

The effects of sensorimotor and linguistic information on the basic-level advantage

Rens van Hoef^{a,*}, Louise Connell^{a,b}, Dermot Lynott^b

^a Department of Psychology, Lancaster University, Lancaster, United Kingdom

^b Department of Psychology, Maynooth University, Maynooth, Ireland

ARTICLE INFO

Keywords:

Sensorimotor simulation
Linguistic distributional knowledge
Conceptual processing
Embodied cognition
Basic-level advantage
Categorisation

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 distributional information (i.e., statistical distribution of words in language) in conceptual processing. In four preregistered word→picture categorisation studies, we examined whether novel measures of sensorimotor and linguistic distance contribute to the basic level-advantage in categorical decision-making. Results showed that overlap in sensorimotor experience between category concept and member concept (e.g., animal→dog) predicted RT and accuracy at least as well as a traditional division into discrete subordinate, basic, and superordinate taxonomic levels. Furthermore, linguistic distributional information contributed to capturing effects of graded category structure where typicality ratings did not. Finally, when image label production frequency was taken into account (i.e., how often people actually produced specific labels for images), linguistic distributional information predicted RT and accuracy above and beyond sensorimotor information. These findings add to our understanding of how sensorimotor-linguistic theories of the conceptual system can explain categorisation behaviour.

1. Introduction

Categorisation is critical to our everyday cognitive functioning. Representative categories and concepts¹ 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 (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 *can fly*), and that categorisation is possible because certain features occur together more frequently than others (e.g., if it *has wings*, *lays eggs* and *can fly*, it is likely a member of the category *bird*; (Cree & McRae, 2003; Hampton, 1993; Malt & Smith, 1984; Posner & Keele, 1968; Rosch, 1973; Smith, Shoben, & Rips, 1974; Tyler, Moss, Durrant-Peatfield, & Levy, 2000). In one popular view, categories are stored in semantic memory through an abstracted summary of how features are shared by category members (prototype theory; Posner & Keele, 1968; Rosch & Mervis, 1975). 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

* Corresponding author at: Department of Psychology, Fylde College, Lancaster University, Bailrigg, Lancaster LA1 4YF, UK.

E-mail address: r.vanhoef@lancaster.ac.uk (R. van Hoef).

¹ 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.

levels to very abstract higher levels. Crucially, while any concept may be categorised at any taxonomic level, the *basic* level (e.g., *bird*) is generally privileged (Rosch, 1978; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). First demonstrated in a series of experiments by Rosch, Simpson, and Miller (1976), the basic-level advantage describes multiple behavioural effects in object recognition, categorisation and naming which show a processing advantage for categories of intermediate abstraction. Most prominent is the finding that people are faster and more accurate at categorising objects which 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 finding is one of the most fundamental effects in categorisation research, and has repeatedly been replicated in subsequent work featuring object typicality (Jolicoeur, Gluck, & Kosslyn, 1984; Murphy & Brownell, 1985), context (Murphy & Wisniewski, 1989), subject expertise (Johnson & Mervis, 1997; Tanaka & Taylor, 1991) and neurological disorders (Rogers & Patterson, 2007).

While the basic-level advantage in object categorisation is a robust effect, its underlying mechanisms have not been conclusively explained. Feature-based 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 (e.g., the feature *lays eggs* is true for all *birds*). Subsequently, Jolicoeur et al. (1984) argue that objects are most easily categorised at the basic level because it is the level at which the taxonomic structure is usually accessed, and so categorisation at a level different to this 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. For instance, 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). Consequently, differentiation accounts suggest an object may 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).

More recently however, 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, Santos, Simons, & Wilson, 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.), though the precise information simulated in a particular instance depends on situational context, task goals, available resources, and participant motivations (e.g., Connell & Lynott, 2014). 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, Wilson, Rizzolatti, & Iacoboni, 2006; Carota, Moseley, & Pulvermüller, 2012; Goldberg, Perfetti, & Schneider, 2006; Hauk, Johnsrude, & Pulvermüller, 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, 2022a 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, McGaw, & Kidd, 2014; Louwerse, 2011). Crucially, while concepts comprise both sensorimotor and linguistic distributional information, linguistic distributional information is computationally cheaper and faster, if less precise, than sensorimotor information (Barsalou et al., 2008; Connell, 2018; Louwerse, 2011). Some debate exists on whether conceptual processing is primarily driven by sensorimotor simulation (e.g., the Language as Simulated Symbols account; Barsalou et al., 2008), or by linguistic distributional information without necessarily requiring sensorimotor activation (e.g., the Symbol Interdependency Hypothesis; Louwerse, 2011), or by either information type depending on a variety of factors including the nature of the task at hand and available processing resources (e.g., the Linguistic Shortcut hypothesis; Connell, 2018; Connell & Lynott, 2014). In the present paper, we take the latter perspective that language may serve as a conceptual shortcut in cases where linguistic association is sufficiently informative to inform decision making.

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. Rather than depending on featural similarity between an object and an abstracted feature summary of its class, category membership may be a product of sensorimotor and linguistic distributional 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 world 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, Connell, Brysbaert, Brand, & Carney, 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, Wingfield, & Connell, 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 (Banks et al., 2021; Connell & Ramscar, 2001; Riordan & Jones, 2011; Wingfield & Connell, 2022a). 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, hierarchical, 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 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 lacking 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 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, 2022a), but feature-concepts are not qualitatively different or subsidiary representations to object-concepts.

Finally, the linguistic-simulation account does not assume that category membership is dependent on similarity to a central tendency or feature abstraction. That is, the category *animal* is not represented through an abstracted feature-summary (i.e., prototype) or salient exemplars, but through the same sensorimotor and linguistic distributional information that other concepts (e.g., *poodle*, *spaniel*) are. Consequently, *poodles* are not categorised as *animals* because they share sufficient features with a prototype, or with familiar poodle exemplars, but because the sensorimotor and linguistic distributional information they activate is similar to that representing the concept *animal*.

The aim of this work is to test whether overlaps in sensorimotor and linguistic distributional experience between member concepts and category concepts can indeed contribute to the behavioural effect of the basic-level advantage in categorisation, without requiring discrete taxonomic levels – regardless of whether these levels are assumed to be explicitly fixed in semantic memory (e.g., entry-level accounts; Jolicoeur et al., 1984) or arise implicitly from the manner in which members' features overlap (e.g., differentiation accounts; Murphy & Brownell, 1985). That is: is categorisation behaviour affected by the sensorimotor and linguistic distributional relationship between a category and member concept? And if so, can it explain the basic-level advantage better than discrete taxonomic levels?

1.1. The current study

In this paper, we report four preregistered experiments (1a, 1b, 2a, 3). We also include a brief discussion of an exploratory analysis (experiment 2b); full details of this study may be found in the supplemental materials on OSF (<https://osf.io/8cjrm>). All studies used a label → picture categorisation task similar to that used by Rosch, Simpson, and Miller (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) more effectively than by discrete taxonomic levels. 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 et al. (2020).² 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).

We expected that sensorimotor and linguistic information would contribute to categorical decision making. In line with the linguistic-shortcut hypothesis (Connell, 2018; Connell & Lynott, 2014), we expected that linguistic information in particular would contribute to the basic level advantage in a category verification task. 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. Experiment 1a: the basic-level advantage

In our first study (preregistration, data, analysis code and results available at <https://osf.io/8cjrm>), 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, Brysbaert, Meyer, & Ghyselinck, 2005). Since the most frequent, earliest-acquired name tends to be the basic-level label (Rosch, Mervis, et al., 1976), we 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 common 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

² 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.

activation (Barsalou et al., 2008; Connell, 2018). Any basic-level advantage would then depend on the extent of overlap in sensorimotor and linguistic distributional experience between the alternative category labels and the picture name.

For example, when a category label (e.g., *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 *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 information alone. We also predicted that continuous sensorimotor and/or linguistic distributional variables 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).

2.1. Method

2.1.1. Participants

Thirty native speakers of English (23 female; $M_{age} = 22$ 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, Wagenmakers, Zehetleitner, & Perugini, 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 $N_{min} = 30$, when analysis A for RT (see Data Analysis) cleared the specified grade of evidence $BF_{10} \geq 3$ (actual $BF_{10} = 2043.09$). 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 preregistered accuracy threshold of 80% correct answers on fillers, 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 preregistration and lower this threshold to 70%; one participant did not pass this new threshold and was replaced.

2.1.2. Materials

Test items consisted of 216 label → picture items, comprising 72 target pictures (depicting natural objects and artefacts in full colour), 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 paired 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 1024×768 pixels. We edited all pictures to display only target objects on a white background (see examples in Fig. 1). All images and labels may be found on OSF (<https://osf.io/8cjrm>).

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 and 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. (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 featured 24 subordinate, 24 basic and 24 superordinate labels and included each picture only once.³

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, to ensure that repeated labels among 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).

2.1.3. Procedure

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. 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 the picture which remained onscreen until a response key was pressed (see Fig. 2). 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 min, including informed consent and debriefing.

2.1.4. Ethics and consent

The study received ethical approval from the Lancaster University Faculty of Science and Technology Research Ethics Committee (ethics code: FST17003). 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.

2.1.5. Critical predictors

As well as a specified taxonomic level (subordinate, basic, superordinate), each label → picture test item had an associated value in two

³ We retrieved Zipf log word frequencies (van Heuven et al., 2014) for all labels except “*sports shirt*”. Average word frequency was lowest for subordinate level labels ($M = 3.10$, $SD = 0.71$) followed by basic-level ($M = 4.57$, $SD = 0.54$) and superordinate-level labels ($M = 4.61$, $SD = 0.44$). Item-level word frequencies may be found in the supplementals on OSF.

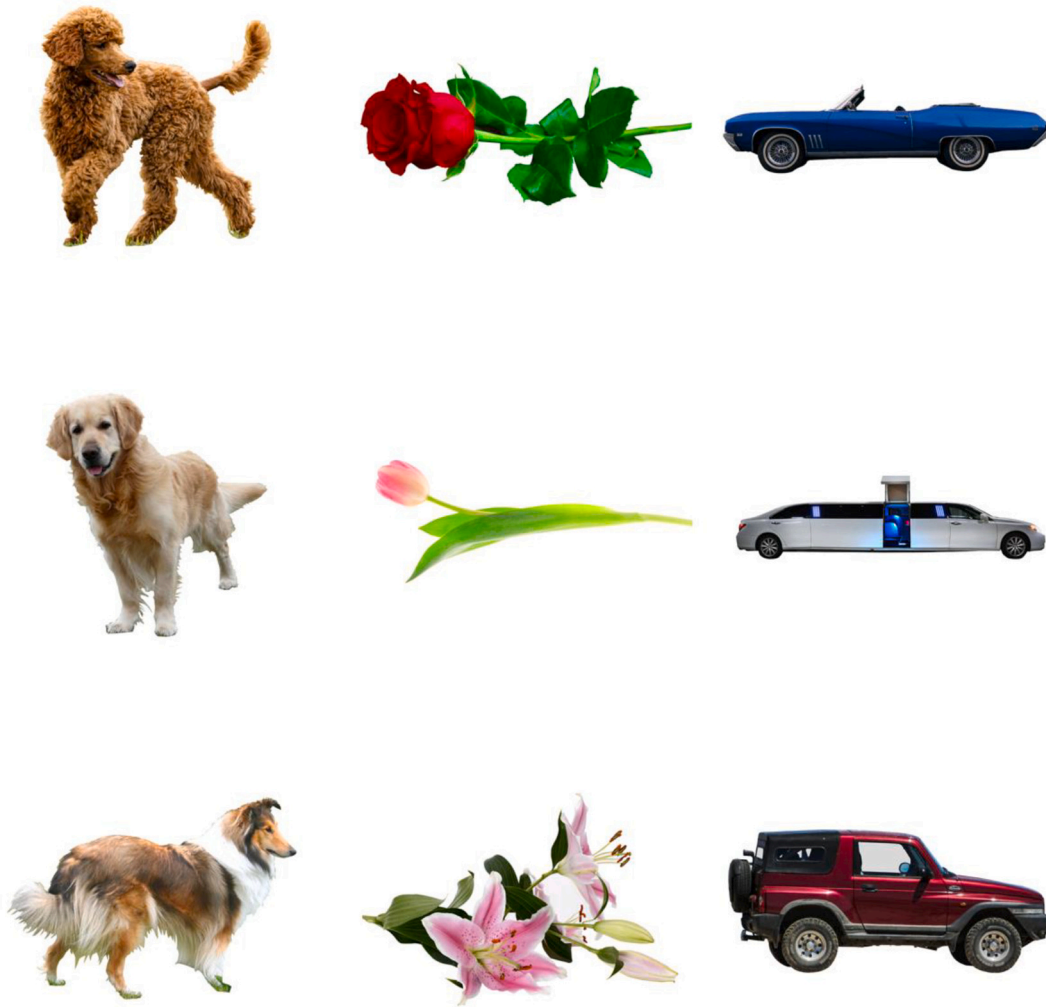


Fig. 1. Examples of “dog”, “flower” and “car” stimuli used in Experiment 1 and 2.

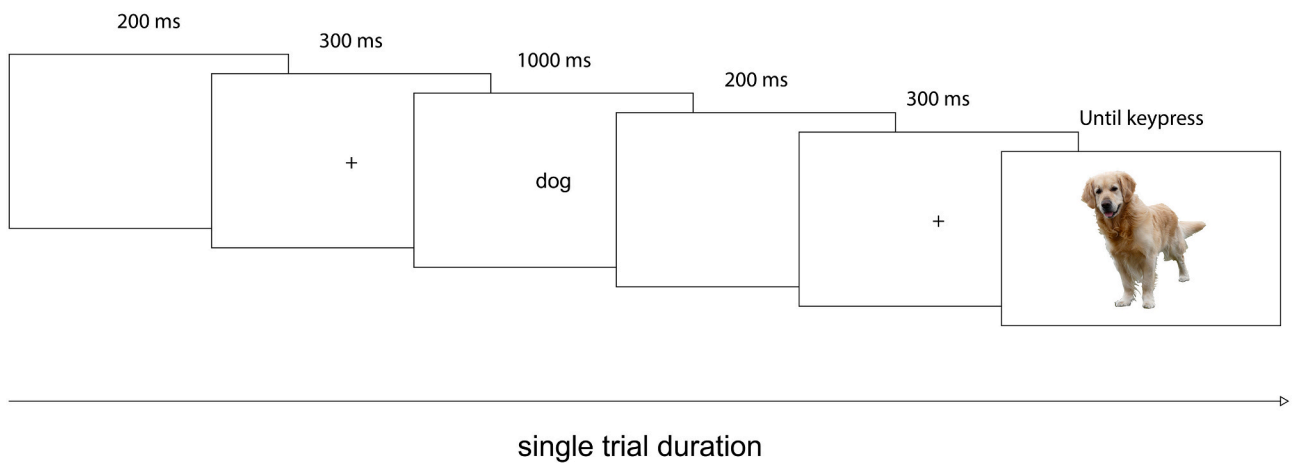


Fig. 2. Trial structure diagram showing trial timings and stimuli as they appeared to participants.

critical predictors that captured the overlap in sensorimotor and linguistic distributional experience between category concept and member concept.

2.1.5.1. *Linguistic distance.* Using a subtitle corpus consisting of 200 million words in British English (see [van Heuven, Mandera, Keuleers, &](#)

[Brysbaert, 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,⁴ 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 (distance = 0.23) than that between two words that appear in very different contexts, such as *dog* and *spaghetti* (distance = 0.46).

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 (e.g., Rosch, Mervis, et al., 1976). In this experiment, 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 = [0.00, 0.83], $M = 0.25$, $SD = 0.21$), with higher values indicating greater distance in linguistic space (i.e., less overlap in the linguistic distributional experience of each word).

2.1.5.2. 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. (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); see Wingfield and Connell (2022b). For example, sensorimotor experience of *dog* and *animal* is more similar than sensorimotor experience of *dog* and *spaghetti*, which is captured by a smaller cosine distance between vectors of sensorimotor experience in the former (0.01) compared to the latter (0.24) example. As for linguistic distance (see above), we calculated sensorimotor distance between sensorimotor vectors for the category label at various taxonomic levels (e.g., *Labrador*, *dog*, *animal*) and the name we assumed would be implicitly associated with the image (e.g., basic-level *dog*).

The final sensorimotor distance measure for each label → picture pair ranged in theory from -1 to +1 (actual range [0.00, 0.29], $M = 0.03$, $SD = 0.05$), 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, $r = 0.38$ (14.4% shared variance).

2.1.6. Data analysis

We planned three sets of analyses to test our hypotheses. All analyses

⁴ For modelling performance in tasks of high conceptual complexity, the use of count-vector (e.g., log co-occurrence) models trained on high-quality corpora is recommended over the use of more complex predict models (e.g., skip-gram) or simpler n-gram models: see Wingfield and Connell (2022a) for a systematic review and analysis.

were run in R (version 4.1.0: R Core Team, 2021) with main packages lme4 (version 1.1–27.1; D. Bates, Maechler, Bolker, & Walker, 2015), lmerTest (version 3.1–3; Kuznetsova, Brockhoff, & Christensen, 2017) MuMIn (version 1.43.17; Bartoń, 2020) and emmeans (version 1.7.0; van Lenth, 2021). A full list of packages used is included in the supplemental materials on OSF.

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 preregistered 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 preregistered 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. In non-nested model comparisons, we 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 null hypothesis significance testing (NHST) statistics of fixed effects in the best-fitting model.

2.2. 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 >2.5SD from the participant's mean. Table 1 shows results of all model comparisons. On average, participants

Table 1

Model comparisons for linear mixed effect regressions of RT and logistic mixed effects regressions of accuracy in Experiment 1a showing change in R² 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)	0.264	–	0.270	–
A	Taxonomic levels vs. null	0.012	2043.09	0.065	49,457.64
B	Sensorimotor distance vs. null	0.011	9521.33	0.016	12.66
	Sensorimotor + Linguistic distance vs. null	0.012	391.87	0.041	300.29
	Sensorimotor + Linguistic distance vs. Sensorimotor-only	0.001	0.04	0.025	23.72
C	Best Sensorimotor-Linguistic model vs. taxonomic levels.	–	4.66	–	0.01

Note: We report conditional model R² for the null model. For all other models, we report marginal (fixed effects only) model R² (see, Nakagawa & Schielzeth, 2013). Note that marginal R² values are estimates reported only for information and are not used for inferencing.

took 764 ms ($SD = 317$ ms) to respond.

2.2.1. 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 participants' RT ($BF_{10} = 2043.09$) and accuracy ($BF_{10} = 49,457.64$). In RT, categorisation decisions made at the subordinate level were 37 ms slower than at the basic level [unstandardised $b = 36.99$, 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 [$b = 83.81$, 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, [$b = -1.01$, 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 [$b = -1.31$, 95% CI = ± 0.44 , $z = -5.83$, $p < .001$] (predicted probability of a correct answer = 91.8%) than at the basic level. Our data thus replicate the classic basic-level advantage in categorisation.⁵

2.2.2. 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 ($BF_{10} = 9521.33$). While a model containing both sensorimotor and linguistic distance was also better than the null ($BF_{10} = 391.87$), model comparisons indicated strong evidence *against* the inclusion of linguistic distance in RT models (i.e., the data were $BF_{01} = 24.30$ 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 (unstandardised $b = 722.97$, 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 ($BF_{10} = 12.66$), but also very strong evidence for the inclusion of linguistic distance alongside sensorimotor distance ($BF_{10} = 300.29$). 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 (unstandardised $b = -3.51$, 95% CI = ± 3.21 , $z = -2.15$, $p = .031$) and as linguistic distance increased ($b = -1.63$, 95% CI = ± 0.82 , $z = -3.87$, $p < .001$) between categories and member concepts. That is, participants were 2.77 times more prone to error for items at the greatest sensorimotor distance (0.29) than for items at the smallest sensorimotor distance (zero). Simultaneously, for items at the greatest linguistic distance (0.83) participants were 4.66 times⁶ more prone to error than for items at the smallest linguistic

⁵ Some subordinate-level labels incorporated the basic-level name (e.g., *steamboat* incorporates *boat*), which one might argue could lead participants to process them more quickly and thus suppress the basic-level advantage over the subordinate level. To check this possibility, we ran an exploratory analysis without these items. Results showed a classic basic-level advantage in accuracy but a *reduced* advantage of the basic-level over the subordinate level in RT, meaning that such items are not responsible for the relatively smaller distinction we observed between basic and subordinate levels. We thank an anonymous reviewer for suggesting this analysis. All data, analyses and results for exploratory analyses may be found on OSF: <https://osf.io/8cjrjm>.

⁶ This value reflects the change in the dependent variable at the maximum sensorimotor (0.29) or linguistic (0.83) distance between categories and members in our dataset, calculated as a proportion of the beta coefficient (e.g., $0.29 \times 722.97 = 209.66$).

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.

2.2.3. Best model

Analysis C showed mixed results as to whether the taxonomic or sensorimotor-linguistic model best explained the data. As predicted, we found positive evidence that sensorimotor distance was $BF_{10} = 4.66$ times better than taxonomic levels in predicting RT. By contrast, and contrary to predictions, there was strong evidence that taxonomic levels were $BF_{01} = 164.70$ times better than sensorimotor and linguistic distance at predicting accuracy. Fig. 3 shows comparisons for each experiment.

2.2.4. 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 did predict response times best, but taxonomic level predicted accuracy best. We address a possible cause in the next experiment.

3. Experiment 1b: modelling categorical gradedness in the basic-level advantage

In experiment 1a, 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 – and such categorical gradedness affects category processing and production (Armstrong, Gleitman, & Gleitman, 1983; Rips, Shoben, & Smith, 1973; Rosch, 1973; Rosch & Mervis, 1975; Rosch, Mervis, et al., 1976; Rosch, Simpson, & Miller, 1976; Smith et al., 1974). That is, categorising typical items that are representative of their category (e.g., *sparrow*, for the category *bird*) tends to be faster than categorising atypical items that are less representative of their category (e.g., *penguin*). However, 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 *penguin* rather than *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 less-representative 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 we

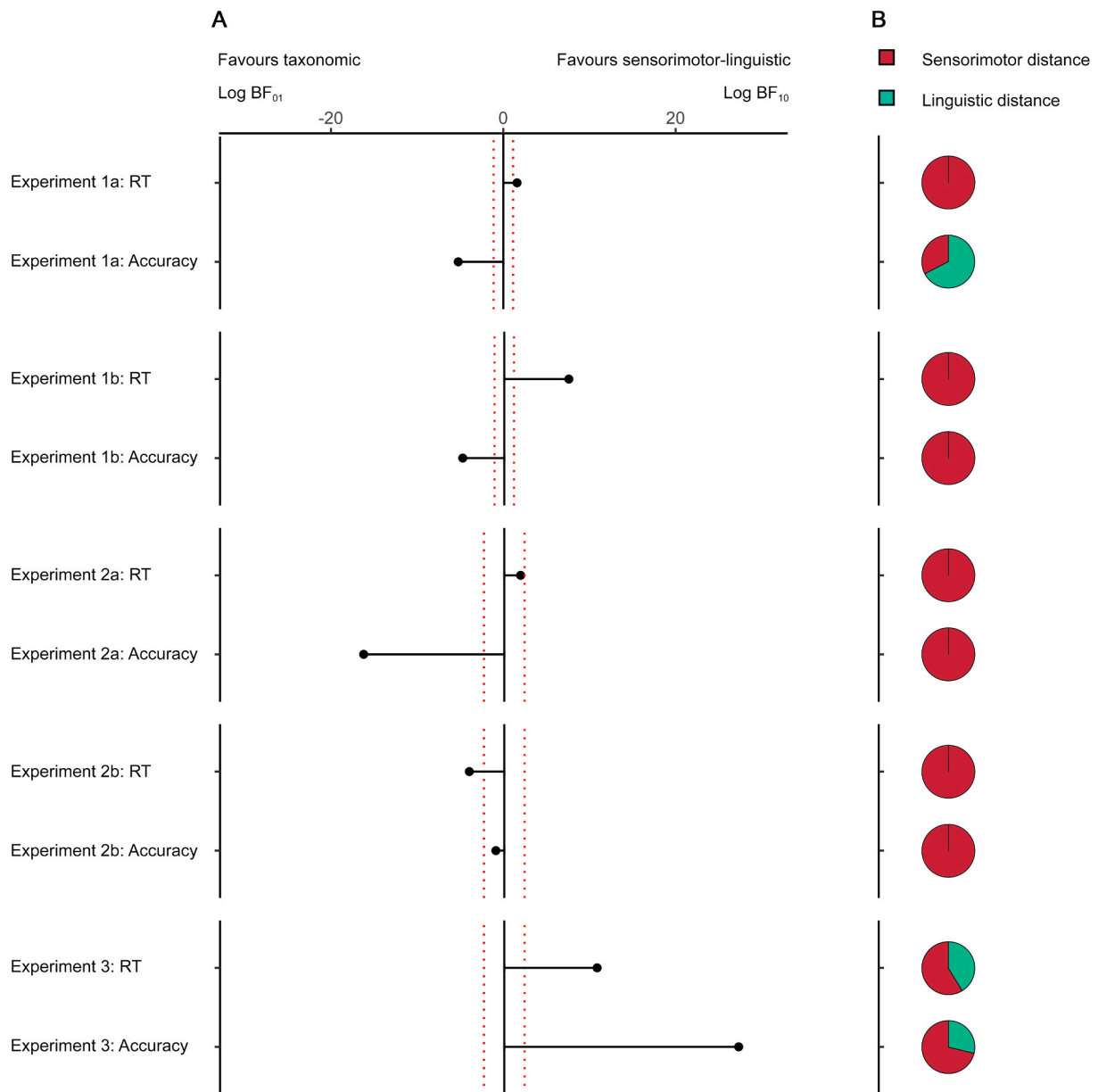


Fig. 3. Log (ln) Bayes Factors for the best sensorimotor-linguistic model versus taxonomic levels for each experiment and dependent variable (Panel A), and relative contribution of sensorimotor and linguistic distributional information to the best-fitting sensorimotor-linguistic model, calculated from standardised regression coefficients (Panel B).

Note: Red dotted line indicates Bayes Factor threshold set for each experiment; evidence falling between dotted lines is equivocal.

calculated distances from *animal* → *dog*. However, for less-representative category members, sensorimotor and linguistic distance should instead be calculated from the category name to the *subordinate* label. If a poodle is a less-representative type of dog, then its picture would be implicitly labelled as *poodle*, and the item *animal* → picture of *poodle* should have distances 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 categoral gradedness by using linguistic distance to determine whether a member concept should be considered a good or poor example of its category. Member concepts close to their category concept (e.g., *salmon* and *fish* appear in very similar linguistic contexts, and have a cosine distance of 0.28) were considered good, highly-representative 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 0.67) were considered poor/less-representative members that would activate subordinate labels and corresponding sensorimotor information. With this adjustment for categoral 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/8cjrm>), we collected traditional 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 categoral decision making using the dataset from Experiment 1a. In line with previous research (e.g., Jolicoeur et al., 1984; Murphy & Brownell, 1985), we expected object 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 traditional 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.

3.1. Method

3.1.1. Materials & dependent measures

We used the categorical decision dataset from Experiment 1a, with new predictors as outlined below.

3.1.2. Critical predictors

As well as a specified taxonomic level (subordinate, basic, superordinate), each label → picture test item had an associated traditional 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 (see below for details).

3.1.2.1. Traditional typicality ratings. We retrieved typicality ratings from 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 in a larger study collecting typicality ratings for 2280 category-member items, where items were divided into lists of 120 items each (Banks and Connell, 2022a); the present 72 category-member items were randomly spread across these lists. Participants rated 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. Consequently, if a participant gave a rating, they were indicating that they were sufficiently familiar with the concept to do so. 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 ($SD = 0.36$, range = [3.42, 5.00]).

3.1.2.2. Gradedness-adjusted linguistic and sensorimotor distance. For the calculation of linguistic distance in Experiment 1a, 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). To incorporate categorical gradedness into linguistic and sensorimotor distance, where pictures of less-representative 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 less-representative 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 less representative of 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 supported by the data by running mixed effect regressions of RT per candidate threshold, with random effects of participant and item and fixed effects of gradedness-adjusted sensorimotor and linguistic distances. Model comparisons showed that the best-fitting model was based on a gradedness-adjusted linguistic distance threshold of 0.33, and that the data favoured this model $BF_{10} = 1908.20$ times more strongly than the original model used in Experiment 1a. In short, a linguistic distance of 0.33 acted as a tipping point between highly representative category members ($N = 23$) that were implicitly named at the basic level and less-representative 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 ($M = 0.30$, $SD = 0.25$, range = [0, 0.83]) and sensorimotor distance ($M = 0.04$, $SD = 0.05$, range = [0, 0.29]) were moderately correlated at $r = 0.53$ (i.e., 28.30% shared variance).

3.1.3. 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 the sensorimotor-linguistic account. Analysis A tested whether including traditional 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 1a, 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.

To test our sensorimotor-linguistic predictions, Analysis C investigated if linguistic distance and typicality ratings 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.

3.2. Results and discussion

Outlier trials were already removed from the dataset in Experiment 1a. 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 ($M = 4.53, SD = 0.45$). While removing this item deviated from the preregistered analysis plan, we felt it was necessary to avoid compromising the robustness of our analyses. Removing responses to this item slightly reduced overall mean RT compared to Experiment 1a ($M = 762$ ms, $SD = 314$ ms). 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 ($BF_{10} = 10,707.19$) and accuracy ($BF_{10} = 82,356.36$). Categorisation at the basic level was faster than at subordinate [unstandardized $b = 42.11, 95\% CI = \pm 29.32, t(1803.90) = 2.81, p = .005$] and superordinate levels [$b = 88.15, 95\% CI = \pm 29.61, t(1809.75) = 5.83, p < .001$]. Similarly, accuracy was greatest at the basic level, followed by subordinate [$b = -1.05, 95\% CI = \pm 0.47, z = -4.43, p < .001$] and superordinate [$b = -1.36, 95\% CI = \pm 0.45, z = -5.87, p < .001$] levels.⁷

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

3.2.1. Taxonomic levels and traditional typicality

In Analysis A, model comparisons showed evidence *against* adding typicality ratings to a model containing taxonomic levels. In RT analysis, the data strongly favoured the taxonomic-only model ($BF_{01} = 40.20$) and typicality had very little effect [$b = 9.48, 95\% CI = \pm 47.36, t(75.05) = 0.39, p = .700$]. In accuracy analysis, the data again favoured the taxonomic-only model ($BF_{01} = 11.29$), with typicality weakly trending in the predicted direction [$b = 0.59, 95\% CI = \pm 0.66, z = 1.76, p = .080$]. 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 & Mervis, 1975; Rosch, Simpson, & Miller, 1976).

3.2.2. Interaction of traditional 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 [unstandardised $b = 46.16, 95\% CI = \pm 75.71, t(1812.60) = 1.19, p = .230$], and data were $BF_{01} = 21.26$ 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 ($BF_{01} = 39.24$), where it had little effect on categorical decision [$b = 0.25, 95\% CI = \pm 0.88, z = 0.56, p = .570$]. 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).

3.2.3. Linguistic distance and traditional typicality

Analysis C found a negative correlation of $r = -0.176$ between our original measure of linguistic distance (i.e., between subordinate items and their basic-level label) and average typicality ratings, with $BF_{10} = 0.79$ representing only an equivocal level of evidence. That is, linguistic distance did not decisively capture the graded structure of categories normally reflected in typicality ratings, where more atypical examples of a category were only weakly associated with greater linguistic distance (i.e., less overlap in linguistic contexts) between category and member

⁷ Full coefficient statistics for all models including and excluding the *gavel* item are included in supplementals (see also Appendix A). In brief, analysis including this outlier item did not affect inferences based on Bayesian model comparisons, but it 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 1b showing change in conditional and marginal R^2 for nested comparisons and Bayes Factors for all comparisons.

Analysis	Model comparison	RT		Accuracy	
		ΔR^2	BF_{10}	ΔR^2	BF_{10}
A	Null model (random effects)	0.262	–	0.261	–
	Taxonomic levels vs. null	0.013	10,707.19	0.070	82,356.36
	Taxonomic levels + Typicality vs. null	0.013	266.36	0.079	7291.12
	Taxonomic levels + Typicality vs. null	<0.001	0.02	<0.001	0.09
B	Taxonomic levels + Typicality + Interaction vs. null	0.014	12.53	0.079	185.78
	Taxonomic levels + Typicality + Interaction vs. null	<0.001	0.05	<0.001	0.02
	Taxonomic levels + Typicality	0.018	31,398,739.40	0.027	752.57
D	Sensorimotor distance vs. null	0.017	2,075,342.90	0.033	133.21
	Graded-Sensorimotor + Graded-Linguistic distance vs. null	<0.001	0.07	0.005	0.18
E	Graded-Sensorimotor + Graded-Linguistic distance vs. Graded-Sensorimotor only	–	118,155.82	–	0.10

Note: We report conditional model R^2 for the null model. For all other models, we report change in marginal (fixed effects only) model R^2 (see Nakagawa & Schielzeth, 2013). Note that marginal R^2 values are estimates reported only for information and are not used for inferencing.

concept.

3.2.4. 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, $BF_{10} = 31,398,739.40$. 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 1a, the data were $BF_{01} = 15.13$ 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), [unstandardised $b = 842.68$, 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, $BF_{10} = 752.52$. However – and unlike Experiment 1a – we also found evidence *against* the inclusion of gradedness-adjusted linguistic distance, whereby the data favoured the model containing sensorimotor distance alone ($BF_{01} = 5.65$). 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 more likely than the shortest distance [$b = -6.66$, 95% CI = ± 2.70 , $z = -4.82$, $p < .001$].

3.2.5. Best model

In analysis E, results followed the mixed pattern of Experiment 1a regarding whether the data were more likely under a model using traditional taxonomic levels with typicality ratings or sensorimotor/linguistic distance (see Fig. 3). 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 $BF_{10} = 118,155.82$ times better at predicting RT than a model containing taxonomic levels and traditional typicality. However, against predictions but consistent with Experiment 1a, there was evidence that the taxonomic-typicality model was $BF_{01} = 9.69$ times better (i.e., $BF_{10} = 0.10$) at predicting accuracy than a model containing gradedness-adjusted sensorimotor distance.⁸

3.2.6. 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 reflected in typicality ratings. While atypical member concepts were slightly 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), the evidence for a systematic relationship was equivocal.

Nevertheless, adjusting for categorical gradedness improved the predictive ability of sensorimotor and linguistic distance measures compared to the unadjusted measures of Experiment 1a. 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. That is, pictures of good, highly representative category members were implicitly named and processed at the basic level (e.g., *fish* → *salmon* processed as *fish* → *fish*) while pictures of less-representative category members were implicitly named and processed at the specific, subordinate level (e.g., *fish* → *sailfish* processed as *fish* → *sailfish*),

⁸ As an exploratory analysis, since typicality performed so poorly as a predictor, we opted to compare the gradedness-adjusted sensorimotor distance model against the taxonomic-only model (i.e., without typicality ratings). Results remained unchanged: sensorimotor distance was the best model of RT ($BF_{10} = 2932.49$ times better than the taxonomic model) and taxonomic levels were the best model of accuracy ($BF_{01} = 109.43$ times better than the sensorimotor model).

and in each case the sensorimotor overlap between the retrieved category and member concept affected latency and accuracy of performance.

In addition, the analysis with typicality produced some unexpected results. Contrary to previous findings in the literature (Rosch & Mervis, 1975; Rosch, Simpson, & Miller, 1976), a traditional measure of categorical gradedness, typicality ratings, 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. Many of the least typical items in our dataset tended to be categorised at the basic rather than subordinate level (e.g., a picture of a *fir* was categorised as a *tree* more quickly and accurately than as a *fir*; a picture of a *chalice* was categorised as a *cup* more quickly and accurately than as a *chalice*) and many of the *most* typical items tended to be categorised at the subordinate level rather than basic (e.g., a picture of an eagle was categorised as an *eagle* more quickly and accurately than as a *bird*; a picture of a *rose* was categorised as a *rose* more quickly and accurately than as a *flower*). 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 less-representative 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 1a, and gradedness-adjusted sensorimotor distance outperformed taxonomic-typical models in predicting RT. Such a pattern of findings suggests that subjective typicality ratings 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.

4. Experiment 2a: replication of Experiment 1a

To replicate the effects found in Experiment 1a (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: we expected that sensorimotor and linguistic distributional information would inform categorisation, that linguistic distributional information would contribute above and beyond sensorimotor information alone, and that a combination of sensorimotor and/or linguistic distributional information would explain categorisation performance better than traditional taxonomic levels (i.e., subordinate, basic, superordinate). Preregistration, data, analysis code, and results are available at <https://osf.io/8cjrj>.

4.1. Method

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

4.1.1. Participants

25 participants (21 female, $M_{\text{age}}, 37.76$, $SD = 10.86$) were recruited through Prolific.co (formerly Prolific.ac), an online crowdsourcing platform, for the sum of £1.75 (i.e., approx. £7 /hour pro rata for an assumed duration of 15 min). On average, participants took 12 min and 44 s to complete the task ($SD = 6$ min and 18 s). 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 1a. Since we stopped sampling at our lower bound of $N = 30$ in Experiment 1a, we used sequential analysis of Experiment 1a's data to determine the number of 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 $N = 14$ ($BF_{10} = 71.88$). Several studies (Crump, McDonnell, & Gureckis, 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. To allow for a higher level of noise in our dataset, we therefore set our present lower bound to be 50% higher at $N_{min} = 21$ and raised our threshold of evidence to a more conservative $BF_{10} \geq 10$. Due to a technical error, an additional 4 participants above N_{min} 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 $BF_{10} = 14,847,327.02$.

4.1.2. Procedure

The experiment ran on web-based platform Gorilla.sc (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020), which handled both collection of informed consent and experimental data collection. Trial presentation was identical to Experiment 1a, except the font used to display labels was lowercase Open Sans.

4.1.3. Ethics and consent

The study received ethical approval from the Lancaster University Faculty of Science and Technology Research Ethics Committee (ethics code: FST17003). As well as the terms specified in Experiment 1a, 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 1a).

4.1.4. Data analysis

We repeated analyses A through C as specified in Experiment 1a.

4.2. Results and discussion

One participant had a very long tail of slow RTs, which indicated inattention, but was not otherwise excluded by the preregistered criteria for outlier removal. As a result, we decided to remove all trials with $RT > 10,000$ ms: 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 >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). On average, participants took 811 ms ($SD = 280$ ms) to respond.

Table 3 shows model comparisons for all analyses.

4.2.1. 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 $b = 35.91$, 95% CI = ± 28.94 , $t(1528.67) = 2.43$, $p = .015$] and superordinate [$b = 105.85$, 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 1a, 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 [$b = -1.52$, 95% CI = ± 0.53 , $z = -5.67$, $p < .001$], as expected. However, people were 1.49 times more likely to respond *correctly* at the subordinate level than

Table 3

Model comparisons for linear mixed effect regressions of RT and logistic mixed effects regressions of accuracy in Experiment 2a 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}
A	Null model (random effects)	0.197	–	0.262	–
	Taxonomic levels vs. null	0.024	14,847,327.02	0.129	9.81×10^{10}
B	Sensorimotor distance vs. null	0.022	92,484,998.15	0.048	12,408.53
	Sensorimotor + Linguistic distance vs. null	0.024	26,398,821.53	0.049	314.23
	Sensorimotor + Linguistic distance vs. Sensorimotor-only	0.002	0.28	0.001	0.02
C	Best Sensorimotor-Linguistic model vs. taxonomic levels.	–	6.23	–	<0.01

Note: We report conditional model R^2 for the null model. For all other models, we report change in marginal (fixed effects only) model R^2 (see, Nakagawa & Schielzeth, 2013). Note that marginal R^2 values are estimates reported only for information and are not used for inferencing.

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 [$b = 0.41$, 95% CI = ± 0.69 , $z = 1.16$, $p = .245$]. 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.

4.2.2. Sensorimotor-linguistic predictors

In RT, Analysis B model comparisons showed overwhelming evidence for the effect of sensorimotor distance over the null model ($BF_{10} = 92,484,998.15$), where RT increased by up to 271 ms for the largest sensorimotor distance compared to the smallest one [$b = 935.20$, 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 $BF_{01} = 3.50$ 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 (i.e., the candidate model with fewer parameters; replicating Experiment 1a) 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 1a – there was strong evidence against the inclusion of linguistic distance alongside sensorimotor distance ($BF_{01} = 39.49$). That is, the best sensorimotor-linguistic model of accuracy was sensorimotor distance alone. People were up to 15.47 times⁶ more likely to respond incorrectly as sensorimotor distance increased [$b = -9.44$, 95% CI = ± 3.58 , $z = -5.17$, $p < .001$].

As predicted, and replicating Experiment 1a, 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.

4.2.3. Best model

Analysis C again showed mixed results as to which model best explained the data. Unlike Experiment 1a, where sensorimotor distance outperformed taxonomic levels in explaining RT, model comparisons in the present analysis showed they performed approximately equivalently (see Figs. 3 and 4). The data favoured the model with sensorimotor distance $BF_{10} = 6.23$ 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 $BF_{01} = 7,902,660.94$ times better at fitting the data than sensorimotor distance, against predictions but replicating Experiment 1a.

4.2.4. Summary

These results are similar but not identical to the findings of Experiment 1a (see Figs. 3 and 4). Nevertheless, they provide further evidence for the effects 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 1a and 2a yielded slightly different results (see Figs. 3 and 4). Specifically, contrasting the findings in Experiment 1a, linguistic distributional distance did not predict accuracy over sensorimotor distance. Moreover, the higher Bayes Factor threshold used in Experiment 2a meant that, while the overall pattern was similar to that observed in Experiment 1a (i.e., strongest evidence for the sensorimotor-only model) evidence for the sensorimotor over taxonomic model was in the equivocal zone ($BF_{10} < 10$).

4.2.5. Experiment 2b

As an exploratory analysis, we examined whether accounting for categorical gradedness (i.e., goodness of membership) affected model fit. We replicated the analysis of Experiment 1b on the Experiment 2a dataset by including measures of object typicality and gradedness-adjusted sensorimotor and linguistic distributional distance. A full description of the method, results and discussion may be found on OSF (<https://osf.io/8cjrjm>)

Results of this exploratory replication were similar but not identical to the findings of Experiment 1b (see Fig. 4). As previously found, although contrary to previous findings in the literature, typicality ratings did not affect categorical decisions (e.g., Rosch, Simpson, & Miller, 1976) nor impact on the level of categorisation (i.e., where atypical members are categorised at the subordinate level: Jolicoeur et al., 1984; Murphy & Brownell, 1985). As predicted, however, gradedness-adjusted sensorimotor distance contributed to both RT and accuracy. Sensorimotor overlap between a category and member concept was a very strong predictor of categorical decision for “good” category members whose pictures were implicitly processed with high-frequency basic labels (e.g., *fish* → *salmon* processed as *fish* → *fish*) and less-representative category members whose pictures were implicitly processed with specific labels (e.g., *fish* → *sailfish* processed as *fish* → *sailfish*). However, counter to our predictions, linguistic distance did not affect performance once sensorimotor distance had been taken into account. These findings are all consistent with Experiment 1b and show that categorical gradedness – as modelled by linguistic distance but not typicality ratings – mediates the basic level advantage in categorical decision making.

Different to Experiment 1b, however, was the best model. Gradedness-adjusted sensorimotor and taxonomic-typicality models performed equally well in explaining both accuracy and RT (see Fig. 4). When we dropped typicality from the taxonomic model due to its lack of effect, accuracy was explained equally well by taxonomic levels and

sensorimotor distance (as opposed to taxonomic levels alone in Experiment 1b), while RT was best explained by taxonomic levels (as opposed to sensorimotor distance in Experiment 1b). Taken together, the results of exploratory Experiment 1a suggest that sensorimotor-linguistic information (i.e., sensorimotor distance categorically graded by linguistic distance) predicts the basic level advantage about as well as discrete taxonomic levels, with the apparent primacy of one model over another varying by participant sample.

5. Experiment 3: categorical gradedness from normed object names

Experiments 1a and 2a were built on the assumption that images of objects would implicitly activate a basic-level name in participants, following research showing that participants most frequently use labels of intermediate abstraction when asked to name objects (Rosch, Mervis, et al., 1976; see also Murphy & Smith, 1982). However, as Experiments 1b and 2b (see supplementary materials on OSF) show, participants’ categorisation behaviour may be better predicted by models that do not rigidly assume the implicit image name is always at the basic level, and adjustment for categorical gradedness is required to capture the fact that some images are named at a more specific, subordinate level. Furthermore, picture naming research suggests that perfect name agreement (i.e., a single name given to an image by all participants) is rarely observed (e.g., only 24 out of 1468 items in the BOSS norms; Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Brodeur, Guérard, & Bouras, 2014) and the most frequently-given name for an object might not be the one assumed by experimenters (e.g., Brodeur and colleagues found that most participants labelled an image of an alligator as a *crocodile*). In short, people do not always label objects as one might expect, which may explain some of the variability in linguistic distributional versus sensorimotor effect sizes in previous experiments.

In this Experiment, therefore, we investigated the effects of sensorimotor and/or linguistic distributional information on category verification using a set of pictures for which the range of associated names was *known*. Specifically, we derived category and object names from a recent set of picture-naming norms that provided both the list of names participants used to label each image, and the frequency with which each name was produced (van Hoef, Lynott, & Connell, 2022). Rather than assuming a single basic- or subordinate-level image label when calculating sensorimotor and linguistic distributional distances for a word → picture pair, we instead incorporated *all* labels people are likely to give an image by averaging their distances, weighted by production frequency. For example, rather than assuming the item *dog* → [picture of a *poodle*] should be treated as *dog* → *dog* (Experiment 1a and 2a) or *dog* → *poodle* (Experiment 1b and 2b), we treated it according to the normed labels for that poodle image: a weighted average of 67% *dog* → *dog* and 33% *dog* → *poodle*. In this way, the graded structure of categories was inherently reflected in the production frequency used to weight the distance calculations, as some member concepts were named at the basic level more often than others (e.g., an image of a Labrador was named as a *dog* more often than was an image of a poodle).

As before, we hypothesised that both sensorimotor and linguistic distributional information would inform categorisation, and that the basic-level advantage would emerge from representational overlap of linguistic and sensorimotor information between a category and member concept. However, we updated our predictions based on our findings in previous experiments. That is, we predicted that weighted average sensorimotor distance would independently predict RT and accuracy, and that weighted average linguistic distributional distance would predict accuracy (but not RT) above and beyond sensorimotor information and do so at least as well as sensorimotor distance. Furthermore, in line with the previous experiments, we predicted that sensorimotor distance alone would predict the basic-level advantage in RT at least as well as taxonomic levels, whereas sensorimotor and linguistic distributional information would not predict accuracy as well as taxonomic

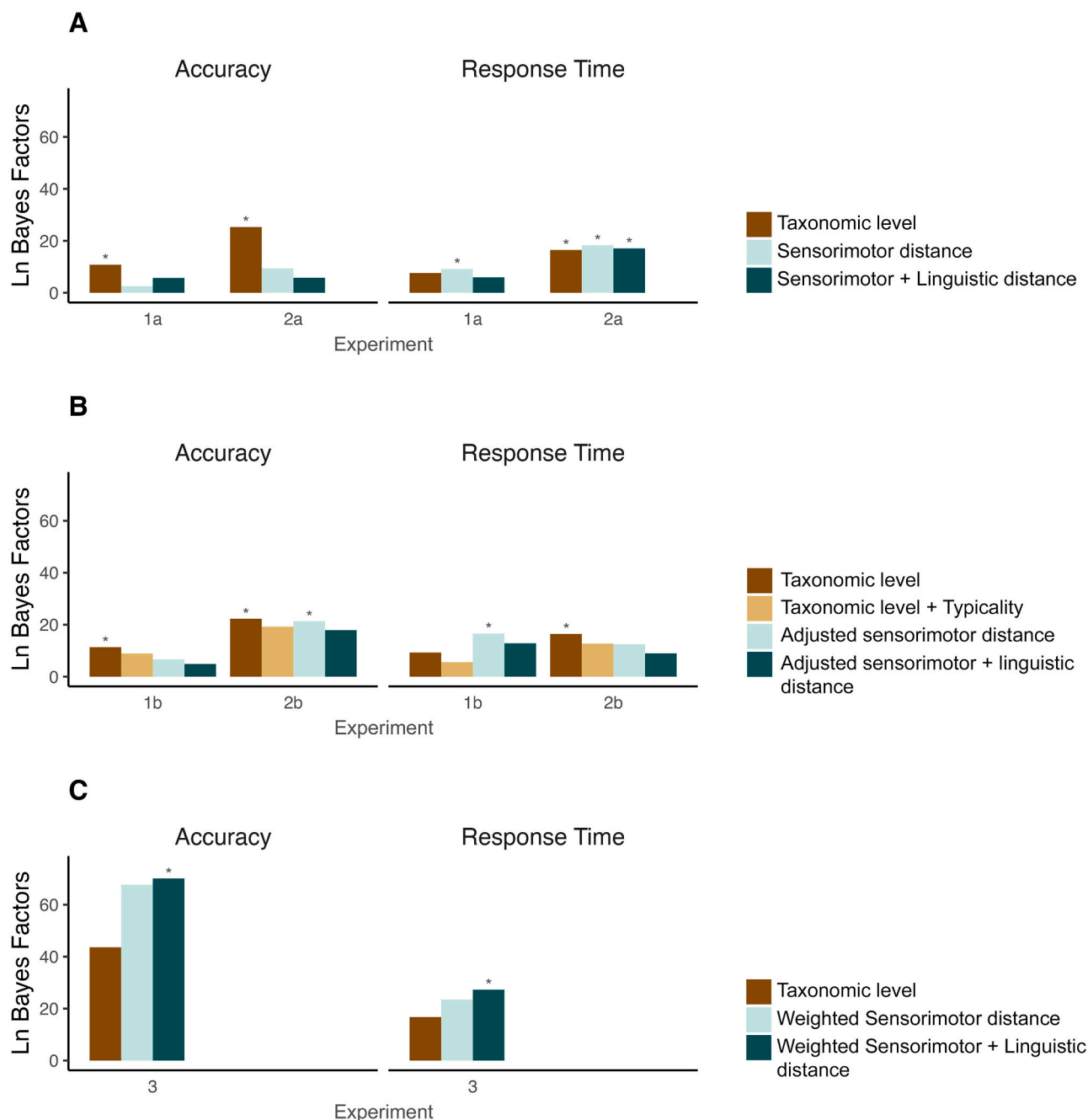


Fig. 4. Log (ln) Bayes Factors for all models for each experiment and dependent variable for unadjusted (Panel A) gradedness-adjusted (Panel B) and production-frequency weighted models (Panel C) compared to a null model containing only random effects of participant and item. *Note:* Asterisks indicate the best-fitting models overall; where bars show multiple asterisks, evidence for models was equivocal.

levels. Preregistration, data, analysis code, stimuli and results are available at <https://osf.io/8cjrm>.

5.1. Method

5.1.1. Participants

Forty-three participants (30 female; $M_{age} = 34.58$ years, $SD = 11.62$) were recruited through web-based crowdsourcing platform Prolific.co, for the sum of £3.55 (i.e., approx. £8.50/ h pro rata for an assumed duration of 25 min). On average, participants took 18 min and 44 s to complete the task ($SD = 4$ min and 20 s), which included giving informed consent and reading a debriefing. Through Prolific’s recruitment filter settings, we ensured that all participants were native

speakers of English, had corrected-to-normal vision, and had not participated in any of our previous web-based experiments. One participant was replaced based on the 70% accuracy threshold specified in the preregistration.

As before, we determined sample size via sequential hypothesis testing using Bayes Factors (Schönbrodt et al., 2017), with N_{min} set at 30 participants, and N_{max} set at 90. We stopped sampling at 43 participants, when hierarchical effects of both sensorimotor distance and linguistic distributional distance in Analysis B cleared the specified grade of evidence of $BF > 10$ (or reciprocal 1/10) for three successive participants for both accuracy and RT (see Results section for actual BFs).

5.1.2. Materials

Test items consisted of 396 label → picture items, comprising 132 target pictures, each of which was paired with three labels that described it at the subordinate, basic and superordinate level. All images were retrieved from the van Hoef et al. (2022) picture naming norms, and their basic- and subordinate-level names were derived from participants' naming responses in these norms. First, we determined basic-level category names, by selecting only those items for which the modal response occurred for the majority of object images as well as for more than one object (e.g., *dog* as the modal name for most *Labrador* and *Spaniel* images). Crucially, we ensured that every image had at least one alternative name that was more specific than the modal response (e.g., *Labrador* for Labrador images). Where more than one alternative name was produced for an item, we used the most frequent alternative as the subordinate-level label in our task. Since abstracted, superordinate-level labels were produced only occasionally in the picture-naming norms, we retrieved superordinate-level category names from WordNet (<https://wordnet.princeton.edu>), a lexical database that includes a hypernymic (i.e., type-of) taxonomy for each sense of a word. For each basic-level name, we determined the set of WordNet hypernyms for the relevant word sense (e.g., for the canine sense of *dog* rather than a figurative meaning), and selected the most easily-comprehensible abstracted hypernym as the superordinate-level label (e.g., for *dog*, we selected the hypernym *animal* rather than more technical alternatives like *mammal* or *chordate*). To create label → picture items, each subordinate-level label was then paired with three images (e.g., *Labrador* → three different images of a Labrador), each basic-level label was paired with between 6 and 27 images (e.g., *dog* → three images of a Labrador, three images of a Spaniel, three images of a collie, etc.) and each superordinate-level label was paired with between 27 and 39 images (e.g., *animal* → three images of a Labrador, three images of a Spaniel, three images of a Lynx etc.). Finally, we divided all 396 test items into three stimulus lists of 132 items, where each list featured 44 subordinate, 9 basic and 4 superordinate labels and included each picture only once.

Filler items consisted of 174 label → picture pairs, comprising similar object pictures and labels to test items, and were seen by all participants. Of these, 48 fillers were false and used category labels featured in test-items, to ensure that repeated labels could not cue participants to respond 'yes' to category membership: 24 basic level (e.g., label *dog* → image *cow*) and 24 superordinate (e.g., label *vehicle* → image *clock*). A further 105 fillers were false and used category labels *not* featured in test items: 27 superordinate (e.g., label *building* → picture *ashtray*), 27 basic (e.g., label *fish* → image *frog*), and 51 subordinate (e.g., label *duck* → image *cockatoo*) labels. All false fillers ranged from easy to difficult to reject. Finally, to balance out the true and false items per stimulus list, we included 21 fillers that were true and did not repeat test-item category labels: 7 basic (e.g., label *saw* → image *jigsaw*), 7 subordinate (e.g., label *kingfisher* → image *kingfisher*), and 7 superordinate (e.g., label *device* → image *keyboard*) labels. As a result, the final stimulus lists each contained 306 label → picture pairs, divided evenly between true and false (132 true test items, 21 true fillers and 153 false fillers). All stimuli may be found in the supplemental materials on OSF.

5.1.3. Procedure

The experiment ran on web-based platform Gorilla.sc (Anwyl-Irvine et al., 2020), handling both collection of informed consent and experimental data collection. Trial presentation and instructions were identical to Experiment 2a, with the exception that the font used to display labels was 36-point lowercase Arial. Participants were randomly assigned to a stimulus list. Test and filler items appeared in random order with a self-paced break after 160 trials.

5.1.4. Ethics and consent

The study received ethical approval from the Lancaster University Faculty of Science and Technology Ethics Committee (ethics code:

FST17003). As per Experiment 1a, all participants read information detailing the purpose and expectations of the study before giving informed consent to take part.

5.1.5. 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.

5.1.5.1. Weighted average linguistic distance. For each image in the item set, we retrieved the set of names produced by participants in a recent picture naming norms, (van Hoef et al., 2022) as well as the associated frequency of production for each name. We then calculated the cosine distance between each category label and each image name, using the same corpus-based measure of linguistic distance as per Experiment 1a. Finally, for every label → picture item, we calculated its weighted average linguistic distance by multiplying each label → name distance by the relevant production frequency of that name for that picture and calculated the mean of these weighted distances. For example, the weighted average linguistic distance for the item *dog* → [Labrador image 1] was calculated as the *dog-dog* distance (0) multiplied by its production frequency weight (81.0%) plus the *dog-Labrador* distance (0.519) multiplied by its production frequency weight (19%), giving a weighted average distance of 0.099. The final weighted average linguistic distance measure for each label → picture pair ranged in theory from -1 to $+1$ (actual range = $[0.00, 0.91]$, $M = 0.30$, $SD = 0.15$), with higher values indicating greater distance in linguistic space (i.e., less overlap in the linguistic distributional experience of each word).

5.1.5.2. Weighted average sensorimotor distance. As for linguistic distance (see above), we calculated the weighted average sensorimotor distance for each item based on the different names produced in the van Hoef et al. (2022) picture naming norms, weighted by their production frequencies. The final weighted average sensorimotor distance measure for each label → picture pair ranged in theory from -1 to $+1$ (actual range $[0.00, 0.16]$, $M = 0.03$, $SD = 0.03$), 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 weakly-to-moderately correlated, $r = 0.311$ (9.67% shared variance).

5.1.6. Data analysis

We ran analyses A and B as specified in Experiment 1a, except we used weighted average sensorimotor distance and weighted average linguistic distributional distance in place of the original distance measures. Analysis C tested whether RT and accuracy were best explained by traditional taxonomic levels or by sensorimotor(–linguistic) information. In non-nested model comparisons, we tested whether the RT data favoured the sensorimotor-only model from Analysis B over the taxonomic model from Analysis A, and whether the accuracy data favoured the sensorimotor-linguistic model from Analysis B over the taxonomic model from Analysis A.

For each analysis, we report the coefficients and null hypothesis significance testing (NHST) statistics of fixed effects in the best-fitting model.

5.2. Results and discussion

In total, we collected data for 5938 trials. We removed 6 trials due to motor error ($RT < 200$ ms) and 21 trials due to inattention ($RT > 5000$ ms); no participants had a mean RT further than 3 SD away from the overall mean. We removed 188 trials from the analysis of accuracy (3.18% from 5911 observations), and 158 trials from the analysis of RT

(2.93% of 5397 correct observations) for having an RT further than 2.5SD away from the participant mean. The final accuracy dataset consisted of 5723 correct and incorrect trials; the final RT dataset consisted of 5239 correct trials. On average, participants took 786 (*SD* = 301) ms to respond.

5.2.1. 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 on participants' RT ($BF_{10} = 19,262,614.75$) and accuracy ($BF_{10} = 8.88 \times 10^{18}$). In RT, categorisation decisions made at the subordinate level were 32.18 ms slower than at the basic level [unstandardized $b = 32.18$, 95% CI = ± 14.60 , $t(5086.31) = 4.32$, $p < .001$]. Furthermore, categorisation decisions at the superordinate level were 53.37 ms slower than at the basic level [$b = 53.37$, 95% CI = ± 14.82 , $t(5088.95) = 7.06$, $p < .001$].

Accuracy was comparable to previous experiments, suggesting the new stimulus set in the present study was of equivalent difficulty. Participants responded most accurately to an image when it was preceded by a basic-level category label (predicted probability of a correct answer = 97.6%). Compared to this basic level, participants were 1.71 times more likely to respond incorrectly when an image was labelled at the subordinate level, [$b = -0.54$, 95% CI = ± 0.28 , $z = -3.69$, $p < .001$] (predicted probability of a correct answer = 96.0%). Finally, participants were up to 3.69 times more likely to respond incorrectly at the superordinate level than at the basic level [$b = -1.30$, 95% CI = ± 0.26 , $z = -9.67$, $p < .001$] (predicted probability of a correct answer = 91.8%). This data thus replicates the classic basic-level advantage, with best categorisation accuracy at the basic level and worst at the superordinate level, in line with Experiments 1a. Table 4 shows all model comparisons.

5.2.2. Sensorimotor-linguistic predictors

Confirming our hypotheses, in RT, Analysis B model comparisons showed very strong evidence for the effect of weighted average sensorimotor distance over a null model containing only random effects ($BF_{10} = 16,023,718,859.37$). However, contrasting our hypotheses, model comparisons also showed strong evidence for the inclusion of weighted average linguistic distributional on top of sensorimotor distance ($BF_{10} = 43.78$). Inspection of the coefficients revealed that RT increased with sensorimotor distance [unstandardised $b = 831.05$, 95% CI = ± 364.55 , t

(1158.64) = 4.47, $p < .001$] as well as with linguistic distance [$b = 105.49$, 95% CI = ± 51.43 , $t(3363.21) = 4.02$, $p < .001$]. That is, although we originally hypothesised in Experiments 1a and 2a that linguistic distance would affect RT, the results of these studies led us to update our hypotheses and predict a null linguistic effect in the present experiment. However, the linguistic effect on RT did indeed emerge, against our preregistered hypothesis, and instead consistent with the joint sensorimotor-linguistic effects we had expected to find in Experiments 1a and 2a. In other words, participants were up to 132.97 ms slower at the greatest average sensorimotor distance (0.16), and up to 95.99 ms slower at the greatest average linguistic distributional distance (0.91) between the category label and image name(s).

In accuracy, there was overwhelming evidence for the effect of sensorimotor distance alone ($BF_{10} = 2.62 \times 10^{29}$), and positive evidence for the inclusion of linguistic distance alongside sensorimotor distance ($BF_{10} = 10.99$). 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 ($b = -24.55$, 95% CI = ± 5.27 , $z = -9.12$, $p < .001$) and as linguistic distance increased ($b = -1.85$, 95% CI = ± 0.96 , $z = -3.77$, $p < .001$) between categories and member concepts. That is, participants were 50.80 times more prone to error for items at the greatest average sensorimotor distance (0.16) than for items at the smallest sensorimotor distance (zero). Simultaneously, for items at the greatest linguistic distance (0.91) participants were 5.38 times more prone to error than for items at the smallest linguistic distance (zero).

As predicted, overlap in sensorimotor experience between category and member concepts (e.g., between *animal* and *dog* concept or *animal* and *Labrador* concept, depending on how an individual participant implicitly names an image of a Labrador) facilitates categorisation RT and accuracy. Similarly, as predicted, overlap in linguistic experience between the distributional patterns of category and member labels also facilitates categorisation accuracy. Indeed, more than we expected in this study, overlap in linguistic distributional experience also facilitates categorisation RT; an effect that failed to emerge as predicted in Experiments 1a and 2a.

5.2.3. Best model

Analysis C tested whether taxonomic or sensorimotor information best predicted RT, and whether taxonomic or sensorimotor-linguistic best predicted accuracy. As predicted, Bayesian model comparisons found very strong evidence that sensorimotor distance was $BF_{10} = 831.87$ times better than taxonomic levels in predicting RT. As a model including linguistic distributional distance on top of sensorimotor distance outperformed a sensorimotor-only model in predicting RT, we ran an exploratory (i.e., not preregistered) comparison between this sensorimotor-linguistic model and taxonomic levels, and found overwhelming evidence that sensorimotor and linguistic distance were $BF_{10} = 36,317.40$ times better than taxonomic levels in predicting RT. Moreover, Bayesian model comparisons found overwhelming evidence that sensorimotor and linguistic distance were $BF_{10} = 3.24 \times 10^{11}$ times better than taxonomic level at predicting accuracy. That is, sensorimotor-linguistic models outperformed taxonomic models in predicting the basic level advantage on both latency and accuracy (see Figs. 3 and 4).

5.2.4. Summary

In conclusion, these results provide further, stronger evidence that sensorimotor and linguistic distributional information contribute to categorical decision making. Critically, even though the classic basic-level advantage was present in our data, results also demonstrate that sensorimotor-linguistic information was systematically better at explaining category verification performance than a taxonomic hierarchy of subordinate, basic, and superordinate levels. The grade of evidence for this superiority effect was generally much stronger in the

Table 4
Model comparisons for linear mixed effect regressions of RT and logistic mixed effects regressions of accuracy in Experiment 3, showing change in conditional R^2 for nested comparisons and Bayes Factors for all comparisons.

Analysis	Model comparison	RT		Accuracy	
		ΔR^2	BF_{10}	ΔR^2	BF_{10}
A	Null model (random effects)	0.332	–	0.295	–
	Taxonomic levels vs. null	0.006	19,242,414.80	0.057	8.88 $\times 10^{18}$
B	Step 1 (Sensorimotor) vs. step 0 (Null model)	0.014	1.60 $\times 10^{10}$	0.116	2.62 $\times 10^{29}$
	Step 2 (Sensorimotor-linguistic) vs step 1.	–0.001	43.78	0.121	10.99
C	Sensorimotor versus taxonomic levels.	–	831.86	–	2.95 $\times 10^{10}$
	Sensorimotor-linguistic versus taxonomic levels	–	36,416.55	–	3.24 $\times 10^{11}$

Note: We report conditional model R^2 for the null model. For all other models, we report change in marginal (fixed effects only) model R^2 . Note that marginal R^2 values are estimates reported only for information and are not used for inferencing, as marginal R^2 values may go down as well as up with additional parameters (see, Nakagawa & Schielzeth, 2013).

present experiment than in previous experiments, suggesting that some of the variable effects in previous experiments may have been due to suboptimal assumptions about how participants implicitly labelled pictured objects. Indeed, in an exploratory analysis, the weighted measures we used in the present study outperformed those we had used in Experiments 1a and 2a (i.e., assuming that all objects were implicitly named at the basic level),⁹ indicating that it is important to consider variety in object naming behaviour when examining categorisation. That is, when we incorporated the normed range of object names that are likely to be activated when people see a particular picture, our weighted measures of sensorimotor and linguistic distributional information offered the best explanation for variation in categorisation performance. People categorised objects more quickly and accurately when the member concept was close to the category concept in both sensorimotor experience and linguistic distributional knowledge.

6. General discussion

The basic-level advantage effect in object categorisation is intrinsically tied to long-standing feature- and/or network-based theories on the nature of categorisation itself and that of conceptual representation (e.g., *Corter & Gluck, 1992; Jolicoeur et al., 1984; Markman & Wisniewski, 1997; Murphy & Brownell, 1985; Rogers & Patterson, 2007; Rosch, Simpson, & Miller, 1976*). In the present work, we have used state-of-the-art measures to show it is possible to express the taxonomic relationships between concepts, that are thought to underly categorisation, in terms of distance in sensorimotor experience and/or linguistic distributional knowledge between category and member concepts, without referring to discrete, binary features (e.g., *has wings, can fly*) or featural dimensions (e.g., *size, weight*). Specifically, we found that the distance in sensorimotor experience between a given category concept (e.g., *DOG*) and a member concept (e.g., *Labrador*) may predict the speed of categorisation at least as well as a division into discrete taxonomic levels (i.e., the basic-level advantage) might (e.g., Experiments 1 and 2a). We also found that adjusting our measure of sensorimotor distance – via linguistic distributional information – to reflect the graded structure of categories enhanced its ability to predict categorisation performance (Experiment 1b, method) to the point that it overall performed about as well in predicting RT and accuracy as models based on the taxonomic level of the category label (Experiments 1b and 2b, although the best model for a given DV varied across studies, see Fig. 4). Finally, we found that weighting our measures of sensorimotor and linguistic distributional distance to reflect the range of names typically produced for a pictured object improved their ability to predict both RT and accuracy well *beyond* that of discrete taxonomic levels (Experiment 3). That is, the present findings show that variation in object naming affects categorisation, and that the latency and accuracy of categorisation 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) more effectively than by explaining the basic-level advantage in terms of discrete taxonomic levels. These findings are in line with recent linguistic-simulation

⁹ We ran an exploratory analysis of sensorimotor and linguistic distance measures that were calculated as per Experiments 1a and 2a (i.e., assuming basic-level object names). Model comparisons favoured sensorimotor distance over a null model containing only random effects for both accuracy ($BF_{10} = 1.90 \times 10^{10}$) and RT ($BF_{10} = 1947.44$) but showed evidence against the addition of linguistic distributional distance for both accuracy ($BF_{10} = 0.05$) and RT ($BF_{10} = 0.20$). The best-fitting weighted average models outperformed the best-fitting of these basic-named models for both accuracy ($BF_{10} = 1.03 \times 10^{20}$) and RT ($BF_{10} = 636,113,631.42$), supporting our weighted-average approach in Experiment 3. We thank an anonymous reviewer for suggesting this analysis; full data and results may be found in the supplemental materials on OSF (<https://osf.io/8cjrm>).

views on the nature of concepts (e.g., *Barsalou et al., 2008; Connell, 2018; Connell & Lynott, 2014; Louwerse, 2011*) as well as recent categorisation research showing similar effects (e.g., *Banks et al., 2021*).

A critical assumption of the present work is not only that concepts are represented by both sensorimotor and linguistic distributional information, but also that categorical relationships between concepts (e.g., taxonomic relationships) may be expressed in terms of their sensorimotor and linguistic distributional similarity. For example, we predict that the sensorimotor and linguistic distributional profiles of taxonomically related concepts (e.g., *Labrador*, and *dog*) are generally more similar to one another than the profiles of taxonomically unrelated concepts (e.g., *guitar* and *dog*). Moreover, the sensorimotor and linguistic distributional profile of a member concept will often – though not always – be more similar to that of a less-generalised category concept (e.g., *Labrador* vs. *dog*) than a more-generalised concept (e.g., *Labrador* vs. *animal*). The taxonomic hierarchy of concepts – and the basic-level advantage in particular – may therefore be a behavioural artefact of sensorimotor and linguistic distributional overlap, but only at the global level and not for every concept individually. This assumption contrasts the perspective traditional taxonomic accounts have taken, such as accounts which argue semantic memory is explicitly structured into taxonomic levels with the basic level as the entry level (e.g., *Jolicoeur et al., 1984*), or accounts which suggest taxonomic levels arise implicitly from the way in which features of member concepts overlap, where the basic level optimally differentiates members from non-members (e.g., *Markman & Wisniewski, 1997; Murphy & Smith, 1982*).

However, the taxonomic and sensorimotor-linguistic approaches to categorisation are not necessarily mutually exclusive. That is, if the taxonomic approach were redefined to describe behavioural phenomena rather than the structure of semantic memory, it can be accommodated within a sensorimotor-linguistic view. For example, in a label → picture category verification task such as the one we employed here, the category label may activate linguistic distributional knowledge as well as sensorimotor representations, which facilitates the categorisation of the subsequent image (*Boutonnet & Lupyan, 2015; Lupyan & Thompson-Schill, 2012*). As a process model of category verification, a sensorimotor-linguistic account is therefore similar to the preparation model proposed by *Murphy and Smith (1982)*, but substitutes functional and perceptual features by simulated modality- and effector-specific sensorimotor experience and/or linguistic distributional knowledge. Importantly, like the original preparation model, the sensorimotor-linguistic preparation model does not assume any taxonomic level is privileged, but rather that performance is solely driven by the degree of overlap in sensorimotor experience and/or linguistic distributional knowledge between category and member concepts. The sensorimotor-linguistic preparation model thus proposes that when participants see a label (e.g., *dog*), it activates a sensorimotor representation of the referent concept, as well as linguistic distributional knowledge about the contexts it may appear in. When the participant sees the subsequent 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.

Importantly, the present work shows that it is possible to capture categorical relations without referring to feature similarity (e.g., *Murphy & Brownell, 1985*). This is an important finding, as there are many reasons why it might be desirable to specify categorical structure and behaviour without assuming feature-based representations of concepts. While various traditional accounts differ in their interpretation of the nature of the similarity processes underlying categorisation (e.g., *Brooks, 1978; Hampton, 1979; Medin & Schaffer, 1978; Nosofsky, 1986; Posner & Keele, 1968; Rosch & Mervis, 1975*), they generally share the assumption that concepts comprise indivisible, static and binary features (e.g., a given concept may or may not possess the features *has wings, can fly*). In categorisation research, such features are frequently derived from participant responses in feature-listing tasks (e.g., *McRae, Cree,*

Seidenberg, & Mcnorgan, 2005), which are assumed to be sufficiently reflective of the correlational structure of the perceived world (Rosch, 1978), in that participants are unlikely to list features for objects that do not possess them. However, therein lie several limitations of a feature-based approach to concepts and categories. Firstly, features generated from feature-listing tasks are necessarily verbalised expressions of people's experience with selected concepts. Consequently, they are typically much better suited to describe concrete concepts (e.g., *dog*, *cup*) than abstract concepts (e.g., *hunger*, *peace*; although see Harpaintner, Trumpp, & Kiefer, 2018). This asymmetry greatly limits the applicability of features as a basis for conceptual representation. If all concepts are represented through features, then why would they be harder to determine for one type of concept compared to the next? Secondly, many listed features are not inherent to the concept they are listed for. For example, in addition to perceptual and functional features, participants may list taxonomic (e.g., *eagle* → *is a bird*), affective (e.g., *wasp* → *is annoying*) and other thematic associations (e.g., *bird* → *builds nests*, *knife* → *used with fork*; see McRae et al., 2005). As a result, it is not clear what exactly is being compared when considering feature overlap between two concepts. Indeed, research shows that measures of feature overlap are among the weaker predictors of concept similarity (Wingfield & Connell, 2022b). Thirdly, features generated in feature-listing tasks are not always uniformly interpretable away from the category or concept they are generated for (e.g., *has a seat* requires knowledge of chairs to be meaningful; Rosch, 1978), and *is large* means something different for metal compared to wooden *spoons* (Medin & Shoben, 1988). In other words, some listed features (e.g., *is large* for *lions*, *limousines*, and *mugs*; McRae et al., 2005) may only become meaningful after the category has been established. It is not clear by what mechanism the meaning of such features is compared between concepts when determining their similarity. Finally, feature-based accounts of categorisation are typically agnostic with regards to the representation of features themselves (e.g., if *bird* is represented by *has wings*, what represents the latter?) and draw a line between features and concepts. It is unclear what mechanism supports this division, and previous arguments that it is warranted on purely operational grounds (e.g., it is useful to assign a feature-status to *has wings* if it allows us to distinguish between the concepts *bird* and *mouse*; Smith & Medin, 1981) are unpersuasive when alternative approaches render it unnecessary.

Of course, like all label → picture categorisation tasks, the present studies focus on concrete concepts and categories, which raises the question of whether the sensorimotor-linguistic account we propose extends to categorisation of abstract concepts. While some have argued sensorimotor grounding is weaker for abstract compared to concrete concepts (e.g., Barsalou & Wiemer-Hastings, 2005; Vigliocco, Meteyard, Andrews, & Kousta, 2009), a growing body of work has shown sensorimotor information is in fact important to representing both concrete and abstract concepts, with complex relationships between various concrete and abstract subdomains and different sensory modalities and motor effectors (Banks & Connell, 2022b; Borghi, Flumini, Cimatti, Marocco, & Scorolli, 2011; Borghi & Zarcone, 2016; Connell & Lynott, 2012; Connell, Lynott, & Banks, 2018; Villani, Lugli, Liuzza, & Borghi, 2019; Villani, Lugli, Liuzza, Nicoletti, & Borghi, 2021). Nonetheless, sensorimotor information is not sufficient alone, particularly for categories that appear to rely on relational information to provide structure (e.g., *art form*: Banks & Connell, 2022b), which is why information from language is also important. Indeed, when examining category production (e.g., name as many types of *fruit* / *science* as possible in 60 s), Banks et al. (2021) found that both sensorimotor and linguistic distributional information contributed to the rank and frequency of listing category members, with an identical pattern of effects for concrete and abstract categories. It therefore appears that the sensorimotor-linguistic account extends to both concrete and abstract categorisation, but the relative use of sensorimotor versus linguistic distributional information varies with the type of categorisation behaviour (i.e., depends on task demands: Connell, 2018; Connell & Lynott, 2014).

A notable deviation from our hypotheses is the unexpectedly weak direct effect of linguistic distributional information in Experiments 1 and 2. It could be the case that sensorimotor simulation is a more reliable source of information in label → picture category verification, as sensory information activated by the category label may easily be verified upon seeing the subsequent image. However, this does not explain why linguistic distributional information predicts categorisation performance above and beyond sensorimotor information in Experiment 3 and ignores the interactive relationship between sensorimotor and linguistic distributional information (Connell, 2018). Indeed, exploratory results from Experiment 3 point towards another explanation: namely that the effect of linguistic distributional information was not adequately captured by our single-point measures in Experiments 1 and 2 (i.e., that assumed a single name for a pictured object). By contrast, calculating a weighted average linguistic distance between the category label and multiple names given to a particular image greatly improved how well it fit the data relative to a single-point variant. Future work may build upon this by incorporating naming distributions into the calculation of distance measures for the purpose of adequately predicting category verification performance via linguistic distributional information.

A possible limitation of the present work is that, while it provides evidence for a sensorimotor-linguistic alternative to existing (feature-based) accounts of processing advantages in categorisation, it does not directly compare the two approaches. An interesting direction for future research would be to explore the relative extent to which sensorimotor-linguistic and feature similarity explain the basic-level advantage in categorisation. Recent work (Wingfield & Connell, 2022b) shows that sensorimotor distance correlates relatively weakly with a measure of feature overlap (derived from Buchanan, Valentine, & Maxwell, 2019), but outperforms it as a predictor of participants' semantic similarity judgments, which suggests that sensorimotor distance is capable of capturing information pertinent to semantic similarity judgments that features cannot. Indeed, that work suggested the same is true of linguistic distributional measures, which also tended to outperform feature overlap in predicting semantic similarity judgments. Future work may build on these findings to determine whether they extend to the basic-level advantage in categorisation.

Of note is that, across Experiments 1b and 2b, we found no evidence that typicality mediated the basic-level advantage; that is, lower subjective typicality ratings did not lead to objects being categorised at the subordinate level. Nonetheless, categorical gradedness *did* affect categorisation when it was modelled via linguistic distance: gradedness-adjusted measures of sensorimotor distance in Experiments 1b and 2b outperformed the unadjusted measures used in Experiments 1a and 2a, respectively. Furthermore, Experiment 3 incorporated the graded structure of categories via production-frequency weighting on sensorimotor and linguistic distributional distance and showed strong effects of both on categorical decision. These results illustrate the ways in which sensorimotor and linguistic distributional information may capture the graded structure of categories. For instance, the gradedness-adjusted measures of Experiments 1b and 2b incorporated the idea that pictures of "good", highly representative, category members are recognised as their basic category concepts (e.g., picture of *jeans* recognised as *trousers*) while pictures of less-representative 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*]), while less-representative member concepts are judged more quickly and accurately when preceded by their specific, subordinate label (e.g., *sweatpants* → [picture of *sweatpants*]), and all other judgments are slower and less accurate according to the sensorimotor distance between the category and member concepts. These findings indicate that categorical gradedness itself – that is, the notion that less-representative category members are implicitly named with a specific (subordinate) label rather

than with a more generic (basic level) label – is valid. The design of Experiment 3 did not permit us to explore items that were predominantly named at the subordinate level, but nevertheless, we showed that measures of sensorimotor and linguistic distributional distance that incorporate weightings of categorical gradedness predict behaviour in a label→picture categorisation task well beyond the division into three discrete taxonomic levels. Finally, since the typicality rating for an object in its basic-level category proved ineffective at detecting how categorical gradedness affects object categorisation, our findings also suggest that typicality ratings may not actually be the best measure of categorical gradedness. Rather – particularly given the fact that our measure of linguistic distance did not meaningfully correlate with the typicality ratings in Experiment 1b – it appears that linguistic distributional information captures aspects of the graded structure of categories that are not captured by subjective ratings of object typicality.

In summary, our measures of sensorimotor and linguistic distributional information successfully captured aspects of the relationship between categories and their members: we have demonstrated that overlap in sensorimotor experience predicts RT and accuracy at least as well as division into discrete taxonomic levels. Moreover, we have shown that sensorimotor experience and linguistic distributional information may capture the graded structure of categories. These findings add to our understanding of sensorimotor-linguistic concepts and categories and provide an alternative to feature- and/or network-accounts of the basic-level advantage.

Funding

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. This article is based on the doctoral dissertation of the first author (van Hoef et al., 2022).

Open access note

All images, code, and data used in this article are available from <https://osf.io/8cjrm> under a Creative Commons Attribution 4.0 International License (CC-BY), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, so long as you give appropriate credit to the original authors and source, provide a link to the Creative Commons license, and indicate if changes were made. All third-party materials included in this article are included in this Creative Commons license. To view a copy of the license, visit <http://creativecommons.org/licenses/by/4.0/>.

CRediT authorship contribution statement

Rens van Hoef: Data curation, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Louise Connell:** Writing – review & editing, Supervision. **Dermot Lynott:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Data availability

All images, code, and data used in this article are available from <https://osf.io/8cjrm> under a Creative Commons Attribution 4.0 International License (CC-BY).

Appendix A. Model statistics for taxonomic-typicality models of accuracy and RT for Experiment 1b datasets excluding and including the item ‘gavel’

Model	Excluding gavel				Including gavel			
	β	95%CI	z	p	β	CI	z	p
Accuracy (taxonomic level, typicality, interaction)								
(Intercept)	3.773	±0.538	–	–	3.715	±0.529	–	–
Subordinate	–1.044	±0.469	–4.362	1.29 × 10–05	–1.010	±0.458	–4.320	1.56 × 10–05
Superordinate	–1.358	±0.455	–5.842	5.17 × 10–09	–1.308	±0.444	–5.777	7.60 × 10–09
Typicality	0.495	±0.733	1.324	0.185	0.601	±0.536	2.195	0.0281*
Subordinate*Typicality	0.254	±0.884	0.562	0.574	0.071	±0.640	0.218	0.8278
RT (taxonomic level, typicality, interaction)								
(Intercept)	721.580	±59.773	–	–	726.789	±60.002	–	–
Subordinate	41.791	±29.318	2.794	0.005	36.792	±29.375	2.455	0.0142
Superordinate	88.163	±29.604	5.837	6.29 × 10–09	83.847	±29.636	5.545	3.36 × 10–08
Typicality	–4.754	±52.788	–0.177	0.86018	–46.695	±43.689	–2.095	0.0382*
Subordinate*Typicality	46.163	±75.710	1.195	0.23222	73.104	±63.086	2.271	0.0232*

* NHST significance test deviating from models based on dataset excluding ‘gavel’.

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](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. (2022a). Category production norms for 117 concrete and abstract categories. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01787-z>
- Banks, B., & Connell, L. (2022b). Multi-dimensional sensorimotor grounding of concrete and abstract categories. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 378(1870), 20210366. <https://doi.org/10.1098/rstb.2021.0366>
- Banks, B., Wingfield, C., & Connell, L. (2021). Linguistic distributional knowledge and sensorimotor grounding both contribute to semantic category production. *Cognitive Science*, 45(10), Article 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.
- Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher, & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). Cambridge University Press.
- Bartón, K. (2020). MuMin: Multi-Model Inference. R Package Version 1.43.17. (1.43.17) [Computer Software]. <https://CRAN.R-project.org/package=MumIn>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., ... 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>
- Borghini, A. M., Flumini, A., Cimatti, F., Marocco, D., & Scorolli, C. (2011). Manipulating objects and telling words: A study on concrete and abstract words acquisition. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00015>
- Borghini, A. M., & Zarcone, E. (2016). Grounding abstractness: Abstract concepts and the activation of the mouth. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01498>
- 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>
- 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), Article 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), Article e106953. <https://doi.org/10.1371/journal.pone.0106953>
- Brooks, L. (1978). Nonanalytic concept formation. In E. Rosch, & B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Lawrence Erlbaum Associates, Publishers.
- Buchanan, E. M., Valentine, K. D., & Maxwell, N. P. (2019). English semantic feature production norms: An extended database of 4436 concepts. *Behavior Research Methods*, 51(4), 1849–1863. <https://doi.org/10.3758/s13428-019-01243-z>
- 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. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, 125(3), 452–465. <https://doi.org/10.1016/j.cognition.2012.07.010>
- Connell, L., & Lynott, D. (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., & Banks, B. (2018). Interoception: The forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 373(1752), 20170143. <https://doi.org/10.1098/rstb.2017.0143>
- Connell, L., & Ramsar, 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>
- 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), Article e57410. <https://doi.org/10.1371/journal.pone.0057410>
- 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](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. (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](https://doi.org/10.1016/S0022-5371(79)90246-9)
- 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.
- Harpaintner, M., Trumpp, N. M., & Kiefer, M. (2018). The semantic content of abstract concepts: A property listing study of 296 abstract words. *Frontiers in Psychology*, 9. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01748>
- 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](https://doi.org/10.1016/S0896-6273(03)00838-9)
- 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>
- 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>
- van Hoef, R., Lynott, D., & Connell, L. (2022). Timed picture naming norms for 800 photographs of 200 objects in English. *Psy ArXiv*. <https://doi.org/10.31234/osf.io/2h6q7>
- JASP Team (2019). JASP (Version 0.9.2)[Computer software].
- 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](https://doi.org/10.1016/0010-0285(84)90009-4)
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–16. <https://doi.org/10.18637/jss.v082.i13>
- 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>
- van Lenth, R. (2021). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means (1.7.0)* [Computer Software].
- 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](https://doi.org/10.1016/S0022-5371(84)90170-1)
- 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>
- 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., & 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](https://doi.org/10.1016/0010-0285(88)90018-7)
- 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](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>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- 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>
- 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>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.r-project.org/>.

- Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, 3(2), 303–345. <https://doi.org/10.1111/j.1756-8765.2010.01111.x>
- 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](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](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., 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](https://doi.org/10.1016/0010-0285(76)90013-X)
- 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](https://doi.org/10.1016/0010-0285(75)90024-9)
- 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>
- 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>
- 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](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>
- Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, 1(2), 219–247. <https://doi.org/10.1515/LANGCOG.2009.011>
- Villani, C., Lugli, L., Liuzza, M. T., & Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition*, 11(3), 403–430. <https://doi.org/10.1017/langcog.2019.23>
- Villani, C., Lugli, L., Liuzza, M. T., Nicoletti, R., & Borghi, A. M. (2021). Sensorimotor and interoceptive dimensions in concrete and abstract concepts. *Journal of Memory and Language*, 116, Article 104173. <https://doi.org/10.1016/j.jml.2020.104173>
- 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. (2022a). Sensorimotor distance: A grounded measure of semantic similarity for 800 million concept pairs. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-01965-7>
- Wingfield, C., & Connell, L. (2022b). Understanding the role of linguistic distributional knowledge in cognition. *Language, Cognition and Neuroscience*, 0(0), 1–51. <https://doi.org/10.1080/23273798.2022.2069278>
- 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>