0.01	-.273	0.01
10b	Linguistic distance*SOA7 vs. step 3	<.001	<0.01	-.274	0.01
10c	Sensorimotor distance*SOA7 + Linguistic distance vs. step 3	<.001	<0.01	-.272	<0.01

Overall, these results nuance our confirmatory findings, showing that the inclusion of psycholinguistic characteristics of the label and image name explain additional variance in RT and accuracy, but that this comes at the cost of the previously observed effect of linguistic distributional information. A potential explanation for these findings might be that missing values in some of the psycholinguistic characteristics meant that the dataset was limited to complete cases only, thus reducing the sample size. However, an exploratory analysis of the truncated dataset with the original non-psycholinguistic SOA, sensorimotor and sensorimotor-linguistic models showed that the data favoured a model including linguistic distributional information over a model including only SOA and sensorimotor information. From this, we might infer that psycholinguistic characteristics account for some of the variance in accuracy previously attributed to linguistic distributional information. This finding mirrors exploratory results from speeded categorisation (van Hoef et al., 2022 – see Chapter 2). In that light, it is worth noting that the current work employed log co-occurrence frequency vectors to calculate linguistic distributional distance between category labels and image names. Such vectors are more likely to correlate with psycholinguistic factors such as word frequency and age of acquisition. Therefore, future research might consider using alternative measures, such as a distance measure based on positive pointwise mutual information (PPMI; Bullinaria & Levy, 2007).

5.6. General discussion

Previous research has shown that effects typically observed in object categorisation (e.g., the basic-level advantage) may disappear when task constraints (e.g., presentation time, time given to respond) change. Here, we investigated whether the effect of sensorimotor and linguistic distributional information we previously observed (van Hoef et al., 2022a – Chapter 2) also changes as a function of task constraints, in particular the time available for perceptual processing.

We hypothesised that categorisation performance would improve, then level off as more time for perceptual processing would become available. Indeed, our analyses showed that accuracy rapidly increased at shorter mask SOA (17 to 50 ms), but began levelling off at mask SOA from 50 ms, where accuracy was no longer significantly different to the average accuracy at longer SOA. By contrast, our analysis of response latencies showed a relatively linear decrease of RT; that is, at each mask SOA, responses were slower than the average of all subsequent SOA. Our findings for accuracy are in line with previous research, which has shown a strong link between the amount of time available for perceptual processing and participants' accuracy in rapid categorisation tasks, but also that accuracy increases more rapidly at shorter compared to longer SOA (Bacon-Macé et al., 2007; Mack & Palmeri, 2011, 2015). Our finding that response time did not level off contrasts some previous work, for example Bacon-Macé et al., 2007, who found that after an initial increase, response times in go/no-go and yes-no categorisation tasks dropped and then stabilised at SOA of around 50 ms. Mack and Palmeri (experiment 3; 2015) reported that response times were significantly shorter with longer SOA, which matches our present findings, but median values reported for each SOA in their study suggest that RT dropped most strongly between the shortest (17 ms) and second-shortest (33) ms SOA. One possible explanation is that the shortest amount of perceptual processing in the present study amounted to roughly 33 ms (i.e., stimulus presentation for 17 ms and blank SOA screen for 17 ms). By contrast, Bacon-Macé et al. used a presentation time of 6ms, with SOA also beginning at 6 ms, bringing the shortest amount of time available for perceptual processing to a mere 12 ms. Similarly, Mack and Palmeri (2015) used a stimulus presentation time of 25 ms, but measured SOA duration from the stimulus onset, such that a SOA of 25 ms meant that the mask appeared immediately after the stimulus, limiting perceptual processing time to 25 ms. As such, it is possible that our manipulation of perceptual processing time, which was linked to the most common screen refresh rate of 60Hz for technical reasons (i.e., online testing), simply was not short enough to observe an initial sharp drop in RT. Nevertheless, our study mostly confirmed our hypothesis that the time available for perceptual processing would affect categorisation performance.

We also found partial confirmation for the hypothesis that sensorimotor and/or linguistic distributional information affect ultra-rapid picture categorisation performance. Our analyses show that increasing the average linguistic distance between the block label and implicit object names resulted in lower accuracy and slower RT. Furthermore, greater average sensorimotor distance between the block label and all implicit object names reduced accuracy but not RT.

While we did not find a consistent effect of sensorimotor information on categorisation performance, we believe the current results are still in line with a sensorimotor-linguistic account of categorisation. The current results are a clear departure from our previous findings using a speeded category verification task (van Hoef et al., 2022a – Chapter 2). That is, where the relative effect of linguistic distributional compared to sensorimotor information was small in speeded verification, it is markedly larger in ultra-rapid categorisation. There are a number of explanations for this. Firstly, it may be the case that, in line with Connell and Lynott (2014), the relative task demands of speeded compared to ultra-rapid categorisation put different emphasis on sensorimotor compared to linguistic distributional information. In speeded object verification, where there is plenty of time to extract perceptual information (e.g., up to several seconds), the match between the sensorimotor representation activated by the label and the perceptual representation activated by the image may have been far more informative than linguistic distributional context. By contrast, in an ultra-rapid paradigm, when perceptual cues are limited, participants may have relied on good-enough linguistic representations. However, this pattern did not occur: sensorimotor information did not have larger effects at longer SOA. It is possible that the mask SOA used in the current study were simply too brief to detect an increase. A future study may address this by adding longer SOA, which may allow us to distinguish more clearly whether more time for perceptual processing activates stronger sensorimotor representations.

Secondly, our blocked design may have dampened the effects of sensorimotor simulation activated by the category label. Mack and Palmeri (2015) found that the superordinate-level advantage in ultra-rapid categorisation disappears when the target category label is presented directly prior to the presentation of the photograph compared, and trials are randomised. In our study, participants judged 80 images for each category label. It is unclear whether the sensorimotor representations activated by the label were sufficiently robust to affect categorisation at the first and the last image equally. However, it is also unclear why sensorimotor information would be disproportionally affected by this compared to linguistic distributional information, which reliably predicted accuracy. A potential follow-up analysis might investigate whether images that were presented earlier on in the block were affected more strongly by sensorimotor information than images that were presented later on, a follow-up experiment might adapt a speeded category-verification design, where labels are presented immediately preceding an image. Incidentally, this might also shed light on the question whether linguistic distributional information was simply more robust to our task design, and make the comparison to data from a speeded verification task stronger.

Our exploratory analyses showed no evidence for a robust effect of the basic-level advantage, which is contrary to what hierarchical accounts would predict (e.g., (Jolicoeur et al., 1984), but our findings were also not in line with previous ultra-rapid categorisation studies with a similar design to our study, which showed a clear superordinate-level advantage in accuracy at shorter SOA (Mack & Palmeri, 2015). By contrast, the finding that categorisation accuracy is better explained by sensorimotor-linguistic information than taxonomic level is consistent with how sensorimotor-linguistic theories conceive of concepts and categorisation. That is, sensorimotor-linguistic accounts argue that it is overlap in sensorimotor and/or linguistic distributional information that determines category membership, and that performance differences in categorisation may stem from the degree of overlap in activation of this information, not from position in a taxonomic hierarchy of concepts.

Our exploratory analyses also revealed that psycholinguistic characteristics of the category label and image names associated with the photograph explained variance in the speed with which participants responded as well as their accuracy. The effects we observed were partially as expected, that is, higher label age of acquisition resulted in longer RT, and lower accuracy. Higher label frequency resulted in lower accuracy. Moreover, images that were given names with higher age of acquisition were also categorised slower and less accurately. However, in contrast to our expectations, longer labels elicited faster responses. Furthermore, participants were slower to categorise items that followed labels that were more familiar. Given the strong relationships between psycholinguistic characteristics such as age of acquisition, frequency and familiarity, we must be cautious to draw strong conclusions regarding individual effects from these exploratory analyses.

The inclusion of psycholinguistic characteristics did not change the previously observed effects of perceptual processing time and sensorimotor information on accuracy and RT. However, in contrast to our confirmatory analyses, linguistic distributional information did not predict accuracy above and beyond the time for perceptual processing, sensorimotor distance and psycholinguistic characteristics. This suggests that at least some of the variance explained by our measure of linguistic distributional information was accounted for by psycholinguistic measures. A potential explanation is the fact that we calculated linguistic distributional distance measures from log co-occurrence vectors, which may correlate with measures of age of acquisition and word frequency. Future work might therefore explore alternative measures of linguistic distributional similarity, for example PPMI derived similarity (Bullinaria & Levy, 2007; Wingfield & Connell, 2019) to determine whether linguistic distributional information and psycholinguistic characteristics explain distinct aspects of categorisation behaviour. Furthermore, we found evidence of an interaction between the shortest SOA (17ms) and linguistic distributional information in predicting accuracy.

In conclusion, our confirmatory findings show that performance differences in ultra-rapid categorisation may at least partially be predicted by overlap in sensorimotor and linguistic distributional information between category and member concepts. Contrasting previous work in speeded categorisation (e.g., van Hoef et al., 2022a - see Chapter 2), our confirmatory analyses suggest that linguistic distributional information predicted accuracy above and beyond sensorimotor information. This suggests that when participants have limited time to perceive an object, they may rely more strongly on linguistic distributional information, however we did not observe an increase in the effect of sensorimotor relative to linguistic distributional information as more time for perceptual processing became available. Therefore, we suggest that further work is needed to determine whether the longest SOA used in this study was still too short to observed flip to reliance on sensorimotor information. While these findings are not exactly as we hypothesised, they nevertheless fit with a sensorimotor-linguistic account of categorisation, which argues categorisation may dynamically depend on overlap in sensorimotor and/ or linguistic distributional overlap.

5.7. References

Anderson, J. R., & Reder, L. M. (1974). Negative judgments in and about semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(6), 664–681. https://doi.org/10.1016/S0022-5371(74)80054-X

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

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

Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (2007). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1013–1026. https://doi.org/10.1037/0096-1523.33.5.1013

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469. https://doi.org/10.1016/j.visres.2005.01.004

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

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

Barsalou, L. W., Santos, A., Simmons, K. W., & Wilson, C. D. (2008). Language and Simulation in Conceptual Processing. In M. de Vega, A. M. Glenberg, & A. Graesser (Eds.), *Symbols, Embodiment and meaning* (pp. 245–283). Oxford University Press.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526. https://doi.org/10.3758/BF03193020

Carota, F., Moseley, R., & Pulvermüller, F. (2012). Body-part-specific Representations of Semantic Noun Categories. *Journal of Cognitive Neuroscience*, *24*(6), 1492–1509. https://doi.org/10.1162/jocn_a_00219

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

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., Brand, J., Carney, J., Banks, B., & Brysbaert, M. (Manuscript in preparation). *Emergent categorical structure from sensorimotor grounding in a full-size human conceptual system. Connell, L., Brand, J., Carney, J., Banks, B., Brysbaert, M., & Lynott, D. (2021).*

Connell, L., & Lynott, D. (2010). Look but don’t touch: Tactile disadvantage in processing modality-specific words. *Cognition*, *115*(1), 1–9. https://doi.org/10.1016/j.cognition.2009.10.005

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & 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

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*(2), 163–201. https://doi.org/10.1037/0096-3445.132.2.163

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

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

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

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*(3), 248–277. https://doi.org/10.1037/0096-3445.126.3.248

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243–275. https://doi.org/10.1016/0010-0285(84)90009-4

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. 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

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The Time-Course of Visual Categorizations: You Spot the Animal Faster than the Bird. *PLOS ONE*, *4*(6), e5927. https://doi.org/10.1371/journal.pone.0005927

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*(1), 28–35. https://doi.org/10.3758/PBR.15.1.28

Mack, M. L., & Palmeri, T. J. (2011). The Timing of Visual Object Categorization. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00165

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*(3), 551–569. https://doi.org/10.1037/a0039184

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 54–70. https://doi.org/10.1037/0278-7393.23.1.54

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 70–84. https://doi.org/10.1037/0278-7393.11.1.70

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical Structure in Concepts and the Basic Level of Categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts and Categories* (pp. 93–131). Psychology Press - Taylor & Francis.

Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 1–20. https://doi.org/10.1016/S0022-5371(82)90412-1

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

Rolls, E. T., Tovée, M. J., & Panzeri, S. (1999). The Neurophysiology of Backward Visual Masking: Information Analysis. *Journal of Cognitive Neuroscience*, *11*(3), 300–311. https://doi.org/10.1162/089892999563409

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

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

Snodgrass, J., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of*

Experimental Psychology: Human Learning and Memory, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

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

van Hoef, R., Lynott, D., & Connell, L. (Manuscript in Preparation). *The Role of Sensorimotor and Linguistic Information in the Basic-level Advantage*.

van Hoef, R., Lynott, D., & Connell, L. (Manuscript in Preparation). *Timed written picture naming for 800 photographs of 200 objects in English*.

VanRullen, R., & Thorpe, S. J. (2001). Is it a Bird? Is it a Plane? Ultra-Rapid Visual Categorisation of Natural and Artifactual Objects. *Perception*, *30*(6), 655–668. https://doi.org/10.1068/p3029

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

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

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Chapter 6: General Discussion

The contribution of this thesis to the literature is discussed in three parts. First, I will outline the thesis aims and corresponding findings. Next, I will address the contributions of this work and its place in the larger literature, by comparing it to previous work on object categorisation

6.1. Thesis aims

Traditional accounts of object categorisation are built on feature-based assumptions about the nature of concepts and the structure of semantic memory. These accounts argue that concepts and categories are represented either as abstract feature summaries (i.e., prototype), exemplars, or nodes in a hierarchically organised semantic memory. On these accounts, categorisation is the process of judging an instance's feature-similarity to the category prototype or stored exemplars. By contrast, current views on the nature of conceptual representation argue that concepts are represented in semantic memory not as collections of abstract, binary features or dimensions or as exemplars, but rather through a combination of (partially reactivated) sensorimotor experience (i.e., perception-action and affective experience of the world around us) and linguistic distributional knowledge (i.e., knowledge about the distribution of words in language). Crucially however, as of yet there exists no explicit simulated-linguistic account of object categorisation, or of the behavioural effects (e.g., taxonomic, typicality) that feature-based accounts were designed to explain. If concepts are not represented in semantic memory as feature-abstractions (e.g., prototype theory), experience with exemplars, or nodes in a hierarchical network, then by what means is category membership determined? Moreover, how can variations in categorisation speed and accuracy that are typically explained by referring to a particular feature-composition be explained?

This thesis aimed to evaluate the central assumptions of traditional accounts of categorisation, and to test a simulated-linguistic account of categorisation that builds on traditional accounts of categorisation in the iterative tradition that is our growing understanding of conceptual representation. To that end, this work in thesis tested the hypothesis that category membership may be determined by assessing the degree of overlap in sensorimotor experience and linguistic distributional knowledge between the concepts that represent instances and categories alike. Specifically, this work has zeroed in on behavioural effects that are synonymous with feature-based theories of conceptual representation; taxonomic and typicality effects, with a primary focus on taxonomic effects. Crucially, the aim of this work was not to discard previous explanations of these effects, but rather to adapt them to work in a sensorimotor-linguistic framework. The research collected in this thesis sought to provide evidence for a simulated-linguistic account of categorisation, and investigated whether sensorimotor and/or linguistic distributional information can explain basic-level advantage effects, exceptions to the basic-level advantage (e.g., due to object typicality or time available for perceptual processing), as well as to test the prediction that the relative importance of sensorimotor over linguistic distributional information varies with task, specifically when there is limited time to perceive an object (e.g., rapid categorisation).

The work reported in Chapters 2 and 3 focuses primarily on the classic basic-level advantage in object categorisation, that traditional accounts typically attribute to the feature-characteristics of categories at specific taxonomic levels (i.e., the basic-level advantage), or the notion of a preferred entry-level into a taxonomically structured semantic memory. I contrasted this with a sensorimotor-linguistic distributional hypothesis, which assigns no particular importance to categorisation at any taxonomic level, but rather argues that in categorisation, member concepts (e.g., Labrador) overlap in that are closer in sensorimotor experience and linguistic distributional context to a category concept (e.g., *dog*, *animal*) are processed faster and more accurately. In Chapter 2, I used a classic label → picture categorisation task, and found that overlap in sensorimotor experience between member and category concepts best predicted the time it took participants to make a decision. Both sensorimotor and linguistic distributional information predicted categorisation, but not above and beyond taxonomic level. Furthermore, adjusting measures of sensorimotor and linguistic distributional information to incorporate object typicality effects improved how well in particular sensorimotor information predicted both RT and accuracy. This finding suggests that sensorimotor overlap may capture categorical graded structure, although imperfectly (see limitations). In Chapter 3, I used a forced-choice response task whereby participants were presented with a cue (e.g., *Labrador*) followed by two correct category options, one basic (e.g., *dog*) and one superordinate (e.g., *animal*). This study showed that sensorimotor and linguistic distributional information affect the processing of basic-level versus superordinate-level labels, reducing the likelihood of a basic-level choice when the overlap in sensorimotor and linguistic distributional information was greater between the cue and the superordinate choice, but also revealed intricate interactions with word frequency.

The effects observed in Chapters 2 and 3 are not clear-cut. For example, the studies reported in Chapter 2 showed that an overlap in sensorimotor experience between a category and member concept was a better predictor of response latencies than linguistic distributional information. By contrast, overlap in linguistic distributional information was a better predictor of accuracy than overlap in sensorimotor experience, but not better than taxonomic level. Moreover, the findings in Chapter 3 suggest that the effects of sensorimotor and or linguistic distributional information may vary with additional predictors of categorisation performance such as word frequency of the presented stimuli. It may be the case that these findings reflect the flexibility of a conceptual system that relies on sensorimotor and/or linguistic distributional information, and is affected by other factors such as word frequency and familiarity. Exploratory analyses indeed hint at a relationship between the time course of the categorisation decision and psycholinguistic characteristics of the category labels, such as their associated age of acquisition, familiarity and length (but see limitations and directions for future research below).

The primary aim of Chapter 4 was to collect a resource researchers may draw upon for future studies on for example object recognition, object naming and object categorisation. However, it also revealed a critical methodological advantage of including the psycholinguistic information (e.g., word frequency) for all names collected for an image rather than just the modal name. This measure proved to be a better predictor of recognition latency. In Chapter 5, we built on this finding, by including the weighted average sensorimotor and linguistic distance between a target category label and *all* names given to a particular image. As such, the picture naming study described in chapter 4 not only gave us a clearer insight into what names were given to an object, it also allowed us to calculate measures of sensorimotor and linguistic distance that were rooted in empirical evidence about the names given to an image rather than an assumption on the part of the researcher.

Previous accounts have suggested that the relative importance of sensorimotor and or linguistic distributional information in conceptual processing may vary depending on task demands (Barsalou et al., 2008; Connell, 2018; Connell & Lynott, 2014). Therefore, the study described in Chapter 5 explored whether limiting perceptual processing and time to respond might affect the degree to which sensorimotor and or linguistic distributional information influence response latencies and accuracy. As in the study described in Chapter 2, linguistic distributional overlap between member and category concepts was a better predictor of categorisation accuracy than sensorimotor overlap, but contrasting the Chapter 2 studies, it also decisively outperformed taxonomic level. These results are complicated by a set of exploratory analyses, which suggest that the effect of linguistic distributional information on accuracy is reduced by the inclusion of psycholinguistic characteristics of the category label (see limitations and future directions below).

6.2. Contributions

In this thesis, I examined the novel idea that category-membership is determined by the degree of overlap in sensorimotor and linguistic distributional knowledge between a given category (e.g., *dog*) and member concept (e.g., *Labrador*). Specifically, I examined two key effects in object categorisation research: the effects of taxonomic level (e.g., the basic-level advantage) and typicality on speed and accuracy in object categorisation. This work takes advantage of the recent development of state-of-the-art measures of sensorimotor experience and linguistic distributional knowledge (e.g., Lynott et al., 2019; Wingfield & Connell, 2019). The results presented here suggest that overlap in sensorimotor and/or linguistic distributional information successfully capture some of the relationship between a category and its members, and predicts aspects of categorisation behaviour, such as taxonomic effects (Chapters 2, 3 and 5) and the interaction between graded category structure and taxonomic effects (Chapter 2) that are traditionally explained by referring to features or networks.

Moreover, these results also suggest that the relative importance of sensorimotor and linguistic distributional information to categorical decision-making varies from task to task, where testing paradigm, stimulus type and amount of time available for perceptual processing affected the degree to which sensorimotor and/or linguistic distributional information predicted categorisation performance (RT and accuracy). This is in line with previous sensorimotor-linguistic accounts of conceptual processing, which suggest the conceptual system may draw differently on both information types depending on task constraints. Furthermore, the findings in Chapter 3, and in the exploratory analyses of in particular Chapters 3 and 5 suggest that in particular tasks object categorisation may draw on sensorimotor and/ or linguistic distributional information as well as on other psycholinguistic variables (e.g., word frequency, age of acquisition and familiarity), to inform a categorisation decision.

The findings in Chapter 2 and 5 provide tentative support for the Linguistic Shortcut Hypothesis (Connell, 2018). The measure of linguistic distributional overlap was slightly better than sensorimotor distance at predicting categorisation accuracy (but not RT) in speeded category verification, but much better in a ultra-rapid categorisation paradigm, where there was limited time to perceive and judge an object, With very little time to accumulate the perceptual information that is argued underpins object recognition (e.g., Busemeyer & Townsend, 1993; Ratcliff & Rouder, 2000), participants may have defaulted to rely on linguistic distributional information instead.

6.2.1. Comparison to traditional accounts of object categorisation

It is important to note that the aim of the work presented in this thesis was not to disavow the merits of feature-based theories of concepts. If one assumes features are verbalised expressions of conceptual representations that have been built from repeated sensorimotor experience (e.g., McRae et al., 2005) then features become a convenient methodological shorthand for testing theories on conceptual processing. That is to say, most recent feature-based work does not argue that features are the stuff that concepts are made off, but rather that they serve as a practical tool in measuring conceptual representations. Indeed, similarity measures derived from feature-norms have been shown to predict conceptual processing in a range of behavioural tasks (Vigliocco et al., 2004). However, it is important to note that there are still limitations to such an account. For example, participants will readily name features that have no clear sensorimotor correlate (e.g., *used to tell time* for *clock*; Yee et al., 2013), suggesting that even if one assumes features reflect sensorimotor information, they do not *only* reflect sensorimotor information. Moreover, critical features may not easily be verbalised for all concepts, in particular abstract concepts (e.g., what are the features of *hunger*?). Furthermore, it is unclear how features are grounded in sensorimotor experience.

Even though some recent feature-based theories of conceptual processing have explicitly argued against the overly simplified interpretation that features make up concepts, traditional accounts of categorisation typically do not nuance the features that underly their theories in this manner. Instead, in such theories the fundamental reliance on features as the building block of concepts and categories is rarely questioned. When faced with fundamental questions (e.g., absence of necessary and sufficient features for certain categories; the disconnect between the processes underlying determining typicality and category membership; basic-level advantage effects), the answer typically involved more features (e.g., characteristic features, distinctive features), different types of features (e.g., functional features), or a new framework to support features (e.g., contextual knowledge to constrain features).

In contrast to the definitional, prototype, and exemplar view of categorisation, the sensorimotor-linguistic view of categorisation assumes no a priori difference between feature, member, and category concepts. All concepts are thought to comprise sensorimotor and or linguistic distributional information, which may be drawn on differently depending on the task at hand. However, to dismiss all of the traditional findings because of a different view on what constitutes a concept, would be to throw out the baby with the bathwater. If we work from the basic assumption that sensorimotor-linguistic similarity between member and member, and member and category, determine category membership, some of the process models and theories outlined by traditional accounts may be repurposed. For example, the process model of label → picture categorisation can be adapted to rely on sensorimotor experience and linguistic distributional knowledge rather than features. Furthermore, our perceptual experience with specific exemplars may form an important part of particular conceptual representations, and relationships between concepts (e.g., *car – driver*), which are the focus of theory-theory accounts, is still very much a part of concepts. I will expand on these examples below.

The results presented in Chapter 2 provide the first evidence that overlap in sensorimotor experience determines how quickly participants decide that a depicted object belongs to a given category better than a division into discrete taxonomic levels. Nevertheless, it was generally true that photographs were categorised faster and more accurately at the basic level. How would a sensorimotor-linguistic account explain this? The simplest explanation would be that basic-level labels generally (but not always) activate sensorimotor simulations which are closer to those activated by the image that follows them than subordinate or superordinate labels. This explanation is inspired by Murphy and Smith's (1982) preparation model, which assumes that a given category label activates a conceptual representation. This representation is subsequently verified upon seeing an image of an object. If the object is a clear match or mismatch, a decision is made. If the object is not a clear match or mismatch, additional processing is required. The preparation model assumes that superordinate-level categorisation is generally slower because there is no single perceptual representation of concepts at this level of abstraction (e.g., no single perceptual representation of *animal*), and requires more information to be active at the same time. The advantage over the subordinate model then is thought to stem from a higher membership threshold because of greater between-category similarity (e.g., the threshold for determining that something is a *bird* or a *dog* is lower than for determining whether something is a *sparrow* or a *finch*). By contrast, on a sensorimotor-linguistic account, an instance's overlap with a conceptual representation comprising sensorimotor experience and linguistic distributional knowledge determines how quickly and accurately it is categorised. That is, reading a label (e.g., *dog*) reactivates aspects of sensorimotor experience across multiple perceptual, motor and affective areas (e.g., the sound of barking, the touch of fur, but also feelings of fondness or fear). Furthermore, the label activates linguistic distributional information about the linguistic contexts *dog* generally appears in (e.g., *bowl*, *walk*, *park*, *leash*, etc.). Together, both types of information form a rich concept of *dog*, which is matched against the sensorimotor and linguistic distributional information that is activated with a subsequent image. On this account, a basic-level advantage may arise because the label *dog* generally (but not always) activates sensorimotor and linguistic distributional information that more closely matches a perceived object. By contrast, superordinate-level labels (e.g., *animal*) activate sensorimotor experience and linguistic distributional information that does not closely match a perceived object, therefore the activation of more sensorimotor and linguistic distributional information is required. Subordinate-level labels on the other hand may require the activation of more specific sensorimotor and linguistic distributional information.

This explanation of taxonomic effects closely matches the preparation model, but notably also retains aspects of influential feature-based accounts of categorisation. That is, sensorimotor representations of categories (e.g., *Labrador*, *dog*, *animal*) are unlikely to encompass all aspects of sensorimotor and affective experience every time they are activated. Rather, they are partial reactivations of stored sensorimotor experience, that through attentional mechanisms may emphasise particular aspects of our experience depending on the task at hand (Barsalou, 1999, Connell & Lynott, 2014). Linguistic distributional information meanwhile condenses our vast experience with language into statistical patterns. In this manner, the activated sensorimotor experience of *dog* may be an abstraction (summary if you will) of our experience with various *dogs*, and involve linguistic distributional knowledge that is an abstraction over the many linguistic contexts in which *dog* occurs. However, our sensorimotor experience of *dog* may also focus particularly on a specific exemplar. For example, people's concept of *dog* may be heavily influenced by experience with a specific *dog* (e.g., their pet corgi), which through regular exposure has become an important aspect of their representation of dogs. That is, when categorising a given instance as dog, they might activate parts of their perceptual and motor experience with their dog, in addition to linguistic distributional knowledge about dogs, and compare them to the target instance. A conceptual system that relies on sensorimotor and linguistic distributional information in this manner is dynamic (Barsalou, 2003b) in that the most accessible conceptual content at a given point in time depends on frequency, recency and context. Determining whether a given object is a *dog*, is *my dog* or is *a large dog*, is likely to activate reasonably different sensorimotor experience and linguistic distributional knowledge. A sensorimotor-linguistic account of categorisation thus flexibly encompasses both summary and exemplar representations, which is compatible with mixed prototype-exemplar accounts (e.g., Smits et al., 2002; Storms et al., 2002; J.D. Smith & Minda, 2000), but does so without the problematic intermediate step of transducing perceptual experience into amodal symbols.

Network-accounts of the basic-level advantage and object typicality (e.g., Jolicoeur et al., 1984; Collins & Loftus, 1975) assume that specific concepts are privileged because they occupy a particular level in a hierarchical structure, or because they are associated with contrasting properties (e.g., when determining that an *ostrich* is a *bird*, processing is slowed down because *ostriches* are linked to the property <u>*cannot fly*</u>, which contrasts the property <u>can fly</u> that is linked to *bird*). The sensorimotor-linguistic account presented here does not assume any particular taxonomic level is privileged, or that object atypicality can be captured fully by the possession of contrasting properties. Taxonomic effects are not bound to one particular level of abstraction, and typicality effects can be observed in concepts that arguably do not possess contrasting properties, such as *uneven number*s (Armstrong et al., 1983). Rather, taxonomic and typicality effects may arise spontaneously from the similarity relationships between patterns of neural activation across sensorimotor, affective and linguistic areas. This assumption adapts the differentiation view of taxonomic effects (Murphy & Medin, 1985), which also argues against the privileged status of any particular taxonomic level, but assumes that taxonomic effects are reflective of a perceived structure of static features and dimensions. Moreover, differentiation theory does not attribute representational value to category labels, whereas the account presented follows sensorimotor-linguistic accounts of conceptual representation in arguing that linguistic distributional knowledge is an integral part of conceptual representation.

The work presented here also shares theoretical ground with connectionist or hub-and-spoke approaches to explaining behavioural effects in categorisation (e.g., Rogers & Patterson, 2008), which similarly suggest stress the importance of sensorimotor and to a lesser extent linguistic information, and argue that taxonomic and typicality effects may originate in similarity-relationships between patterns of activation. However, connectionist accounts view sensorimotor experience, linguistic knowledge and semantic representations as related but separate entities (Meteyard et al., 2012). They argue that concepts are not represented through distributed reactivation of the modality-specific systems involved in perception and handling of their referents, but rather that information from these systems forms the input for a separate amodal system or “hub”, which mediates the communication between modality and language-specific areas of the brain (Patterson et al., 2007), in the same way a hidden layer of nodes mediates between input and output layers in a neural network that inspired this line of research (e.g., Rogers & McClelland, 2004). A proposed candidate for the neural location of such a hub is the anterior temporal lobe (ATL), as damage to this particular brain-region has been shown to result in multi-modal impairments (Lambon Ralph et al., 2017; Patterson et al., 2007), which lead to behavioural changes that are argued to be consistent across tasks and stimuli (Lambon Ralph et al., 2017). The present work cannot make such architectural claims, but note that a sensorimotor-linguistic account of categorisation is not incompatible with hub-spoke accounts, as it may well be the case that similarity processes based on sensorimotor and linguistic distributional information are governed by particular brain areas.

Finally, a sensorimotor-linguistic account incorporates the idea central to theory-theory, namely that perceptual similarity by itself is insufficient to determine category membership in many cases (Murphy & Medin, 1985). Indeed, both sensorimotor and linguistic distributional aspects of concept representations have been argued to represent *relationships* between concepts. That is, a sensorimotor representation of *car* may include experience of *driving*, and depending on context may activate information about *drivers*. The same relationship can be expressed in terms of linguistic distributional knowledge, where *driver* and *car* may occur in the same linguistic contexts (e.g., *steering wheel*, *driver's seat*, *driving permit, vehicle*). In another example, both sensorimotor and linguistic information have been argued be able to represent spatial relationships (Louwerse, 2008; Zwaan & Yaxley, 2003), whereby the canonical spatial relationships between objects (e.g., *branches* are higher than *roots*) is represented through perceptual experience *and* linguistic information. It is possible that knowledge about relationships informs categorisation (see also Gentner & Kurtz, 2005). For example, when determining whether a particular object is a *car*, the process of matching the perceived object to the category may include the activation of *driver* as a relevant aspect of cars. Likewise, when perceiving an unknown object in the sky, we may activate sensorimotor information as well as linguistic distributional information to constrain categorical information (e.g., the perceptual information that *birds* and *planes* typically appear in the sky, or that they appear in the same linguistic contexts).

6.3. Limitations and future directions

6.3.1. Object typicality

Barsalou (1982, 1983) suggests that the mechanisms for determining object typicality differ from those for determining category membership. Indeed, the task demands for a category membership judgment and a typicality judgment could be considered to be quite different. Where the first is generally an open question with a binary response (e.g., does this object belong to the category x, yes or no?), the second is less open (i.e., category membership is generally already determined) and often requires an answer that is not binary (e.g., ratings). As a consequence, people may make radically different judgments and engage different contrast sets (e.g., members vs. non-members for categorisation, and comparison to other members object typicality judgments). Indeed, there are examples of atypical items that most people would decisively place within a given category (e.g., *penguin* comes to mind). However, while it may be the case that category-membership verification and object typicality judgments require a different comparison, they are not unrelated (e.g., Jolicoeur et al., 1984; Murphy & Brownell, 1985). If category-verification is argued to depend on a judgment of similarity in terms of sensorimotor experience and/or linguistic distributional knowledge, it must also explain aspects of graded structure.

The work in described in Chapter 2 (experiment 2 and 3b) showed little effect of the interaction between rated object typicality and subordinate-level categorisation. This finding is in contrast with previous work which found that categorisation performance for atypical items improved at the subordinate compared to the basic level (Jolicoeur et al., 1984; Murphy & Brownell, 1985). A possible explanation may be found in the limited range of typicality ratings for the object our study. Nevertheless, adjusting sensorimotor distance to reflect graded structure via linguistic distributional distance greatly improved its ability to predict both RT and accuracy, suggesting that in this particular study, variations in sensorimotor overlap captured aspects of graded structure that object typicality ratings did not. What does this finding mean for a sensorimotor-linguistic account of typicality effects in categorisation? One possible answer is that object typicality ratings one the one hand and sensorimotor and/or linguistic distributional information on the other hand reflect distinctly different phenomena. This explanation is supported by Banks et al., (2020) who found that object typicality ratings (collected in the same experiment as the typicality ratings used in the present work) predicted label production frequency, rank and RT in a category naming study, whereas sensorimotor and linguistic distributional information predicted the ordinal naming position of a given member concept. Taken together, these results suggest that in particular sensorimotor information captures some aspects of graded structure, but that these aspects may not necessarily be equal to those captured by object typicality ratings. However, more work is needed to flesh out the relationship between measures of overlap in sensorimotor experience and linguistic distributional knowledge on the one hand, and ratings of object typicality on the other hand.

It is not clear why evidence for the relationship between linguistic distributional information and object typicality ratings was negative, while adjusting measures of sensorimotor distance to reflect graded structure via linguistic distributional distance and in accordance with traditional accounts of typicality interactions with taxonomic level worked to improve its predictive power. This particular finding contrasts previous work on the relationship between linguistic distributional information from a count-vector model (LSA) and object typicality ratings (Connell & Ramscar, 2001) and is particularly remarkable since Banks et al. (2021) found evidence that a similar measure of linguistic distributional information and object typicality ratings from the same study *did* correlate. As stated before, it is possible that the lack of an effect of object typicality observed in Chapter 2 was the result of a particular stimulus selection. For example, one possible cause for the lack of a typicality-effect may have been that the stimuli we used in the Chapter 2 studies were all rated moderately to highly typical (with one exception, *gavel*, which had to be excluded from the analysis for deviating too much from the other stimuli), which may have been because they were not explicitly selected to range in object typicality. As such, we may not have been able to detect the effect object typicality has on categorisation judgments, or of its relationship to linguistic distributional information.

To explore this matter further, a future study might explore both the individual effects of object typicality ratings, overlap in sensorimotor experience and overlap in linguistic distributional knowledge on category verification, as well as their relationships on a set of items with a broader object typicality range. The normed image set described in Chapter 4 might prove a useful resource in that regard, as it allows for the calculation of weighted average sensorimotor and linguistic distance between a given category label (e.g., *bird*) and the names given to a particular image (e.g., *ostrich*, *emu, bird, animal, creature* etc.). Theoretically, images that contain atypical objects should be named more frequently with specific names, which might mean that weighted average sensorimotor and linguistic distributional information for atypical items might show greater overlap with sensorimotor and linguistic distributional information activated by a subordinate (e.g., *ostrich*) rather than a basic-level label. Therefore, a measure sensorimotor and linguistic distributional similarity between weighted average image names and subordinate-level labels might correlate with object typicality ratings. To that end, a future study might also collect typicality ratings for the normed image set described in chapter 4.

Another possibility for the lack of object typicality effects is that our attempt at finding an effect of object typicality did not focus on any particular taxonomic level. That is, typicality has traditionally been linked to categorisation performance at the basic and subordinate level. However, we tested object typicality on a full dataset, which included superordinate items on which basic-level object typicality arguably may not have much of an effect (Murphy & Brownell, 1985), for example, the typicality of *Labrador* as a member of *dog* might not be relevant to the category of *Labrador* as an *animal*. It may be the case that a wider variety of objects, specifically manipulating a range of object typicality ratings, may allow for a clearer comparison between the effect of typicality ratings and sensorimotor-linguistic measures on RT and accuracy in categorisation.

6.3.2. Taxonomic information occasionally outperforms sensorimotor-linguistic information

Throughout the studies described in Chapters 2, 3 and 5, taxonomic level predicted various aspects of performance (RT and/or accuracy) above and beyond overlap in sensorimotor and/or linguistic distributional information between category and member. That is, we found the classic basic-level advantage in either RT or accuracy in all three studies (and the replications of the studies described of chapter 2). In and of itself, the fact that taxonomic levels were good predictors of categorisation performance does not constitute evidence against a sensorimotor-linguistic account of categorisation. As I have argued in Chapter 2, just as feature-based theories assumed features were organised in such a way (e.g., matching a perceived structure of features) that they gave rise to taxonomic effects, it is possible that taxonomic effects arise spontaneously from sensorimotor and linguistic distributional information. In fact, Connell et al., (2021) found that the relative extent to which different perceptual modalities and action effectors underlie sensorimotor experience of concepts is sufficient to cluster concepts into an inclusive hierarchy (e.g., the concepts *sparrow*, *crow*, *eagle*, and *bird* tend to be moderately strong in visual, auditory and head-action experience; moderately weak in haptic and hand-action strength and very weak on all other sensorimotor dimensions, and hence tend to cluster together).

If taxonomic levels simply encode sensorimotor and linguistic distributional information, the question is why our predictor of taxonomic level explains additional variance to sensorimotor-linguistic measures in accuracy (Chapter 2) and RT (Chapter 3, chapter 5). Possible explanations lie in word frequency and object familiarity, which I did not explicitly control for (with the exception of word frequency in Chapter 3). Previous work in has suggested that response times in categorisation may be affected by stimulus familiarity (McCloskey, 1980), whereby more familiar stimuli are processed faster than less familiar ones. Similarly, Barsalou (2003b) argues that a concept's frequency of instantiation affects how accessible information is. While often conflated with object typicality, some research suggests that familiarity affects categorisation separately (Malt & Smith, 1982). As such, it is possible that additional explanatory value of taxonomic levels in predicting RT or accuracy comes from the fact that participants were simply faster at making judgments for objects they were more familiar with.

6.3.3. Linguistic distributional information and psycholinguistic characteristics

Any study that employs labels either implicitly or explicitly is sensitive to variations in the psycholinguistic characteristics of those labels. In studying taxonomic effects in particular, word frequency, age of acquisition, familiarity and even word length might all be variables that inform the ease of categorical decision making (e.g., Rosch, 1976). The confirmatory analyses in this work suggest that taxonomic level indeed interacts with word frequency in predicting RT (Chapter 3) in a forced-choice categorisation task. Furthermore, while they were not explicitly designed to accurately determine the (often related) effects of psycholinguistic characteristics on object categorisation, the exploratory analyses included in Chapters 2,3 and 5 suggest that characteristics such as age of acquisition, word frequency, familiarity and word length also inform categorical decision-making to varying degrees.

Overall, psycholinguistic characteristics of the category label did not predict categorisation performance better than overlap in sensorimotor experience or than taxonomic level. However, the predictive effect of linguistic distributional information was reduced when psycholinguistic characteristics were included in the model.

Previous research (Murphy & Smith, 1982) has provided some evidence against the explanations based of the basic-level advantage based on word-frequency and order-of-learning. Consequently, Murphy and Smith argue that the assumption that basic-level labels are preferred because of their frequency and the order in which they are acquired ignores the underlying structure of concepts and categories. Indeed, this assumption seems to create more questions than it answers. What is it about basic-level labels that means they are used more frequently, and learned earlier? It could indeed be the case that psycholinguistic characteristics reflect rather than determine conceptual structure (e.g., patterns of activation of sensorimotor and/or linguistic distributional knowledge). However, where psycholinguistic characteristics informed performance, they did so separately from sensorimotor information. Linguistic distributional information meanwhile was affected by the inclusion of psycholinguistic characteristics. This may indicate that psycholinguistic predictors and linguistic distributional knowledge explain similar aspects of categorisation behaviour. However, the present work used a linguistic distributional measure based on vectors of log co-occurrence, which may correlate with psycholinguistic characteristics such as age of acquisition and word frequency. As such, future work might explore the use of linguistic distributional measures that are less strongly related to psycholinguistic characteristics, to explore whether linguistic distributional information and psycholinguistic characteristics explain distinct variance or whether they overlap. Furthermore, such a study would need to take care to use orthogonal designs to separate strongly related psycholinguistic characteristics such as age of acquisition, familiarity and word frequency.

6.3.4. Direct comparison to features

The work presented in this thesis does not directly compare feature-based predictors of categorisation to sensorimotor-linguistic predictions. Instead, it focuses on explaining effects that have traditionally been explained from a feature-based perspective, for example the basic-level advantage, and the superordinate-level advantage in ultra-rapid categorisation. As stated earlier, the aim of this work has not been to dismiss the notion that features have been a useful measure of operationalising and thinking about conceptual knowledge, but rather to stress that an alternative explanation is possible: sensorimotor-linguistic overlap. Nevertheless, now that we have established a framework for how a sensorimotor-linguistic account might explain categorisation, future studies might explore direct comparisons, such as the extent to which the measures of sensorimotor experience and linguistic distributional knowledge used in this study overlap with the predictive power of feature-based alternatives, such as similarity measures derived from feature-norms.

6.3.5. Technical improvements

Due to COVID-19, the study described in Chapter 5, which was initially intended to be run in the lab, was moved to a web-based testing environment. While the SOA patterns in accuracy and RT that we observed are generally in line with what previous researchers found (e.g., Bacon-Macé et al., 2005; Mack & Palmeri, 2015), we had to forego some of the tighter controls on stimulus presentation time that would have been available to us in a lab, such as having control over the screen refresh rate and size, as well as ruling out lag due to internet connection, browser type and installed plugins, or background processes. To counter this, we did extensive pilot testing of the experiment across multiple browsers (Chrome, Firefox, Edge, Safari, Opera), operating systems (Windows, MacOS and Linux) and displays. Furthermore, we applied a strict screening and accuracy threshold, replacing a considerable number of participants, and set a higher upper boundary for our sequential hypothesis testing. In addition to this, we used a script to load all images into a buffer at the beginning of each block, to ensure loading times did not affect stimulus presentation. A future, lab-based study, could aim to replicate the results described in Chapter 5, while maintaining tighter control over external variables and timings.

6.4. Concluding remarks

This thesis explored object categorisation from a sensorimotor-linguistic perspective on conceptual representations. Crucially, it has shown that overlap in sensorimotor and/or linguistic distributional information between a member and category concept predicts aspects categorisation performance in speeded and ultrarapid categorisation. The relative importance of either information type may change depending on task constraints. As such, the findings in this work are in line with sensorimotor-linguistic accounts of conceptual processing, and are the first step in providing an alternative to traditional accounts of object categorisation. Sensorimotor-linguistic accounts allow for a great richness of information to come to bear on cognitive processes, and we may only be scratching the surface of what such accounts can explain.

Consolidated bibliography

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of Clinical and Experimental Neuropsychology*, *31*(6), 731–753. https://doi.org/10.1080/13803390802488103

Aerts, D., Broekaert, J., Gabora, L., & Sozzo, S. (2016). Generalizing Prototype Theory: A Formal Quantum Framework. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00418

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140–155. https://doi.org/10.3758/BF03195559

Anderson, J. R., & Reder, L. M. (1974). Negative judgments in and about semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(6), 664–681. https://doi.org/10.1016/S0022-5371(74)80054-X

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. https://doi.org/10.1037/a0016261

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*(3), 263–308. https://doi.org/10.1016/0010-0277(83)90012-4

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning: From Acquiring Specific Items to Forming General Rules. *Current Directions in Psychological Science*, *21*(3), 170–176. https://doi.org/10.1177/0963721412436806

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

Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (2007). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human*

Perception and Performance, *33*(5), 1013–1026. https://doi.org/10.1037/0096-1523.33.5.1013

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469. https://doi.org/10.1016/j.visres.2005.01.004

Banks, B., & Connell, L. (2021). Category production norms for 117 concrete and abstract categories in British English. *Manuscript in Preparation*.

Banks, B., Wingfield, C., & Connell, L. (2021). Linguistic Distributional Knowledge and Sensorimotor Grounding both Contribute to Semantic Category Production. *Cognitive Science*, *45:* e13055. https://doi.org/10.1111/cogs.13055.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. Context-predicting semantic vectors. In K. Toutanova & H. Wu (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (pp. 238–247). Association for Computational Linguistics. doi:0.3115/v1/P14-1023

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart Pictures: Effects of Age of Acquisition, Frequency, and Name Agreement. *Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 560–585. https://doi.org/10.1080/027249897392026

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*(1), 82–93. https://doi.org/10.3758/BF03197629

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227. https://doi.org/10.3758/BF03196968

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *11*(4), 629–654. https://doi.org/10.1037/0278-7393.11.1-4.629

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

Barsalou, L. W. (2003a). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1435), 1177–1187. https://doi.org/10.1098/rstb.2003.1319

Barsalou, L. W. (2003b). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*(5–6), 513–562. https://doi.org/10.1080/01690960344000026

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., Santos, A., Simmons, K. W., & Wilson, C. D. (2008). Language and Simulation in Conceptual Processing. In M. de Vega, A. M. Glenberg, & A. Graesser (Eds.), *Symbols, Embodiment and meaning* (pp. 245–283). Oxford University Press.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380. https://doi.org/10.3758/BF03196494

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical-semantic competition hypothesis. *Cognition*, *96*(2), B45–B54. https://doi.org/10.1016/j.cognition.2004.11.006

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*(1), 38–64. https://doi.org/10.1016/0010-0285(88)90024-2

Bohn, R., & Short, J. (2009). *How Much Information?* University of California.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, *93*(1), 89–114. https://doi.org/10.1348/000712602162463

Bonin, P., Méot, A., Laroche, B., Bugaiska, A., & Perret, C. (2019). The impact of image characteristics on written naming in adults. *Reading and Writing*, *32*(1), 13–31. https://doi.org/10.1007/s11145-017-9727-3

Bonner, M. F., & Grossman, M. (2012). Gray Matter Density of Auditory Association Cortex Relates to Knowledge of Sound Concepts in Primary Progressive Aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991. https://doi.org/10.1523/JNEUROSCI.6241-11.2012

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

Boutonnet, B., & Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience*, *35*(25), 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Bramão, I., Reis, A., Petersson, K. M., & Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica*, *138*(1), 244–253. https://doi.org/10.1016/j.actpsy.2011.06.010

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS One*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLOS ONE*, *9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953

Brodie, E. E., Wallace, A. M., & Sharrat, B. (1991). Effect of Surface Characteristics and Style of Production on Naming and Verification of Pictorial Stimuli. *The American Journal of Psychology*, *104*(4), 517–545. JSTOR. https://doi.org/10.2307/1422939

Brooks, L. (1978). Nonanalytic concept formation. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Lawrence Erlbaum Associates, Publishers.

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, *13*(7–8), 992–1011. https://doi.org/10.1080/13506280544000165

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

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526. https://doi.org/10.3758/BF03193020

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instruments, & Computers*, *30*(2), 188–198. https://doi.org/10.3758/BF03200643

Burgess, C., & Lund, K. (1997). Representing abstract words and emotional connotation in a high-dimensional memory space. In *Proceedings of the Cognitive Science Society* (pp. 61–66). Erlbaum.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295X.100.3.432

Carota, F., Moseley, R., & Pulvermüller, F. (2012). Body-part-specific Representations of Semantic Noun Categories. *Journal of Cognitive Neuroscience*, *24*(6), 1492–1509. https://doi.org/10.1162/jocn_a_00219

Catling, J. C., Dent, K., & Williamson, S. (2008). Age of acquisition, not word frequency affects object recognition: Evidence from the effects of visual degradation. *Acta Psychologica*, *129*(1), 130–137. https://doi.org/10.1016/j.actpsy.2008.05.005

Catling, J. C., & Johnston, R. A. (2006). The effects of age of acquisition on an object classification task. *Visual Cognition*, *13*(7–8), 968–980. https://doi.org/10.1080/13506280544000138

Clarke, A., & Tyler, L. K. (2015). Understanding What We See: How We Derive Meaning From Vision. *Trends in Cognitive Sciences*, *19*(11), 677–687. https://doi.org/10.1016/j.tics.2015.08.008

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

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., Brand, J., Carney, J., Banks, B., & Brysbaert, M. (Manuscript in preparation). *Emergent categorical structure from sensorimotor grounding in a full-size human conceptual system*.

Connell, L., & Lynott, D. (2010). Look but don't touch: Tactile disadvantage in processing modality-specific words. *Cognition*, *115*(1), 1–9. https://doi.org/10.1016/j.cognition.2009.10.005

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & 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

Connell, L., & Ramscar, M. (2001). Using Distributional Measures to Model Typicality in Categorization. In *Proceedings of the Twenty-Third Annual Meeting of the Cognitive Science Society* (pp. 226–231). Lawrence Erlbaum Associates.

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, *111*(2), 291–303. https://doi.org/10.1037/0033-2909.111.2.291

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*(2), 163–201. https://doi.org/10.1037/0096-3445.132.2.163

Crick, F. (1989). The recent excitement about neural networks. *Nature*, *337*(6203), 129–132.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Crutch, S. J., & Warrington, E. K. (2008). Contrasting patterns of comprehension for superordinate, basic-level, and subordinate names in semantic dementia and aphasic stroke patients. *Cognitive Neuropsychology*, *25*(4), 582–600. https://doi.org/10.1080/02643290701862290

D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture Naming and Lexical Access in Italian Children and Adults. *Journal of Cognition and Development*, *2*(1), 71–105. https://doi.org/10.1207/S15327647JCD0201_4

Dent, K., Catling, J. C., & Johnston, R. A. (2007). Age of acquisition affects object recognition: Evidence from visual duration thresholds. *Acta Psychologica*, *125*(3), 301–318. https://doi.org/10.1016/j.actpsy.2006.08.004

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

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 515–523. https://doi.org/10.1037/0278-7393.24.2.515

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, *18*(4), 500–549. https://doi.org/10.1016/0010-0285(86)90008-3

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013). Parkinson's Disease Disrupts Both Automatic and Controlled Processing of Action Verbs. *Brain and Language*, *127*(1), 65–74. https://doi.org/10.1016/j.bandl.2012.07.008

Fodor, J. A. (2008). Modularity of the mind. In J. E. Adler & L. J. Rips (Eds.), *Reasoning. Studies of Human Inference and Its Foundations* (pp. 878–914). Cambridge University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Forde, E. M. E., & Humphreys, G. W. (2002). The Role of Semantic Knowledge in Short-term Memory. *Neurocase*, *8*(1–2), 13–27. https://doi.org/10.1093/neucas/8.1.13

Geeraerts, D. (2016). Prospects and problems of prototype theory. *Diacronia*, *4*, 1–16. https://doi.org/10.17684/i4A53en

Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin*. (pp. 151–175). American Psychological Association. https://doi.org/10.1037/11156-009

Gerlach, C. (2009). Category-specificity in visual object recognition. *Cognition*, *111*(3), 281–301. https://doi.org/10.1016/j.cognition.2009.02.005

Glass, A., & Holyoak, K. (1974). Alternative conceptions of semantic theory. *Cognition*, *3*(4), 313–339. https://doi.org/10.1016/0010-0277(74)90002-X

Glenberg, A. M., & Robertson, D. A. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, *43*(3), 379–401. https://doi.org/10.1006/jmla.2000.2714

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

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*(2), 125–157. https://doi.org/10.1016/0010-0277(94)90065-5

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

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-Space Models of Semantic Representation From a Cognitive Perspective: A Discussion of Common Misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Hampton, J. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 441–461. https://doi.org/10.1016/S0022-5371(79)90246-9

Hampton, J. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition*, *9*(2), 149–156. https://doi.org/10.3758/BF03202329

Hampton, J. (1993). Prototype models of concept representation. In I. van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67–95). Academic Press.

Hampton, J. (1995). Testing the Prototype Theory of Concepts. *Journal of Memory and Language*, *34*(5), 686–708. https://doi.org/10.1006/jmla.1995.1031

Hampton, J. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, *65*(2), 137–165. https://doi.org/10.1016/S0010-0277(97)00042-5

Hampton, J. (2007). Typicality, Graded Membership, and Vagueness. *Cognitive Science*, *31*(3), 355–384. https://doi.org/10.1080/15326900701326402

Hampton, J., & Jönsson, M. L. (2012). Typicality and Compositionality: The Logic of Combining Vague Concepts. In M. Werning, W. Hinzen, & E. Machery (Eds.), *The Oxford handbook of compositionality* (pp. 385–402). Oxford University Press.

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

Heit, E., & Barsalou, L. W. (1996). The Instantiation Principle in Natural Categories. *Memory*, *4*(4), 413–452. https://doi.org/10.1080/096582196388915

Heuer, S. (2016). The influence of image characteristics on image recognition: A comparison of photographs and line drawings. *Aphasiology*, *30*(8), 943–961. https://doi.org/10.1080/02687038.2015.1081138

Hilbig, B. E. (2016). Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: A unique clinicopathological syndrome. *The Lancet Neurology*, *6*(11), 1004–1014. https://doi.org/10.1016/S1474-4422(07)70266-1

Hodges, J. R., Patterson, K., & Tyler, L. K. (1994). Loss of semantic memory: Implications for the modularity of mind. *Cognitive Neuropsychology*, *11*(5), 505–542. https://doi.org/10.1080/02643299408251984

Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, *125*(3), 293. https://doi.org/10.1037/rev0000094

Hofmann, B. (2017). Do health professionals have a prototype concept of disease? The answer is no. *Philosophy, Ethics, and Humanities in Medicine*, *12*(1), 6. https://doi.org/10.1186/s13010-017-0047-7

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The Role of Surface Information in Object Recognition: Studies of a Visual Form Agnosic and Normal Subjects. *Perception*, *23*(12), 1457–1481. https://doi.org/10.1068/p231457

Humphreys, G. W., & Forde, E. M. E. (2005). Naming a giraffe but not an animal: Base-level but not superordinate naming in a patient with impaired semantics. *Cognitive Neuropsychology*, *22*(5), 539–558. https://doi.org/10.1080/02643290442000176

Jefferies, E., Thompson, H., Cornelissen, P., & Smallwood, J. (2020). The neurocognitive basis of knowledge about object identity and events: Dissociations reflect opposing effects of semantic coherence and control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20190300. https://doi.org/10.1098/rstb.2019.0300

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, *120*(1), 113–139. https://doi.org/10.1037/0033-2909.120.1.113

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*(3), 248–277. https://doi.org/10.1037/0096-3445.126.3.248

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243–275. https://doi.org/10.1016/0010-0285(84)90009-4

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, *112*(3), 500–526. https://doi.org/10.1037/0033-2909.112.3.500

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

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

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, *99*(2), 199–208. https://doi.org/10.1037/h0034633

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What categories reveal about the mind*. The University of Chicago Press.

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Laurence, S., & Margolis, E. (1999). Concepts and Cognitive Science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core Readings* (pp. 3–81). The MIT Press.

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, *21*(6), 335–346. https://doi.org/10.1038/s41583-020-0277-3

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & 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. 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. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. https://doi.org/10.1016/j.cognition.2009.09.002

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The Time-Course of Visual Categorizations: You Spot the Animal Faster than the Bird. *PLOS ONE*, *4*(6), e5927. https://doi.org/10.1371/journal.pone.0005927

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*(1), 28–35. https://doi.org/10.3758/PBR.15.1.28

Mack, M. L., & Palmeri, T. J. (2011). The Timing of Visual Object Categorization. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00165

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*(3), 551–569. https://doi.org/10.1037/a0039184

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), 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

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition*, *10*(1), 69–75. https://doi.org/10.3758/BF03197627

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 250–269. https://doi.org/10.1016/S0022-5371(84)90170-1

Mandler, J. M., & Bauer, P. J. (1988). The cradle of categorization: Is the basic level basic? *Cognitive Development*, *3*(3), 247–264. https://doi.org/10.1016/0885-2014(88)90011-1

Mandler, J. M., & McDonough, L. (2000). Advancing Downward to the Basic Level. *Journal of Cognition and Development*, *1*(4), 379–403. https://doi.org/10.1207/S15327647JCD0104_02

Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an Integration of Deep Learning and Neuroscience. *Frontiers in Computational Neuroscience*, *10*. https://www.frontiersin.org/article/10.3389/fncom.2016.00094

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 54–70. https://doi.org/10.1037/0278-7393.23.1.54

Martínez-Cuitiño, M., & Vivas, L. (2019). Category or diagnosticity effect? The influence of color in picture naming tasks. *Psychology & Neuroscience*, *12*(3), 328–341. https://doi.org/10.1037/pne0000172

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310. https://doi.org/10.1038/nrn1076

McCloskey, M. E. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 485–502. https://doi.org/10.1016/S0022-5371(80)90330-8

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, *6*(4), 462–472. https://doi.org/10.3758/BF03197480

McRae, K., Cree, G. S., Seidenberg, M. S., & Mcnorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

Medin, D. L. (1989). Concepts and Conceptual Structure. *American Psychologist*, *44*(12), 1469–1481. https://doi.org/10.1037/0003-066X.44.12.1469

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*(2), 158–190. https://doi.org/10.1016/0010-0285(88)90018-7

Medin, D. L., & Smith, E. E. (1984). Concepts and Concept Formation. *Annual Review of Psychology*, *35*(1), 113–138.

Mervis, C. B., & Crisafi, M. A. (1982a). Order of Acquisition of Subordinate-, Basic-, and Superordinate-Level Categories. *Child Development*, *53*(1), 258–266. https://doi.org/10.2307/1129660

Mervis, C. B., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, *32*(1), 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513

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

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv:1301.3781 [Cs]*. http://arxiv.org/abs/1301.3781

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775–799. https://doi.org/10.1037/0278-7393.27.3.775

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An Ecological Alternative to Snodgrass & Vanderwart: 360 High Quality Colour Images with Norms for Seven Psycholinguistic Variables. *PLOS ONE*, *7*(5), e37527. https://doi.org/10.1371/journal.pone.0037527

Morrison, C. M., Hirsh, K. W., & Duggan, G. B. (2003). Age of Acquisition, Ageing, and Verb Production: Normative and Experimental Data. *The Quarterly Journal of Experimental Psychology Section A*, *56*(4), 1–26. https://doi.org/10.1080/02724980244000594

Murphy, G. L. (1982). Cue validity and levels of categorization. *Psychological Bulletin*, *91*(1), 174–177. https://doi.org/10.1037/0033-2909.91.1.174

Murphy, G. L. (2002). *The Big Book of Concepts*. MIT Press.

Murphy, G. L. (2016). Is there an exemplar theory of concepts? *Psychonomic Bulletin & Review*, *23*(4), 1035–1042. https://doi.org/10.3758/s13423-015-0834-3

Murphy, G. L., & Allopenna, P. D. (1994). The Locus of Knowledge Effects in Concept Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 904–919. https://doi.org/10.1037/0278-7393.20.4.904

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *11*(1), 70–84. https://doi.org/10.1037/0278-7393.11.1.70

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical Structure in Concepts and the Basic Level of Categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts and Categories* (pp. 93–131). Psychology Press - Taylor & Francis.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316. https://doi.org/10.1037/0033-295X.92.3.289

Murphy, G. L., & Ross, B. H. (2005). The two faces of typicality in category-based induction. *Cognition*, *95*(2), 175–200. https://doi.org/10.1016/j.cognition.2004.01.009

Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 1–20. https://doi.org/10.1016/S0022-5371(82)90412-1

Murphy, G. L., & Spalding, T. L. (1995). Knowledge, Similarity, and Concept Formation. *Psychologica Belgica*, *35*(2–3), 127–144. https://doi.org/10.5334/pb.882

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 572–586. https://doi.org/10.1037/0278-7393.15.4.572

Navarrete, E., Arcara, G., Mondini, S., & Penolazzi, B. (2019). Italian norms and naming latencies for 357 high quality color images. *PLOS ONE*, *14*(2), e0209524. https://doi.org/10.1371/journal.pone.0209524

Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods*, *37*(3), 398–416. https://doi.org/10.3758/BF03192709

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M., & Palmeri, T. J. (1997). An Exemplar-Based Random Walk Model of Speeded Classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Stanton, R. D. (2005). Speeded Classification in a Probabilistic Category Structure: Contrasting Exemplar-Retrieval, Decision-Boundary, and Prototype Models. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 608–629. https://doi.org/10.1037/0096-1523.31.3.608

Ostarek, M., & Vigliocco, G. (2016). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 579. https://doi.org/10.1037/xlm0000318

Ostergaard, A. L., & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 579–587. https://doi.org/10.1037/0278-7393.11.3.579

Palmer, S. E. (1977). Hierarchical Structure in Perceptual Representation. *Cognitive Psychology*, *9*(4), 441–474. https://doi.org/10.1016/0010-0285(77)90016-0

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying Different-Modality Properties for Concepts Produces Switching Costs. *Psychological Science*, *14*(2), 119–124. https://doi.org/10.1111/1467-9280.t01-1-01429

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin & Review*, *11*, 164–167. https://doi.org/10.3758/BF03206477

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, *51*(6), 2533–2545. https://doi.org/10.3758/s13428-018-1100-1

Perret, C., & Laganaro, M. (2013). Why are written picture naming latencies (not) longer than spoken naming? *Reading and Writing*, *26*(2), 225–239. https://doi.org/10.1007/s11145-012-9365-8

Pind, J., & Tryggvadóttir, H. B. (2002). Determinants of picture naming times in Icelandic. *Scandinavian Journal of Psychology*, *43*(3), 221–226. https://doi.org/10.1111/1467-9450.00290

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363. https://doi.org/10.1037/h0025953

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. MIT Press.

Ratcliff, R., & Rouder, J. N. (2000). A Diffusion Model Account of Masking in Two-Choice Letter Identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 127–140. https://doi.org/DOI: 10.10371/0096-1523.26.1.127

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407. https://doi.org/10.1016/0010-0285(72)90014-X

Rey, G. (1983). Concepts and stereotypes. *Cognition*, *15*(1–3), 237–262. https://doi.org/10.1016/0010-0277(83)90044-6

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

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General*, *122*(4), 468–486. https://doi.org/10.1037/0096-3445.122.4.468

Rips, L. J., Shoben, E., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20. https://doi.org/10.1016/S0022-5371(73)80056-8

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. The MIT Press. https://doi.org/10.7551/mitpress/6161.001.0001

Rogers, T. T., & McClelland, J. L. (2008). Précis of Semantic Cognition: A Parallel Distributed Processing Approach. *Behavioral and Brain Sciences*, *31*(6), 689–714. https://doi.org/10.1017/S0140525X0800589X

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

Rolls, E. T., Tovée, M. J., & Panzeri, S. (1999). The Neurophysiology of Backward Visual Masking: Information Analysis. *Journal of Cognitive Neuroscience*, *11*(3), 300–311. https://doi.org/10.1162/089892999563409

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233.

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition & Categorization*. Lawrence Erlbaum Associates, Publishers.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502. https://doi.org/10.1037/0096-1523.2.4.491

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's Object Pictorial Set: The Role of Surface Detail in Basic-Level Object Recognition. *Perception*, *33*(2), 217–236. https://doi.org/10.1068/p5117

Roth, E. M., & Mervis, C. B. (1983). Fuzzy set theory and class inclusion relations in semantic categories. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 509–525. https://doi.org/10.1016/S0022-5371(83)90310-9

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, *15*(3), 346–378. https://doi.org/10.1016/0010-0285(83)90012-9

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Rumelhart, D. E., McClelland, J. L., & Group, P. R. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. https://doi.org/10.7551/mitpress/5236.001.0001

Rumelhart, D. E., & Norman, D. (1988). Representation in Memory. In R. Atkinson, R. Herrnstein, G. Lindzey, & R. Luce (Eds.), *Steven's handbook of experimental psychology: Vol. 2. Learning and cognition* (p. 137). Wiley.

Salmon, J. P., Matheson, H. E., & McMullen, P. A. (2014). Photographs of manipulable objects are named more quickly than the same objects depicted as line-drawings: Evidence that photographs engage embodiment more than line-drawings. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01187

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 537–555. https://doi.org/10.3758/BF03200541

Sanocki, T., Bowyer, K. W., Heath, M. D., & Sarkar, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 340–349. https://doi.org/10.1037/0096-1523.24.1.340

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

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). 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

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*(4), 1241–1260. https://doi.org/10.3758/s13428-016-0783-4

Severens, E., Lommel, S. V., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*(2), 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Siew, C. S. Q., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive Network Science: A Review of Research on Cognition through the Lens of Network Representations, Processes, and Dynamics. *Complexity*, *2019*, e2108423. https://doi.org/10.1155/2019/2108423

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Harvard University Press.

Smith, E. E., Shoben, E., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214–241. https://doi.org/10.1037/h0036351

Smith, J. D. (2002). Exemplar Theory's Predicted Typicality Gradient Can Be Tested and Disconfirmed. *Psychological Science*, *13*(5), 437–442. https://doi.org/10.1111/1467-9280.00477

Smith, J. D. (2014). Prototypes, exemplars, and the natural history of categorization. *Psychonomic Bulletin & Review*, *21*, 312–331. https://doi.org/10.3758/s13423-013-0506-0

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3–27. https://doi.org/10.1037/0278-7393.26.1.3

Smits, T., Storms, G., Rosseel, Y., & De Boeck, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. *Psychonomic Bulletin & Review*, *9*(4), 836–844. https://doi.org/10.3758/BF03196343

Snodgrass, J., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Snodgrass, J., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 516–536. https://doi.org/10.3758/BF03200540

Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*, 330–336.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598–605. https://doi.org/10.3758/BF03193891

Storms, G., Boeck, P. D., & Ruts, W. (2000). Prototype and Exemplar-Based Information in Natural Language Categories. *Journal of Memory and Language*, *42*, 51–73.

Székely, A., D’Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 621–633. https://doi.org/10.3758/BF03195542

Tanaka, J., & Presnell, L. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*(6), 1140–1153. https://doi.org/10.3758/BF03207619

Tanaka, J., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*(5), 211–215. https://doi.org/10.1016/S1364-6613(00)01626-0

Therriault, D. J., Yaxley, R. H., & Zwaan, R. A. (2009). The role of color diagnosticity in object recognition and representation. *Cognitive Processing*, *10*(4), 335. https://doi.org/10.1007/s10339-009-0260-4

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792–14797. https://doi.org/10.1073/pnas.94.26.14792

Torrance, M., Nottbusch, G., Alves, R. A., Arfé, B., Chanquoy, L., Chukharev-Hudilainen, E., Dimakos, I., Fidalgo, R., Hyönä, J., Jóhannesson, Ó. I., Madjarov, G., Pauly, D. N., Uppstad, P. H., van Waes, L., Vernon, M., & Wengelin, Å. (2018). Timed written picture naming in 14 European languages. *Behavior Research Methods*, *50*(2), 744–758. https://doi.org/10.3758/s13428-017-0902-x

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, *113*(2), 169–193. https://doi.org/10.1037/0096-3445.113.2.169

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual Structure and the Structure of Concepts: A Distributed Account of Category-Specific Deficits. *Brain and Language*, *75*(2), 195–231. https://doi.org/10.1006/brln.2000.2353

Urooj, U., Cornelissen, P. L., Simpson, M. I. G., Wheat, K. L., Woods, W., Barca, L., & Ellis, A. W. (2014). Interactions between visual and semantic processing during object recognition revealed by modulatory effects of age of acquisition. *NeuroImage*, *87*, 252–264. https://doi.org/10.1016/j.neuroimage.2013.10.058

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

van Hoef, R., Lynott, D., & Connell, L. (Manuscript in Preparation). *The Role of Sensorimotor and Linguistic Information in the Basic-level Advantage*.

van Hoef, R., Lynott, D., & Connell, L. (Manuscript in Preparation). *Timed written picture naming for 800 photographs of 200 objects in English*.

VanRullen, R., & Thorpe, S. J. (2001). Is it a Bird? Is it a Plane? Ultra-Rapid Visual Categorisation of Natural and Artifactual Objects. *Perception*, *30*(6), 655–668. https://doi.org/10.1068/p3029

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*(4), 422–488. https://doi.org/10.1016/j.cogpsych.2003.09.001

Vinson, D., Thompson, R. L., Skinner, R., & Vigliocco, G. (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language*, *82*, 56–85. https://doi.org/10.1016/j.jml.2015.03.002

Voorspoels, W., Vanpaemel, W., & Storms, G. (2008). Exemplars and prototypes in natural language concepts: A typicality-based evaluation. *Psychonomic Bulletin & Review*, *15*, 630–637. https://doi.org/10.3758/PBR.15.3.630

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

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, *27*(4), 635–657. https://doi.org/10.1080/14640747508400525

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

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

Wisniewski, E. J., & Medin, D. L. (1994). *On the Interaction of Theory and Data in Concept Learning*. *18*(2), 221–281. https://doi.org/10.1016/0364-0213(94)90002-7

Wittgenstein, L. (1958). *Philosophical Investigations* (2nd ed.). Basil Blackwell Ltd.

Yee, E., Chrysikou, E. G., & Thompson-Schill, S. L. (2013). Semantic Memory. In K. Ochsner & S. Kosslyn (Eds.), *The Oxford Handbook of Cognitive Neuroscience: Vol. 1: Core Topics*. Oxford University Press. https://yeelab.uconn.edu/wp-content/uploads/sites/1236/2019/02/YeeChrysikouThompsonSchill2013Chapter.pdf

Zadeh, L. A. (1996). Fuzzy sets. In G. J. Klir & B. Yuan (Eds.), *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh* (pp. 394–432). World Scientific. https://doi.org/10.1142/9789814261302_0021

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*(1), 81–88. https://doi.org/10.1080/01638539909545070

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., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

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

List of tables

Chapter 1

Table 1. Illustration of defining features shared between near and far superordinates according to a definitional account.

Chapter 2

Table 1. Experiment 1: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Table 2. Experiment 2: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Table 3. Experiment 3a: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Table 4. Experiment 3b: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Table 5. Experiment 4a: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy for web- and lab-based data, showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Table 6. Experiment 4b: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy for web- and lab-based data, showing change in R^2 for nested comparisons and Bayes Factors for all comparisons.

Chapter 3

Table 1. Analysis 1: Model comparisons for confirmatory and exploratory logistic mixed effects regressions of choice, showing change in R^2 and Bayes Factors for all comparisons.

Table 2. Analysis 1: Estimates and standard errors of choice for sensorimotor-linguistic bias (0 = pro-basic), log word frequency of the chosen and rejected label, and the interaction between sensorimotor-linguistic bias and word frequencies. Derived from the exploratory step 5 model.

Table 3. Analysis 2: Model comparisons for confirmatory and exploratory linear mixed effects regressions of RT (full dataset), showing change in R^2 and Bayes Factors for all comparisons.

Table 4. Analysis 2: Estimates and standard errors of RT for choice (0 = basic, 1 = superordinate), sensorimotor-linguistic bias (0 = pro-basic), log word frequency of the chosen and rejected label, and the interaction between choice and word frequencies. Derived from the exploratory step 9 model.

Table 5. Analysis 2: Model comparisons for confirmatory and exploratory linear mixed effects regressions of RT (dataset split by sensorimotor-linguistic bias condition), showing change in R^2 and Bayes Factors for all comparisons.

Table 6. Analysis 2: Estimates and standard errors of RT in a pro-superordinate dataset for choice (0 = basic, 1 = superordinate), log word frequency of the chosen and rejected label, and the interaction between choice and word frequencies. Derived from the exploratory step 5 model.

Chapter 4

Table 1. Examples of the listed names for artefacts and natural objects, ranging from low to high name agreement and h-statistic.

Table 2. Image-level summary statistics: average name agreement, h-statistic, modal word frequency, modal word length, weighted word frequency, weighted word length and modal production frequency.

Table 3. Comparison of summary statistic to other norming studies.

Table 4. Linear regression coefficients of recognition RT for H-statistic, modal word frequency and length, and weighted word frequency and length.

Chapter 5

Table 1. Bayesian correlations between participant accuracy and average RT

Table 2: Analysis A: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for the comparison of each step against the preceding one.

Table 3. Analysis B: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested

comparisons and Bayes Factors for each step and the previous one (steps 1, 3) and for all interaction models compared to the non-interaction model (Step 3 vs. 4,5,6,7,9,10a-c).

Table 4. Analysis C: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for the comparison of each step against the preceding one.

Table 5. Exploratory taxonomic analyses: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R^2 for nested comparisons and Bayes Factors for Taxonomic levels vs. null, and sensorimotor-linguistic vs. taxonomic levels.

Table 6. Exploratory psycholinguistic analysis D stage one: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R2 for nested comparisons and Bayes Factors for the comparison of each step against the previous best-fitting one.

Table 7. Exploratory psycholinguistic analysis D stage two: Model comparisons for linear mixed effects regressions of RT and logistic mixed effects regressions of accuracy, showing change in R2 for nested comparisons and Bayes Factors for the comparison of each step against the previous one.

Table 8.

Exploratory psycholinguistic analysis E, model comparisons showing change in R-squared and Bayes Factors for the comparison for each step and the previous one (steps 1, 3) and for all interaction models compared to the non-interaction model (Step 3 vs. 4,5,6,7,9,10a-c).

List of Figures

Chapter 1

Figure 1. Schematic representation of the vertical dimension in categorisation, whereby categories range along a dimension of abstraction and inclusiveness.

Figure 2. Schematic representation of the horizontal dimension in categorisation, whereby categories range along a dimension of typicality or exemplariness of the category.

Chapter 2

Figure 1. Bayes factors for models of accuracy and RT, compared to null (random effects) for experiments 1, 3a, and 4a, showing evidence for models containing taxonomic level, sensorimotor distance only and sensorimotor and linguistic distance. Note: best-fitting models are marked with an asterisk. In the case of equivocal performance between two models both models are marked. Experiment 4a denotes the analysis of a combined web- and lab-based dataset (experiments 1 and 3a) rather than data from a new experiment.

Figure 2. Bayes factors for models of accuracy and RT, compared to null (random effects) for experiments 2, 3b, and 4b, showing evidence for models containing taxonomic level only, taxonomic level and object typicality, sensorimotor distance only and sensorimotor and linguistic distance. Note: best-fitting models are marked with an asterisk. In the case of equivocal performance between two models both models are marked. Experiment 4b denotes the analysis of a combined web- and lab-based dataset (experiments 2 and 3b) rather than data from a new experiment.

Chapter 3

Figure 1. Trial design.

Figure 2. Predicted probabilities of choosing basic-level categories by sensorimotor-linguistic bias for high (.5) average (0) and low (-.5) values of centred log word frequency of the chosen and rejected label (with 95% confidence intervals). Dotted line indicates chance threshold.

Figure 2. Predicted RT (step 3 model) by choice (basic, superordinate) and sensorimotor-linguistic bias (pro-basic, pro-superordinate), with 95% confidence intervals.

Figure 3. Predicted RT (step 9 model), for high (.5) average (0) and low (-.5) values (values chosen for illustrative purposes) of centred log word frequency by choice for pro-basic trials with a non-chosen label of average frequency, with 95% confidence intervals.

Figure 4. Predicted RT for pro-superordinate and pro-basic datasets, by taxonomic level chosen, and non-chosen word frequency. With 95% confidence intervals.

Chapter 4

Figure 1. Example of processed images for the category *dog*.

Figure 2. Breakdown of removal of trial-level response data, and corresponding image/name combinations.

Chapter 5

Figure 1. Mask creation process, showing an example of an original image, an image to which a cut-out filter was applied, and the final mask after using the forward warp tool.

Figure 2. Trial-diagram, showing the variable SOA (blank screen), between the rapidly displayed target-image and the perceptual mask. The block label (e.g., *bird*) was only displayed once per block.

Figure 3. Coefficient plots for analysis A, step 7 models of accuracy and RT, displaying estimated change in log-odds of answering correct and RT in ms per Helmert-coded SOA (with error bars showing ±1SE)