INTERFACES BETWEEN LANGUAGE AND COGNITION:

METAPHOR, ICONICITY, AND MULTIMODAL NUMERICAL COMMUNICATION

by

GREG WOODIN

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Supervised by

Professor Jeannette Littlemore, Dr Bodo Winter, and Dr Marcus Perlman

Department of English Language and Linguistics

School of English, Drama and Creative Studies

College of Arts and Law

University of Birmingham

September 2022

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

This thesis explores metaphor, iconicity, and multimodal numerical communication, and their connection to the cognitive processes that are involved in language comprehension and production. Focusing on the English language, this thesis discusses the idea that linguistic metaphors reflect deeper conceptualisations of the world, which are expressed multimodally through metaphoric and iconic gesture, prosody, and vocalization. It also studies the representation of these metaphors in cultural artifacts, like line graphs, and how these representations can make data visualizations easier to comprehend. By focusing on number metaphors and then expanding its scope to include numerical communication more broadly, this thesis explores how communication about numbers reflects numerical cognition. It also argues that sensorimotor simulation motivates iconic and metaphoric expression in different modalities, and discusses the implications of this argument for language evolution. By synthesising insights from a wide range of research areas and methodologies from across the spectrum of linguistics, cognitive science, and psychology, this thesis demonstrates that metaphor, iconicity, and multimodal numerical communication are interfaces between language and cognition.

ACKNOWLEDGMENTS

I was supported by the Economic and Social Research Council [grant number ES/P000711/1].

LIST OF PAPERS

This thesis incorporates the following published journal articles:

1. **Woodin, G.**, Winter, B., Perlman, M., Littlemore, J. & Matlock, T. (2020). 'Tiny numbers' are actually tiny: Evidence from gestures in the TV News Archive. *PLOS One*, 1–21. https://doi.org/10.1371/journal.pone.0242142
2. **Woodin, G.**, Winter, B. & Padilla, L. (2022). Conceptual metaphor and graphical convention influence the interpretation of line graphs. *IEEE Transactions on Visualization and Computer Graphics*, 28(2): 1209–1221. http://doi.org/10.1109/TVCG.2021.3088343

This thesis also incorporates the following unpublished journal articles:

3. **Woodin, G.**, Winter, B., Littlemore, J., Perlman, M. & Grieve, J. (under review). Large-scale patterns of number use in spoken and written English.
4. **Woodin, G.**, Winter, B. & Perlman, M. (in prep). The iconic impetus of language: From sensorimotor simulation to the lexicon.

The first of these papers has been submitted to an academic journal and is under review. The second is in a complete state, but additional work is needed to adapt the paper for journal submission. Detailed authorship contributions for each paper are presented in the chapters associated with each paper. All other chapters are entirely my own work. References for each paper are included at the end of each paper chapter. The bibliography at the end of the thesis is for citations in the non-paper chapters.

During the PhD, the following papers were also published or accepted for publication:

- Perlman, M. & **Woodin, G.** (2021). A complete real-world theory of language should explain how iconicity remains a stable property of linguistic systems. *Journal of Cognition*, 4(1): 43, 1–4. http://doi.org/10.5334/joc.166
- Banks, B., Borghi, A., Fargier, R., Fini, C., Jonauskaite, D., Mazzuca, C., Montalti, M., Villani, C., **Woodin, G.** (accepted). Consensus paper: Current perspectives on abstract concepts and future research directions. *Journal of Cognition.*

CONTENTS

1. INTRODUCTION 1

2. LITERATURE REVIEW 9

2.1. Conceptual Metaphor Theory 9

2.2. Experimental Evidence for Conceptual Metaphors 13

2.3. Metaphoric Gestures 21

2.4. Conceptual Metaphors in Cultural Artifacts 29

2.5. Origins of Conceptual Metaphors 32

2.6. Number Metaphors and Numerical Communication 38

2.7. Iconicity in Multimodal Communication 41

2.8. Sensorimotor Simulation in Language Comprehension 47

2.9. Sensorimotor Simulation in Iconic Language Production 50

2.10. Sensorimotor Simulation in Conceptual Metaphor 53

3. PAPER 1 61

4. PAPER 2 64

5. PAPER 3 65

6. PAPER 4 66

7. DISCUSSION 67

BIBLIOGRAPHY 78

1. INTRODUCTION

Metaphor and iconicity can both express meaning by drawing on associations between different domains, which is evident in communication about numbers. For example, the metaphor 'tiny number' describes the magnitude of a number using words that denote physical size, despite numbers not literally being small or large. This phrase draws on the natural correlation between physical size and number: quantities with fewer items tend to be physically smaller (see Hurewitz et al., 2006). When using the phrase 'tiny number', speakers may produce a gesture in which the forefinger and thumb are pinched together, with the space between these fingers representing the small size of the number. In this example, when numerical quantity is described using words related to size (e.g., 'tiny'), we call it a metaphor. When the size of an entity is depicted by the spatial parameters of manual gestures (rather than size-related words), we call it iconicity. Furthermore, because the meaning of the iconic pinching gesture is not size itself, but numerical quantity, this iconic gesture is also metaphoric. This example demonstrates how speakers communicate metaphorically and iconically about concepts like numerical magnitude, using different communicative channels (e.g., vocal and manual) to achieve this task.

Metaphor is a figure of speech, pervasive in spoken languages, that defines one concept in terms of another. Take, for example, this famous example from Shakespeare's *As You Like It:* "All the world's a stage, and all the men and women merely players". The world is not literally a stage, but this metaphor implies that life is full of drama, we all play roles – social, professional, or otherwise – and our lives constitute a brief appearance on this stage in the context of the entire play. Other metaphors are less obvious but are

nonetheless prevalent, such as in words indicating spatial relations, which are used to denote concepts like numerical quantity, time, and emotional valence (good/bad). For example, English speakers talk of numbers as 'high' and 'low', look 'forward' to the future and look 'back' at the past, and express feeling 'up' or 'down' based on their mood (e.g., Casasanto & Boroditsky, 2008; Lakoff & Johnson, 1980b; Winter et al., 2015). These expressions are metaphorical because they describe non-spatial domains (numerical magnitude, time, and emotional valence) in terms of spatial locations, mapping concrete spatial language onto more abstract concepts.

Iconicity is the phenomenon where a communicative form (e.g., a word or sign) resembles its meaning in some way (Peirce, 1931). An example of iconicity in spoken languages is onomatopoeia, where the sound of a word resembles its auditory meaning (Bladon, 1977; Bredin, 1996). For example, 'ka-boom!' depicts the sound of an explosion (see Taylor, 2007). Some spoken languages have a grammatical category of iconic words called ideophones, including the Japanese 'gorogoro' and 'korokoro', which respectively mean 'a heavy/light object rolling repeatedly' (Kita, 1997; Perniss et al., 2010). In these ideophones, the contrast in voicing between /g/ and /k/ depicts the weight of the rolling object, while the reduplication of syllables depicts the iterative nature of the rolling movement. Even in spoken languages without a dedicated category of ideophones, like English (Nuckolls, 1999), certain sounds are systematically used to convey certain meanings, which has been termed sound symbolism. In English and other spoken languages, for instance, high front vowels such as /i/ and /ɪ/ tend to appear in words for 'small' meanings (see Blasi et al., 2016; Ultan, 1978; Winter & Perlman, 2021), like 'tiny' /taɪni/, 'itsy-bitsy' /ɪtsɪbɪtsi/, and 'mini' /mɪni/, whereas back vowels such as /a/, /u/, and /ɔ/

tend to appear in words for ‘large’ meanings (see Johnson, 1967; Ohala, 1983; Winter & Perlman, 2021), such as in the word ‘large’ /laːdʒ/ itself, ‘humongous’ /hjuːmʌŋgəs/, and ‘enormous’ /ɪnɔːməs/. This phonological pattern represents size by tapping into the association between the size of smaller (larger) entities and the higher (lower) frequency sounds that these entities, particularly animals (e.g., mice versus lions), tend to produce (Fitch, 1994; Ohala, 1994; Winter et al., 2021).

Some of these forms of iconicity, such as onomatopoeia, use a communicative channel (e.g., the voice) to represent meanings in the same modality (e.g., auditory meanings). In comparison, cross-modal associations motivate the use of iconic sound symbolism like the use of the voice to represent physical size, a non-auditory meaning. Because sound symbolism maps meaning from one domain (e.g., vowel frequency) to another (e.g., size), cross-modal iconicity is sometimes treated as being equivalent, or similar, to metaphor (Marks, 2014). In fact, both metaphor and cross-modal iconicity may derive from associations formed through noticing correlations in our experience. For example, upper space may become associated with greater quantities because quantities (e.g., water in a glass) tend to rise as they increase (Lakoff, 1987, p. 276), forming the basis for the metaphoric description of numbers as ‘low’ and ‘high’. Analogously, low frequency sounds may become associated with large entities because large entities (e.g., elephants) tend to make low frequency sounds (Perlman, Clark, et al., 2015).

This link between metaphor and iconicity is made clearer if we consider that the expression of metaphors via gestures can be seen as a kind of iconicity (see Bouissac,

2008). For example, in Figure 1, a speaker uses the linguistic metaphor[1] 'huge numbers' and gestures by moving his hands outward from one another, the space between his flat palms representing the 'huge' size of these numbers. Metaphoric gestures such as this are also iconic gestures, as they iconically depict the metaphor's source domain (Cienki & Müller, 2008), which is the domain (e.g., size) used to represent the target domain (i.e., literal meaning) of the metaphor (e.g., numerical quantity). If the speaker were depicting, for example, the size of a fish he had caught, we would call this gesture iconic, with the hands acting as boundaries that mark the extremities (face and tail) of the fish, and the space between these hands depicting the fish's size. However, because the speaker is talking about a number, which is abstract and therefore does not have a physical size, his gesture is metaphoric, as well as iconic.

[1] In this thesis, I treat words in spoken languages and signs in signed languages as 'linguistic', and anything else (e.g., gesture, prosody) as non-linguistic. It should be noted that some authors have advocated gestures being treated as part of language, rather than being separate from it (Kendon, 2004; e.g., McNeill, 2000). For the purposes of this thesis, this debate is not important, and I should stress that calling gesture 'non-linguistic' is not intended to disregard the importance and prevalence of this channel of communication, or the ways in which speech and gesture may be combined in speakers' utterances (e.g., Enfield, 1999), planned in tandem (e.g., Melinger & Levelt, 2004), and so on.



Figure 1. CNBC presenter produces a metaphoric gesture while using the linguistic metaphor 'huge numbers'. As with all metaphoric gestures, this gesture is also iconic. White arrows have been added to this image to illustrate the movement direction of the gesture. Source: CNBC Worldwide Exchange, accessed via the TV News Archive: https://tinyurl.com/5efhrdzw.

Other forms of communication can also be both metaphoric and iconic, such as prosody, which includes modulating the pitch, speed, rhythm, and timbre of one's voice during speech (e.g., Ćwiek & Fuchs, 2019; Nygaard et al., 2009; Perlman, Clark, et al., 2015). For example, Shintel et al. (2006) found that people raised their vocal pitch when describing the upward movement of a dot on a screen, reflecting the association between

height and pitch instantiated in English and other languages (e.g., German) in the description of pitches as ‘high’ and ‘low’. Pitch is an auditory phenomenon and so is not literally ‘high’ or ‘low’, so the expression of verticality through pitch can be considered metaphoric (e.g., Dolscheid et al., 2012, 2013). Because this use of vocal pitch iconically depicts sounds of different pitches, it is also iconic.

Another form of communication that can be both metaphoric and iconic is non-linguistic vocalization, which involves depicting sounds using non-verbal vocal noises (e.g., Edmiston et al., 2018; Lemaitre et al., 2016; Sarvasy, 2016). For instance, Perlman et al. (2022) reported that deaf and hearing Chinese children and adults asked to produce vocalizations to denote different items produced louder vocalizations for larger versus smaller items (e.g., big versus small ball). These vocalizations tap into the association between larger objects and the louder sounds they tend to make relative to smaller objects (e.g., Grassi, 2005; see Spence, 2011). These vocalizations are iconic as they depict the sounds that objects of different sizes tend to make, but they are also metaphoric, as they depict the source domain of volume to denote the target domain of physical size.

Metaphoric and iconic gesture, prosody, and vocalization open a window onto the mind that is not afforded by conventional words. Scholars have previously argued that because metaphoric and iconic gestures are produced less deliberately and more spontaneously than conventionalized speech, they allow insight into the conceptualization processes that underlie language production (De Ruiter, 2007; McNeill, 1992, 2000, 2005). A similar argument can be made about metaphoric and iconic prosody and vocalizations. With regard to the cognitive processes that metaphoric and iconic gestures, prosody, and vocalizations reveal, these three forms of communication may be outward expressions of

sensorimotor simulations. Sensorimotor simulation is the mental reactivation of previously experienced perceptions and actions, which recruits the same brain areas active during perception and action but to a lesser degree (e.g., Barsalou, 1999, 2007; Decety & Jeannerod, 1995). These simulations are productive and hence can be used to imagine an infinite number of experiences by combining previously experienced perceptions and actions (Barsalou, 1999; Barsalou & Prinz, 1997). Metaphoric and iconic gestures, prosody, and vocalizations may reveal the sensorimotor simulations that are active during language production, which is an idea that is explored throughout this thesis.

This thesis investigates metaphor and iconicity, exploring the idea that linguistic metaphors reflect deeper conceptualisations of the world, which are communicated multimodally via metaphoric and iconic gesture, prosody, and vocalization. It also studies the representation of these metaphors in cultural artifacts, such as line graphs, and how these representations can make data visualizations easier to interpret. The thesis focuses on abstract concepts such as numerical quantity, time, and emotional valence, because, traditionally, conceptual metaphors have been seen as grounding abstract concepts in embodied, concrete experience (e.g., Gibbs, 1994; Kövecses, 2002; Lakoff & Johnson, 1980b). In particular, the thesis provides an in-depth illustration of the varied multimodal strategies with which people communicate about numerical quantities, and how these strategies may reflect numerical cognition, demonstrating the multiple ways with which abstract concepts may be cognitively represented and communicated about. It also puts forward the argument that sensorimotor simulation motivates iconic and metaphoric expression in different modalities, and discusses the implications of this idea for theories of language evolution. This thesis focuses on the English language, making comparisons with

other spoken and signed languages (e.g., Dutch, Japanese, American Sign Language) when helpful to do so.

Overall, this thesis explores the connection of metaphor, iconicity, and multimodal numerical communication to the cognitive processes that are involved in language comprehension and production, revealing these phenomena as interfaces between language and cognition. To expound this thesis, Chapter 2 presents a review of the literature, covering the main topics explored, and Chapters 3–6 contain the four main papers, with Papers 1–3 (§3–5) reporting empirical research, and Paper 4 (§6) making a theoretical contribution based on prior research. Chapter 7 concludes the thesis by discussing the papers in light of the main themes of the thesis, and suggesting future research directions.

2. LITERATURE REVIEW

This literature review covers the foundational knowledge needed to understand the main topics of interest in this thesis, and provides context for the papers presented. First, I introduce the idea that linguistic metaphors reflect the way we think (§2.1), and report experimental evidence to support this point (§2.2). In the next chapters, I discuss the expression of these conceptual metaphors through gestures (§2.3) and cultural artifacts such as data visualizations (§2.4), before addressing the potential origins of these conceptual metaphors (§2.5). Following this, I discuss how conceptual metaphors of numerical quantity can give insight into numerical cognition and communication more generally (§2.6). Iconicity in multimodal communication is then explored in detail (§2.7), before the role of sensorimotor simulation in language comprehension, especially the comprehension of iconic vocabulary, is discussed (§2.8). Then, I argue that sensorimotor simulation is implicated in the production of iconic communication (§2.9), and that sensorimotor simulation is also involved in conceptual metaphor (§2.10). The specific relevance of these ideas to the main papers presented in this thesis is discussed at the beginning of each of these chapters. The main papers that comprise the thesis are then introduced.

2.1. Conceptual Metaphor Theory

This chapter introduces the notion that linguistic metaphors may reflect conceptual metaphors that structure our conceptualisation of the world. It focuses on metaphors of numerical quantity, which are investigated in Papers 1 and 2, and metaphors of time and emotional valence, which are investigated in Paper 2.

Conceptual Metaphor Theory (CMT) (e.g., Gibbs, 1994; Kövecses, 2002; Lakoff & Johnson, 1980b) is a theoretical framework affiliated with cognitive linguistics (e.g., Evans & Green, 2011; Fauconnier, 1994; Langacker, 2008) that describes how people understand and reason about abstract concepts. CMT starts with the observation that people often use linguistic metaphors to describe abstract concepts in terms of more concrete ones.[2] For example, English speakers talk about 'low numbers', 'high numbers', 'plummeting' share prices, and 'skyrocketing' property prices (examples from British National Corpus; BNC Consortium, 2007). In these examples, numerical quantity is described as increasing upward through physical space. The numbers themselves are not literally 'low' or 'high'; hence these expressions are deemed to be metaphorical. Each of these expressions is based on a mapping from the source domain of space to the target domain of numerical magnitude. The key insight from CMT is that this mapping is instantiated in a range of conventionalized linguistic expressions: as well as the expressions already described, people talk about 'rising inflation', 'falling sales', and about prices going 'up' or 'through the roof' (BNC Consortium, 2007). This mapping is productive in that it can generate novel expressions in which numerical quantity is described using any terms relating to vertical space, such as if a person were to state that the cost of living is 'on a steep upward trajectory', or that interest rates have 'nosedived'.

Space is a common source domain for the mapping of abstract target domains in English. In addition to numerical quantity, vertical space terms are used to describe

[2] While CMT focuses on abstract concepts, it should be noted that not all linguistic metaphors are about abstract concepts. For example, Shakespeare's metaphor "All the world's a stage, and all the men and women merely players" is about the world, which is not usually considered abstract.

emotional valence, a term referring to the perception of concepts as good or bad (e.g., Barrett & Russell, 2016; Lang et al., 1997; Russell, 2003; Warriner et al., 2013). For instance, being in a good mood can be described as being 'on a high', and a bad mood as 'feeling low', and likewise, stating that a situation is 'looking up' expresses optimism, whereas claiming that things are 'going downhill' is pessimistic. Other abstract target domains, such as time, are described in terms of other spatial axes, including the sagittal (front-back) axis. For example, people often describe looking 'forward' to future events and 'back' at past ones, put experiences 'behind' them, and reschedule events 'forward' or 'backward' in time. Words like 'hindsight' and 'retrospective', which describe reflecting on past events, also tap into the same mapping, with 'hind-' and 'retro-' denoting backward locations or motion. Some abstract target domains, like numerical quantity, are described using size-based spatial terms, as well as the vertical spatial terms discussed above. For instance, lesser quantities can be described as 'tiny' or 'small' and greater quantities as 'large' or 'huge', and numerical changes can be characterised as 'growing' or 'shrinking' (Winter et al., 2013).[3]

In this thesis, I focus on the spatial metaphors of numerical quantity, time, and emotional valence described above, with a particular focus on the various ways in which numerical quantity is conceptualised and communicated about multimodally. However,

[3] Throughout this thesis, I use the terms 'greater' and 'lesser' to describe numbers of a higher and lower magnitude, respectively, to differentiate discussion about the numbers themselves from the spatial metaphors that people use to describe them. This choice of language is an attempt to use a relatively neutral, non-metaphorical form of expression, although I recognise that it is practically impossible to do so in English, as 'greater' and 'lesser' still have some (albeit weaker, I would argue) metaphorical associations with physical size.

people use many metaphors other than these. For example, people often describe similarity in terms of proximity, with people describing political views as ‘very far from’ (different) or ‘very close to’ (similar) their own (Winter & Matlock, 2017, p. 106). Similarly, people discussing two similar shades of blue might say they ‘aren’t identical, but they’re close’ (Casasanto, 2008, p. 1047). Moreover, evil, scary, or depressing things are often described metaphorically in terms of darkness (Winter, 2014), such as in the expressions ‘dark times’ (Kövecses, 2002, p. 85) and ‘black comedy’, and in opposite expressions such as ‘brighten up!’ (Kövecses, 2002, p. 85) and ‘lighten the mood’ (BNC Consortium, 2007). People also describe arguments using war-related language in sentences such as ‘Your claims are indefensible’, ‘His criticisms were right on target’, and ‘I demolished his argument’ (Lakoff & Johnson, 1980a, p. 454). As a final example, people often describe time in terms of money in expressions such as ‘How do you spend your time these days?’, ‘He’s living on borrowed time’, and ‘I’ve invested a lot of time in her’ (Lakoff & Johnson, 1980a, p. 456).

As we can see, metaphorical mappings underlie a wide range of linguistic expressions, and they are also productive in the generation of new expressions. For example, a poet might use a space-time metaphor of focusing on her rear-view mirror (reminiscing about the past) instead of keeping her eyes on the road ahead and driving forward (being present). In addition, a vaccine developer might talk about the chances of adverse side effects being ‘microscopic’, tapping into the size-number mapping. The systematicity with which these metaphorical mappings are instantiated in linguistic expressions has led CMT theorists to posit that the linguistic expressions themselves are the surface-level manifestation of deeper conceptual mappings. In other words, these

linguistic metaphors are not merely linguistic but reflect the ways in which people conceptualise the world.

According to this line of thinking, the linguistic metaphors in which people describe numbers as 'low' and 'high' reflect an underlying *conceptual metaphor* in which greater numbers are associated with upper space, and lesser numbers are associated with lower space. Similarly, the metaphorical framing of numbers as 'small' and 'large' may reflect an association between larger physical sizes and greater numbers, and smaller sizes and lesser numbers. Emotional valence metaphors like 'on a high' and 'feeling low' may be surface-level representations of an association between upper space and positive valence (good), and lower space and negative valence (bad). In the domain of time, people may associate frontward space with the future (consistent with expressions like 'looking forward') and backward space with the past (consistent with expressions like 'put it behind you'). In all these examples, conceptual metaphors have primacy, and are expressed through linguistic metaphors.

2.2. Experimental Evidence for Conceptual Metaphors

As we have seen, proponents of CMT argue that linguistic metaphors reflect conceptual metaphors that structure our conceptualization of the world. This chapter reports non-linguistic evidence for the conceptual nature of the linguistic metaphors described in the previous chapter, which are explored in Paper 1 (size-based metaphors of numerical quantity) and Paper 2 (vertical metaphors of numerical quantity and emotional valence). Because these experiments are non-linguistic, they also reveal evidence for associations that are not expressed through language, but that can nonetheless be described as

metaphoric. These non-linguistic metaphors are explored in Paper 1 (horizontal metaphors of numerical quantity) and Paper 2 (horizontal metaphors of numerical quantity and emotional valence, horizontal and vertical metaphors of time).

The CMT claim that language represents deeper conceptual metaphors was originally supported through reference to language itself: the systematicity with which metaphorical mappings (e.g., between vertical space and numerical quantity) are attested in language, and the productivity of these mappings to generate novel expressions, was interpreted as evidence for conceptual metaphors. Rightly, this line of reasoning was criticised for its circularity, as language was being interpreted as 'revealing' conceptual metaphors, but was also being put forward as evidence for the reality of conceptual metaphors in the first place (Murphy, 1996, 1997). Since then, researchers have probed conceptual metaphors through experimental, non-linguistic methods (see Gibbs, 1994; Landau et al., 2010; Winter & Matlock, 2013). Also, researchers focused on numerical cognition whose hypotheses have often not always been directly inspired by CMT have nevertheless reported evidence that can be interpreted as supporting the core claims of CMT.

Experiments show evidence for a vertical conceptual metaphor of numerical magnitude, where people often think about numerical magnitude as increasing upward through space, consistent with linguistic metaphors that describe numbers in terms of vertical space. For example, people asked to verbalise numbers as randomly as possible choose greater numbers when they are moved upward in a body-lifting device than when they are moved downward (Hartmann et al., 2012), and when they look upward rather than downward (Winter & Matlock, 2013). Other studies show that when adults are tasked with

categorising numbers as odd or even (Hartmann, Gashaj, et al., 2014), and when children are asked to select the greater of two numbers (Cooney et al., 2021), both groups are quicker when the response button for the greater number is positioned higher than the response button for the lesser number. Moreover, when people are asked to place the quantity words ‘least’, ‘less’, ‘more’, and ‘most’ on a piece of paper, they tend to place the words vertically, ascending upward from ‘least’ to ‘most’ (Woodin & Winter, 2018). People’s gaze also tends to shift upward when people count aloud in an ascending sequence (Hartmann et al., 2016) and when processing greater as opposed to lesser numbers (Felisatti et al., 2022).

Experimental evidence also shows that people tend to conceptualise numerical magnitude in terms of physical size, associating lesser and greater numbers with smaller and larger sizes, respectively, consistent with size-number linguistic metaphors. For example, in magnitude comparison tasks, participants are quicker to select the greater of two numbers when the greater number is written in a larger typeface (Henik & Tzelgov, 1982), and to select the dot display containing more dots when the display with more dots spans a larger area (Hurewitz et al., 2006). Lesser and greater numbers have also been linked to manual actions associated with the manipulation of smaller and larger objects, respectively. For instance, people tend to increase the distance between their thumb and forefinger when reaching to grasp wooden blocks inscribed with greater numbers, regardless of the actual size of the blocks (Andres et al., 2008). Participants also initiate a precision grip (thumb and forefinger approaching one other) more quickly when responding to lesser numbers, whereas they initiate a power grip (firm, whole-hand) more quickly when responding to greater numbers (Lindemann et al., 2007). A random number generation task

showed that merely observing hand-opening movements caused participants to verbalize greater numbers compared to when they observed hand-closing movements (Badets et al., 2012). Even infants have been found to associate closed hand shapes with lesser numbers (Decarli et al., 2022).

Consistent with linguistic metaphors of emotional valence, studies also support the idea that emotional valence tends to be associated with vertical space, with people associating upper space with positive valence and lower space with negative valence. A seminal study conducted by Meier and Robinson (2004) showed that people were quicker to respond to positively valenced words such as ‘pride’ when they appeared in a higher position on a computer screen, and to negatively valenced words such as ‘liar’ when they appeared in a lower screen position. Another study showed that participants were quicker to judge the valence of words with vertically oriented response buttons when the upper button was used to indicate positive valence and the lower button to indicate negative valence (Castaño et al., 2018). In addition, Casasanto and Dijkstra (2010) showed that participants recalled more positive memories when moving marbles upward from a lower box to a higher box than when they moved marbles in the reverse, downward direction (see also Seno et al., 2013). Similarly, Globig et al. (2019) found that people recalled more positively valenced words from a list of nouns when they moved their head upward, and more negatively valenced words when moving their head downward. Finally, Sasaki et al. (2015) showed that participants rated images more positively after dragging a dot to an upper location of a screen.

If linguistic metaphors are surface-level expressions of conceptual metaphors, it stands to reason that there may be conceptual metaphors that are not instantiated

linguistically. Hence, conceptual metaphor may be a more prevalent strategy with which people conceptualise different domains than language alone would suggest. In CMT, the term 'metaphor' is still used to describe these instances, despite language not being involved, as a target domain is still being conceptualised in terms of a source domain. For example, there are no linguistic metaphors in which people describe numerical magnitude in horizontal terms – for example, English speakers do not usually state that a number is 'lefter' or 'righter' than another. Despite this lack of linguistic representation, experiments show that people conceptualise numerical magnitude using the horizontal axis along a 'mental number line'. The *Spatial-Numerical Association of Response Codes* (SNARC) effect demonstrates that people tend to respond more quickly to lesser numbers with a left-sided response, and to greater numbers with a right-sided response (e.g., Aleotti et al., 2020; Dehaene et al., 1993; Fischer & Shaki, 2014), which has also been replicated in an online context (Cipora et al., 2019). While this effect was initially shown to be related to left- and right-handed responses, Dehaene (1993, Experiment 6) showed that when participants' hands were crossed over, they responded more quickly to lesser numbers with their right hand (on the left side) and to greater numbers with their left hand (on the right side). Other studies show that processing a greater number followed by a lesser number induces leftward eye movements (Loetscher et al., 2008), and leftward saccades can be used to predict that the next number verbalised by a participant in a random number generation task will be of a lesser magnitude than the previous one (Loetscher et al., 2010). When counting in an ascending sequence, people's gaze position also tends to shift rightward (Hartmann et al., 2016). Moreover, in a free placement task, participants tend to place the numbers '2', '4', '7', and '9' from left to right across a piece of paper (Woodin & Winter, 2018). Finally, a study

required participants to reproduce the location of a numeral they had just seen within a rectangular space, finding that greater numbers were reproduced in a more rightward position on average than lesser numbers (Aulet et al., 2021).

Like numbers, emotional valence may be conceptualised in horizontal terms, even though this conceptualisation is not usually expressed through linguistic metaphors. In particular, Casasanto's (2009) *body-specificity hypothesis* predicts that people associate positive valence with the side of space corresponding to the dominant side of their body, whereas negative valence is associated with the non-dominant side. Motivating this hypothesis, Casasanto found that right-handers tend to sort items framed as 'good' into a right-side box on a piece of paper and 'bad' items into a left-side box, whereas left-handers tend to do the opposite. Casasanto and Henetz (2012) obtained similar results for children as young as five years old, also finding that right-handed children attribute more positive traits to animals presented on the right, whereas left-handed children evaluate animals presented on the left more positively. In another study, strongly right-footed participants were quicker to categorise the valence of positive words with their right foot, and negative words with their left foot (de la Vega et al., 2015). Li (2020) also found general support for the body-specificity hypothesis in patients with unilateral stroke and individuals who had suffered the loss of a limb. Despite the lack of linguistic metaphors that express this horizontal conceptualisation of emotional valence, the use of 'right' as a synonym for 'correct' may derive from the predominance of right-handers in society and the resultant dominance of the association between positive valence and right-side space. In addition, etymological analysis shows that 'sinister', a word with negative connotations, derives from the Latin word for 'left' (Casasanto, 2009). The Bible also states that Jesus sits at the right

hand of God (Plevnik, 1991), a place of honour and stature, whereas being at God's left hand is often seen as a place of disrepute (Marshall, 1968). Therefore, although this conceptual metaphor is not expressed through linguistic metaphors, we find evidence of this conceptual metaphor in other domains.

Similar to numerical quantity and emotional valence, people tend to conceptualise time in horizontal terms, despite not expressing this association through linguistic metaphors. For example, it is unacceptable in English to talk about earlier times as 'lefter' than later ones, or to reschedule a meeting 'rightward' in time. Analogous with the SNARC effect, the *Spatial-Temporal Association of Response Codes* (STEARC) effect shows that, when Westerners[4] are asked to indicate whether a sound was heard earlier or later than expected based on the timing of previous sounds, they are faster at responding to 'earlier' sounds with a left-side button and to 'later' sounds with a right-side button (Ishihara et al., 2008; Vallesi et al., 2008, 2011). Comparable results have been reported for short and long stimuli durations (Conson et al., 2008) and sequential time-related concepts such as days of the week (Gevers et al., 2003, 2004). In addition, in a free response task, participants tend to arrange time-related words ('past', 'future', 'earliest', 'earlier', 'later', 'latest') in a left-to-right direction across a piece of paper (Woodin & Winter, 2018; see also Leone et al., 2018). These results show that Western people often conceptualise time as passing from left to right across a horizontally oriented *mental timeline.* The idea of a mental timeline with a rightward direction is coherent with the observation in the visual semiotics literature that

[4] As discussed in chapter 2.5, some non-Western cultures have been shown to conceptualise time in the opposite, rightward direction (Fuhrman & Boroditsky, 2010; Tversky et al., 1991).

'given' information (i.e., something the reader already knows) in a text usually appears on the left, whereas 'new' information (i.e., something the reader does not yet know) tends to appear on the right (Van Leeuwen & Kress, 1995).

Regarding the vertical axis, there is debate about whether Western[5] people think about time flowing from bottom to top – another representation that is not instantiated in English. One study reported that Westerners are quicker at responding to a square positioned in upper space that is paired with a future-related word, and to a lower square paired with a past-related word (Ruiz Fernandéz et al., 2014). Eye-tracking studies have also reported that downward versus upward saccades are triggered when Western people hear past- versus future-related words (Stocker et al., 2016), or think about the past versus the future (Hartmann, Martarelli, et al., 2014). Another study showed that Western people sometimes organise time concepts (past, present, and future) chronologically upwards (Leone et al., 2018). However, the response medium in all these studies was a screen, which could not detect sagittal associations due to its two-dimensionality. Because more distant objects along the sagittal plane appear higher than objects that are closer (Ooi et al., 2001; Yang & Purves, 2003), a near-past and distant-future association could resemble a low-past and high-future association on a two-dimensional screen (Stocker et al., 2016). Thus, the putatively 'vertical' associations reported may actually be sagittal representations in disguise (Winter et al., 2015). Indeed, one study showed that Western participants are quicker at categorising verbal stimuli as past- versus future-related when experiencing

[5] The terms 'Western' and 'Westerners' are used deliberately, as some studies suggest that Mandarin Chinese speakers have vertical time representations (e.g., Boroditsky, 2001; Boroditsky et al., 2011; Miles et al., 2011).

backward and forward motion, whereas vertical movement did not have a significant effect (Hartmann & Mast, 2012). Null or inconsistent vertical effects have also been reported by other studies (e.g., Dalmaso et al., 2022; Fuhrman et al., 2011; Ishihara et al., 2008; Miles et al., 2011; Woodin & Winter, 2018). As a consequence, it is unclear whether Westerners do in fact conceptualise time along the vertical axis; at any rate, vertical representations of time are reported less consistently than horizontal ones.

2.3. Metaphoric Gestures

In the previous section, we saw some conceptual metaphors (horizontal metaphors of numerical quantity, time, and emotional valence; vertical metaphors of time) are not expressed via linguistic metaphors at all, implying that linguistic metaphors are just one outlet through which conceptual metaphors may be expressed. This chapter introduces gesture as another communicative strategy via which conceptual metaphors can be expressed. These *metaphoric gestures* are the focus of Paper 1, which looks at size-based and horizontal metaphoric gestures of numerical magnitude. Paper 1 explores metaphoric gestures that are produced alongside linguistic metaphors, where the gesture reflects the semantic content of the metaphor. It also investigates whether speakers perform gestures that express a metaphor other than the one mentioned in their speech. Both these types of metaphoric gesture are discussed in this chapter, in addition to metaphoric gestures that are produced in the absence of metaphorical language. Moreover, this chapter explores metaphoric gestures as a source of additional evidence for the conceptual metaphors studied in Paper 2: vertical and horizontal metaphors of numerical magnitude, time, and emotional valence. As discussed in Chapter 1, metaphoric gestures are also a subtype of

iconic gestures, where the source domain iconically depicted in the gesture stands metaphorically for a target domain.

Some scholars have argued that linguistic metaphors are so conventionalised that interlocutors may sometimes understand them as literal rather than as metaphorical (e.g., Black, 1993; Hanks, 2006). For instance, when describing a number as 'small', speakers may not be consciously thinking of the spatial meaning of 'small', but may regard this linguistic metaphor as "just the way to say it" (Cameron, 2003: 100). However, the performance of a metaphoric gesture that matches the meaning of the metaphor may suggest that the speaker is indeed interpreting the linguistic metaphor in a genuinely metaphorical sense (e.g., Cienki & Müller, 2008; Goldin-Meadow et al., 1993; Müller, 2008). In other words, metaphoric gestures may indicate that the metaphor used in someone's speech is *conceptually active* (see Müller, 2008; Woodin, 2019). For example, Figure 2 shows a speaker gesturing by moving his hands close to each other, with the space between his flat palms metaphorically depicting the 'small' size of the number. The presence of this metaphoric gesture along the linguistic metaphor 'small number' may suggest that this metaphor is conceptually active for the speaker. Metaphoric gestures are sometimes interpreted as stronger evidence for conceptual metaphors than linguistic metaphors because metaphoric (and iconic) gestures may be produced less deliberately and more spontaneously than language, providing a window into how the speaker is conceptualising the topics being discussed (McNeill, 1992, 2000, 2005).



Figure 2. RT presenter performs a metaphoric gesture while using the linguistic metaphor 'small number'. This gesture may indicate that the linguistic metaphor 'small number' is conceptually active for this speaker. White arrows have been added to this image to illustrate the movement direction of the gesture. Source: RT, accessed via the TV News Archive: https://tinyurl.com/ycxzhwhs.

The following discussion should, however, be prefaced by acknowledging that the relationship between language and thought may be more complicated than the line of reasoning presented so far implies. As de Ruiter (2007) discusses, gestures may not be unadulterated expressions of thought, but may be influenced by factors such as the language one speaks (e.g., Kita & Özyürek, 2003), the concurrent planning of speech and

gesture (e.g., Melinger & Levelt, 2004), and concerns about how to make one's gestural communication maximally interpretable by one's addressee (e.g., Bavelas et al., 2002). Other research on multimodal construction grammar (e.g., Hinnell, 2018; Steen & Turner, 2013; Zima & Bergs, 2017) has discussed the idea that certain speech-gesture combinations may be stored as a single multimodal construction in the brain, rather than directly expressing the conceptualisation processes underlying language production. Furthermore, it should be acknowledged that some gestures are highly conventionalized, such as emblems like the 'OK' gesture (e.g., Kendon, 2004), but the focus here is on metaphoric and iconic gestures that communicate imagistic content, which McNeill (1992) argues are unconventionalized. However, it is likely that even metaphoric and iconic gestures differ in their degree to which they are conventionalized, although we cannot say for sure *how* conventionalized they are. Nonetheless, this thesis subscribes to the notion that metaphoric and iconic gestures are expressions of psychological processes, notwithstanding the idea that there are other relevant factors that may influence these gestures.

Gesture studies have discussed examples of metaphoric gestures pertaining to numerical magnitude. For example, Winter et al. (2013) discuss a size-based gesture performed by a speaker accompanying the phrase 'tiny number' in which the speaker gestures with her forefinger and thumb held close together, as if holding the number between her fingertips. The same speaker also uses the metaphor 'huge amount', and in doing so performs a size-based gesture in which she moves her hands away from each other along the horizontal axis, with the space between her open palms representing the size of the 'huge' quantity (similar to the gesture shown in Figure 1, see §1). Winter et al.

also discuss examples of vertical gestures in which people gesture with a downward movement when saying ‘low number’ and with an upward movement when saying ‘high number’. They show that speakers may perform vertical gestures even when they do not use vertical linguistic metaphors. For example, one speaker gestures with a downward movement when using the phrase ‘less than’, evidencing an association between lower space and lesser quantities despite not expressing this association linguistically. Also discussed by Winter et al. are gestures that reveal a different conceptual metaphor to the one expressed verbally, with one speaker gesturing with a rightward movement along the horizontal axis while describing numbers using the vertical metaphor ‘going up’. As discussed in chapter 2.2, this horizontal conceptual metaphor is not usually expressed linguistically, but these metaphoric gestures suggest that it can still be conceptually active during language production, and can be expressed simultaneously with other conceptual metaphors (e.g., vertical number metaphors).

Researchers have also reported evidence for metaphoric gestures of emotional valence. For instance, Woodin (2019) found that speakers tend to perform gestures with upward and downward movements respectively when using the linguistic metaphors ‘raise the standard’ and ‘lower the standard’, which describe valence changes – improvement (positive) and deterioration (negative) of quality – in vertical terms. Figure 3 shows an example of one of these gestures, where the speaker uses the phrase ‘raise the standards’ while moving his hand upward, the open palm of this hand facing downward. For the horizontal axis, consistent with the body-specificity hypothesis, Casasanto and Jasmin (2010) reported that left-handed political candidates gesture more frequently with their left hand during positively valenced clauses, and with their right hand during negatively

valenced clauses. The opposite pattern was found for right-handed candidates. As discussed in chapter 2.2, this horizontal conceptual metaphor of emotional valence is generally unproductive in language (aside from, e.g., 'right' being a synonym for 'correct'; Casasanto, 2009), so gesture may be the only outlet available for this conceptual metaphor to be expressed during verbal communication, offering a non-verbal indication that it is conceptually active.

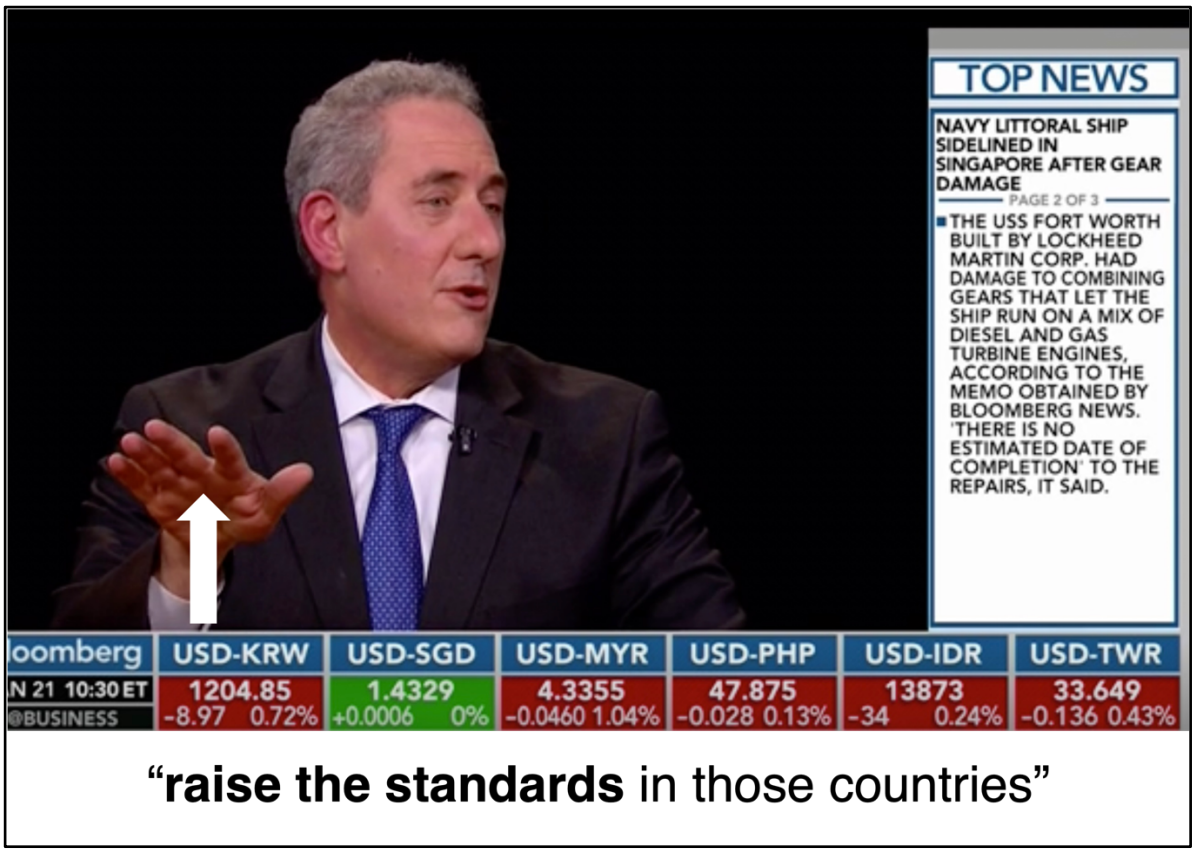


Figure 3. Michael Froman performs a metaphoric gesture while using the linguistic metaphor 'raise the standards'. This gesture may indicate that Froman is conceptualising emotional valence in terms of vertical space. A white arrow has been added to this image to illustrate the movement direction of the gesture. Source: Bloomberg, Charlie Rose, accessed via the TV News Archive: https://tinyurl.com/yuk6yx2e.

Investigating metaphoric gestures of time, Cienki and Müller (2008) discuss a speaker who gestures from left to right when talking about two events that were earlier (preparing to take an exam) and later (taking the exam) in time (see also Walker & Cooperrider, 2016). Speakers also sometimes gesture with a horizontal movement even when using sagittal linguistic metaphors, such as when describing earlier times using the word 'back' with an accompanying leftward gestural movement (Casasanto & Jasmin, 2012). Alternatively, speakers sometimes combine the horizontal and sagittal axes in a diagonal movement, gesturing with a leftward-backward movement when discussing the past, and a rightward-forward movement when talking about the future (Walker & Cooperrider, 2016). Studies tend to show that vertical gestures are less common than horizontal or sagittal ones and/or are excluded from the main analyses (Casasanto & Jasmin, 2012; Walker & Cooperrider, 2016), and in general, these vertical gestures have received less attention because vertical conceptual metaphors of time are not as well-attested when compared with horizontal conceptual metaphors.

Gestures that depict the metaphor used in speech are often considered to be metaphoric. However, Bouissac (2008) argues that these gestures only imitate speech, making them iconically imitative of speech, but not metaphoric. In instances where a linguistic metaphor is used, the linguistic metaphors themselves could prime the performance of metaphoric gestures in some instances. However, this cannot be the full story, because there are examples of metaphoric gestures that are performed without speakers expressing these metaphors linguistically, and indeed where linguistic representation is not conventional in English (Casasanto & Jasmin, 2010, 2012; Cienki & Müller, 2008; Walker & Cooperrider, 2016; Winter et al., 2013; see also Calbris, 1990;

Cienki, 1998). For instance, Walker and Cooperrider (2016) cite an example of a speaker saying ‘looking to the future’ while moving his right hand in a rightward direction, where speech does not contain spatial language, but the gesture nonetheless has a horizontal movement.

According to Cienki and Muller (2008), there are four ways in which speech and gesture correspond to conceptual metaphors (1–4). These four categories have been supplemented by two more categories (5–6) proposed by Winter et al. (2013). The six categories are as follows:

1. Both speech and gesture express the conceptual metaphor, e.g., Michael Froman gesturing upward while saying ‘raise the standards’ (see Figure 3).
2. Only speech expresses the conceptual metaphor, e.g., speakers using the phrase ‘high number’ while not gesturing, or performing a gesture that is incompatible with the metaphor or any other conceptual metaphor (e.g., Woodin, 2019).
3. Only gesture expresses the conceptual metaphor, e.g., speakers gesturing with their left hand during positively valenced clauses and with their right hand during negatively valenced clauses, while not expressing this horizontal metaphor verbally (indeed, there are no conventional linguistic metaphors available to express this conceptual metaphor) (Casasanto & Jasmin, 2010).
4. Speech and gesture express different conceptual metaphors, e.g., speakers gesturing with a horizontal movement when describing earlier times with the sagittal term ‘back’ (Casasanto & Jasmin, 2012).

5. Gesture expresses different metaphors consecutively, e.g., speakers performing a rightward gesture and then performing an outward ‘expansion’ gesture reflecting large physical size (Winter et al., 2013).
6. Gesture expresses multiple metaphors at the same time, e.g., speakers gesturing with a leftward-backward movement when discussing the past, and a rightward-forward movement when discussing the future (Walker & Cooperrider, 2016).

2.4. Conceptual Metaphors in Cultural Artifacts

The previous chapters have framed language and gesture as two outlets through which conceptual metaphors are expressed. Another outlet via which conceptual metaphors are represented is cultural artifacts, such as number lines, timelines, and data visualizations. These cultural artifacts represent size-based (Paper 1) and horizontal (Papers 1 and 2) number metaphors, vertical number metaphors (Paper 2), and horizontal time metaphors (Paper 2). Paper 2 explores the idea that these conceptual metaphors laid the groundwork for the conventions that have emerged in the field of data visualization design, with regard to the representation of numerical magnitude and time. It also tests whether horizontal and vertical conceptual metaphors of emotional valence influence how people interpret data visualizations, despite valence metaphors not conventionally being represented in this medium.

Cultural artifacts representing conceptual metaphors have their roots in cultural practices that have existed for millennia. Precursors to metaphoric cultural artifacts have been discovered that date back to 75,000 BP (beads) and 28,000 BP (a stick with notches carved into it), which appear to be items with which past cultures counted numerical

information (discussed by Coolidge & Overmann, 2012; Overmann, 2013). According to Zhang and Norman (1995, p. 274), the “first numeration systems invented in nearly all nations of antiquity were […] systems represented by simple physical objects, such as stones, pebbles, sticks, tallies, etc.”. These nascent numerical representations were necessarily spatial, as they predate symbolic communication of numerical information afforded by the invention of numerals circa 3200 BCE. Other spatial-numerical representations include finger counting, written tallies, and abaci (De Cruz, 2008). Representing concepts externally in a physical, spatial medium offloads mental computations onto the physical world, perhaps making these concepts easier to process, remember, manipulate, and communicate.

Counting beads and message sticks are only spatial insofar as mapping concepts onto physical objects inherently involves a spatial relationship between the objects. In contrast, many cultural artifacts intentionally represent properties of concepts with spatial parameters (see Figure 4). For example, pie charts (Fig. 4A) map quantity onto size, with larger slices representing larger quantities, while vertical and horizontal bar charts map quantity onto upward (Fig. 4B) and rightward (Fig. 4C) extent (see Fischer et al., 2005). Similar to horizontal bar charts, number lines (Fig. 4D) represent differences in quantity along the horizontal axis, as do timelines (Fig. 4E) in the representation of time (Tversky, 2011; Woodin & Winter, 2018). Calendars (Fig. 4F) represent days progressing from left to right and weeks from top to bottom, while analogue clocks (Figure 4G) represent time as passing in a circular direction, rotating rightward (see Duffy, 2014). In time series charts (Figure 4H), horizontal representations of time and vertical representations of quantity are used in the same chart, showing changes in quantity (y-axis) across time (x-axis). These

cultural artifacts are consistent with size-based metaphors of numerical magnitude (pie charts), horizontal metaphors of numerical quantity (horizontal bar charts, number lines), vertical metaphors of numerical quantity (vertical bar charts, time series graphs), and horizontal metaphors of time (timelines, calendars, clocks[6], time series graphs).

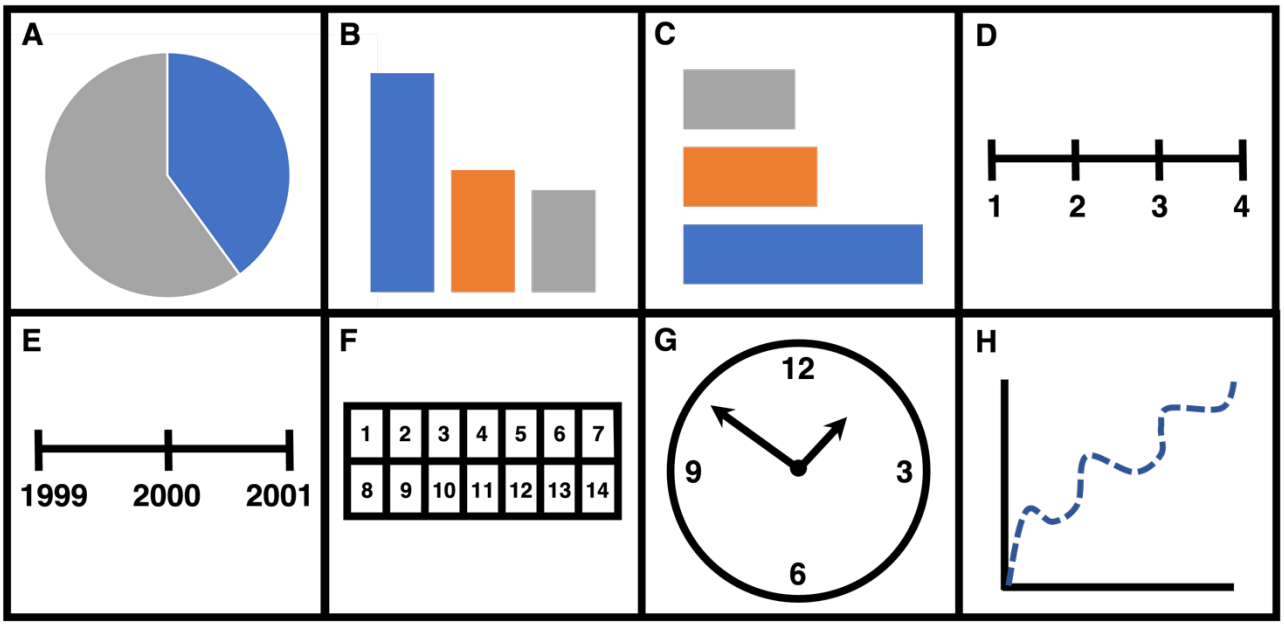


Figure 4. Examples of cultural artifacts that represent numerical quantity and time spatially: pie charts (A), vertical bar charts (B), horizontal bar charts (C) number lines (D), timelines (E), calendars (F), analogue clocks (G), and time series charts (H).

Designing data visualizations to be consistent with conceptual metaphors may make these visualizations easier to understand (see Parsons, 2018). For example, a graph

[6] Clocks do not show an axial horizontal representation, but the hands on a clock nonetheless turn in a rightward direction: from the perspective of the hands themselves, facing outward from the centre, an anticlockwise turn is in a leftward direction, and a clockwise turn is in a rightward direction.

showing numerical quantity increasing up the y-axis may be easier to interpret than a graph that shows numerical quantity increasing from top to bottom. This argument can be understood through the lens of *cognitive fit theory*, which concerns the cognitive effort needed to interpret different presentation formats of information (Vessey, 1991). According to cognitive fit theory, the strategy people use to understand information is based on a mental model of what they expect to see (Padilla et al., 2018), which may be motivated in part by conceptual metaphors. For instance, a metaphorical association between upward space and greater quantities may motivate an expectation that quantities will increase up the y-axis. When a graph subverts this expectation – for example, by reversing its y-axis so greater numbers appear lower down – graph users may have to mentally transform the graph so it conforms with their mental model, in this case by re-ordering the y-axis. This mental transformation may incur a processing cost, making the graph more difficult to understand and potentially eliciting incorrect conclusions about the information depicted. Initial evidence confirms that inverting the y-axis makes graphs more difficult to interpret (Pandey et al., 2015). The potential value of factoring conceptual metaphors into the design of cultural artifacts may explain the conventions that have emerged in fields like data visualization (see Cox, 2004; Fishwick, 2006; Tversky, 2011).

2.5. Origins of Conceptual Metaphors

So far, I have argued that conceptual metaphors have primacy and are then expressed through language, gesture, and cultural artifacts, rather than linguistic metaphors creating the associations from scratch. This begs the question of from where conceptual metaphors originate, if not from language. This chapter argues that the conceptual metaphors

investigated in Papers 1 and 2 often originate via the mechanism of Hebbian learning – although there may be a different mechanism for the size-based conceptual metaphors of numerical magnitude studied in Paper 1. Furthermore, as we will see, investigating the origins of conceptual metaphors casts more doubt on the vertical conceptual metaphor of time studied in Paper 2. This chapter also makes the argument that the expression of conceptual metaphors through language, gesture, and cultural artifacts can motivate or reinforce these same conceptual metaphors.

Where do conceptual metaphors come from? One possibility is that they are acquired by observing domains that are correlated in experience and learning to associate these domains with one other (e.g., Casasanto, 2014; Pitt & Casasanto, 2020). One mechanism by which these associations might be formed is through *Hebbian learning*, a term that describes the idea that 'neurons that fire together wire together' (Hebb, 1949; see Winter et al., 2015 for a discussion in relation to conceptual metaphors of numerical quantity). The human brain is geared toward pattern recognition, which allows us to make associations and make predictions based on these associations. When one domain (e.g., numerical magnitude) is associated with another (e.g., space), neurons encoding both of these domains fire in unison every time we experience this association. Through repeated neuronal firing, both groups of neurons become 'wired' together, and thereafter, experiencing one domain in isolation triggers the firing of neurons associated with the other domain. When one domain is abstract and the other concrete, this process leads to the formation of what we call conceptual metaphors.

Hebbian learning may be implicated in the emergence of vertical conceptual metaphors of numerical quantity. It is thought that vertical conceptual metaphors emerge

via exposure to environmental correlations: as noted by Lakoff (1987), adding water to a glass makes the water level of the glass rise, while tipping water out of the glass makes the level fall. Similarly, adding stones to a pile increases its height, whereas the opposite happens if we remove stones from the pile. By noticing this correlation, we come to associate upper space with greater quantities, and lower space with lesser quantities. The vertical valence metaphor may also stem from noticing correlations in experience and Hebbian learning taking place. For instance, people are likely to assume a 'low', slumped body posture when experiencing negative emotions (i.e., when we are 'feeling down'), whereas they may assume a 'high', upright posture when happy and confident (i.e., when 'on a high') (Lakoff & Johnson, 1999). Moreover, to be smaller and so in a lower position than another person is usually to be physically inferior, whereas to be taller and so in a higher position usually allows one to exert more control (see discussion in Winter, 2014). Observing this correlation in others and ourselves may lead us to associate upper space with positive valence, and lower space with negative valence.

The formation of conceptual metaphors of emotional valence may also involve Hebbian learning. The horizontal valence metaphor may stem from the positively valenced experience of motor fluency with the dominant side of the body (Casasanto, 2009). Supporting this idea, Casasanto and Chrysikou (2011) found that patients with a disabled left hand who were right-handed prior to having a stroke continue to associate positive valence with right-side space, while previously right-handed patients with a disabled right hand come to associate positive valence with left-side space. These authors also reported that right-handers switched from right-is-good to left-is-good mappings after wearing a cumbersome glove on their right hand for about twelve minutes (for a review, see

Casasanto, 2011). Moreover, Chen and Lin (2021) found that right-handed people express higher intention to purchase food presented with tableware to its right, whereas this effect reverses when their right hand is under high motor resource load. Therefore, a bodily-experienced correlation between fluency and either side of the body leads people to associate the right or left side of space with positive valence, and vice versa for negative valence.

Horizontal representations of time and numerical quantity may be motivated by culture-specific reading and writing conventions, which are another source from which Hebbian associations may be formed. This putative origin of horizontal mappings is supported by evidence showing that Westerners asked to arrange cards in a temporal sequence prefer a light-to-right mapping (Bergen & Chan Lau, 2012; Boroditsky & Gaby, 2010; Tversky et al., 1991), while speakers of Hebrew and Arabic (which are both written from right to left) prefer a leftward mapping (Fuhrman & Boroditsky, 2010; Tversky et al., 1991). In another study, Casasanto and Bottini (2014a) showed that Dutch participants' reaction times were consistent with a left-to-right mental timeline when judging phrases written in standard orthography, whereas their mental timelines were reversed after brief exposure to mirror-reversed orthography. Autry et al. (2020) also demonstrated that the tendency to place stickers depicting a temporal sequence (breakfast, lunch, dinner) in a rightward direction is stronger in older, English-speaking kindergarteners (mean age = 6.05 years) than younger, English-speaking pre-schoolers (mean age = 3.87 years), and is correlated with emergent literacy skills. Similarly, another study reported that the SNARC effect is reversed in Arabic monoliterates, weakened in Arabic-English biliterates, and absent in illiterate Arabic speakers (Zebian, 2005), showing a graded influence of literacy

on horizontal conceptual metaphors of numerical quantity. Other research has corroborated the idea that the SNARC effect is culturally relative (e.g., Chatterjee et al., 1999; Padakannaya et al., 2002; Shaki & Fischer, 2008; Shaki et al., 2009; Azhar et al., 2021; see Göbel et al., 2011 for a review).

Regarding the vertical conceptual metaphor of time, chapter 2.2 discussed the idea that it could be explained by a transformation of the sagittal axis onto a two-dimensional plane (Stocker et al., 2016), making it a sagittal mapping in disguise rather than a vertical mapping proper (Winter et al., 2015) (see §2.2). If true, the ultimate source of these putatively vertical (but actually sagittal) mappings may be the embodied experience of moving through space, or observing objects travelling through space, and experiencing a positive correlation between distance travelled and time taken (Casasanto, 2014). Experiencing this correlation could motivate a Hebbian association between time and sagittal space. Regarding what correlations could directly motivate a Hebbian association between lower space and the past, and upper space and the future, physical growth (correlated with the passing of time) is associated with higher vertical positions (e.g., Cooperrider et al., 2022; Dancygier & Sweetser, 2014; Forker, 2020; see also Safar, 2020), as can be observed in the growth of a plant or a teenager during a growth spurt.

When it comes to size-based conceptual metaphors, numerical quantities may become associated with larger sizes because more numerous sets often span a greater extent, which may lead us to conflate numerosity with physical size (see Hurewitz et al., 2006), creating a Hebbian association between physical size and numerical magnitude. Alternatively, de Hevia (2021) argues that this association may be innate, and hence may be heritable rather than being acquired through Hebbian learning. In making this argument,

de Hevia cites research showing that neonates up to three days old, who have had extremely limited experience with which to form new associations, nevertheless evidence the association between size and number. These neonates preferred to look at a large rather than a small object after experiencing an increase in number (hearing six syllables, then eighteen syllables), and vice versa for decreases in number. If this association is functional from birth, it would speak to the evolutionary advantage of being able to make approximate assessments of number based on size, due to these domains being highly correlated.

As discussed, once established, conceptual metaphors come to be represented through linguistic metaphors (§2.1), metaphoric gestures (§2.3), and/or cultural artifacts, such as data visualizations (§2.4). These outward expressions of conceptual metaphors then become new correlations that serve as new input for the formation of conceptual metaphors, or the reinforcement of existing ones in what Winter (2014) and Winter and Matlock (2017) characterise as a cultural feedback loop. For example, histograms in which 'more' is represented by bars with a larger area (see Tversky, 2011) may reinforce the association between size and number. Moreover, the thumbs up/down gesture might reinforce the association between upper space and positive valence, and the cultural convention of raising one's right hand when taking an oath may reinforce the association between right-side space and positive valence in right-handers, or even motivate this association in left-handers (see Casasanto, 2017; Casasanto & Bottini, 2014b). Linguistic metaphors like 'Petrol prices are stratospheric' or 'The pound has hit an all-time low' may also *reinforce* conceptual metaphors – in this case, the vertical conceptual metaphor of numerical quantity. In other instances, linguistic metaphors may strengthen particular conceptual metaphors at the expense of others. For example, Dutch speakers tend to

discuss musical pitch in vertical terms ('low' and 'high'), while adult Farsi speakers tend to discuss pitch in thickness terms ('thick' and 'thin') (Dolscheid et al., 2013). Despite these differences, research with prelinguistic infants shows that both vertical and thickness conceptual metaphors are present cross-culturally, with Dutch infants tending to look longer at stimuli that are congruent with both verticality and thickness metaphorical mappings (Dolscheid et al., 2012). Therefore, the conceptual metaphor expressed in one's native language using linguistic metaphors may be strengthened over time and become the dominant conceptual metaphor in that particular culture. It is plausible that metaphoric gestures and cultural artifacts have a similar strengthening effect on the conceptual metaphors they represent. Hebbian learning may be at the heart of the strengthening of conceptual metaphors described here, with repeated exposure to linguistic metaphors, metaphoric gestures, and cultural artifacts 'wiring together' groups of neurons associated with different domains more strongly.

2.6. Number Metaphors and Numerical Communication

Numerical magnitude has featured prominently in the discussion of conceptual metaphors so far, and is a primary topic of interest in this thesis, being investigated in both Papers 1 and 2. Paper 3 explores numerical communication more generally, looking into the frequency with which different numbers are used in English, and tying these findings to research in numerical cognition. This chapter discusses additional research into number metaphors as a way of contextualising the research presented in Paper 3, and making explicit the connection of Paper 3 to the other papers presented in this thesis. Specifically,

this chapter introduces the communicative practice of rounding, which is one of the factors investigated in Paper 3 that affect number frequencies.

As discussed in previous chapters, empirical evidence shows that people often conceptualise numbers on a mental number line, which is oriented from left to right in Western cultures. This mental number line is furthermore proposed to be logarithmically scaled (e.g., Dehaene, 1992). Therefore, the space between greater numbers on this number line is smaller than the space between smaller numbers – there is more compression on the number line at greater magnitudes (see Figure 5). The logarithmic scaling of the mental number line may explain other experimental findings in numerical cognition. For example, the *size effect* shows that, for a given difference in magnitude, comparing greater numbers is harder than comparing lesser numbers (e.g., Merten & Nieder, 2009; Moyer & Landauer, 1967; Nieder & Miller, 2003). One possible explanation for this finding is that greater numbers are harder to discriminate because they are closer together on the mental number line, so a greater difference in magnitude is needed to compare the magnitudes of greater numbers. This notion aligns with the Weber-Fechner law: as the intensity of a stimulus (in this case, the magnitude of a number) increases, a greater difference in intensity is needed for us to notice the increase (DeWind et al., 2015; Shepard et al., 1975). The Weber-Fechner law shows that our ability to quantify sets becomes more imprecise at higher magnitudes. Other potential evidence for logarithmic scaling of the mental number line comes from random number generation tasks showing that lesser digits (1, 2, or 3) are used more often than greater digits (4, 5, or 6). In these tasks, the use of smaller digits is associated with leftward attentional biases, as if looking at the location of numbers on the mental number line (Loetscher & Brugger, 2007). Smaller

numbers may be selected more often in these tasks because they are further apart on the mental number line, and so are represented more distinctly (see Dehaene, 1992).

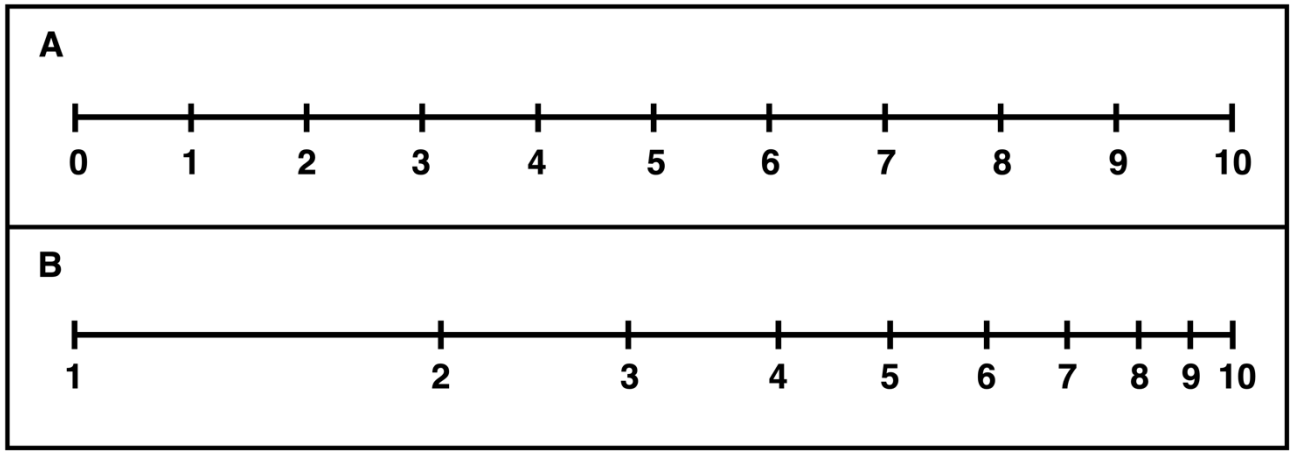


Figure 5. A comparison of number lines up to 10 that are linearly scaled (A) versus logarithmically scaled (B). On the logarithmic number line, greater numbers are closer together than lesser numbers – they are logarithmically compressed.

The logarithmic representation of numbers may affect the frequency with which different numbers are used. More specifically, it may influence the linguistic practice of *rounding*. In decimal systems (i.e., based on the number ten, tenths, and powers of ten), round numbers are multiples of ten and sometimes five (e.g., Dorogovtsev et al., 2005; Sigurd, 1988). Rounding is when the nearest *round number* is used in place of the real number (Cummins, 2015, p. 20). For example, rather than say there are 491 animals at Twycross Zoo, a speaker may round this number up to 500. Alternatively, round numbers may be used as a benchmark indicating a level of completion (Dehaene, 1997, Ch. 4; Gunasti, 2016). For instance, 491 animals may be framed as being ‘almost 500’, while 506 animals might be framed as ‘over 500’. In other contexts, a round number may be used

when the exact value is unknown (Ruud et al., 2014), such as when guessing the number of pebbles on a beach. If our mental representation of numbers becomes less precise at higher magnitudes, due to a logarithmic representation of numbers, we may round to a greater extent at higher magnitudes – for instance, rounding 77 up to 80, but rounding 177 up to 200 (rather than 180). Hence, the increasing imprecision of numerical cognition at higher magnitudes may yield an increasingly imprecise use of numbers at higher magnitudes, evident in rounding practices.

2.7. Iconicity in Multimodal Communication

This literature review has thus far focused on metaphor. As discussed in Chapter 1, metaphor is related to the phenomenon of *iconicity* – the phenomenon whereby communicative forms resemble their meanings (see Peirce, 1931). Paper 4 explores the historical evolution of iconicity, its expression via words, gestures, non-linguistic vocalizations, and prosodic modulations, and its role in language evolution. The focus of this chapter is on the expression of iconicity in different modalities: auditory (words, vocalizations, prosody) and visuospatial (gestures).

Historically in linguistics, iconicity has been downplayed as a feature of language, with an emphasis placed on so-called 'arbitrariness', a term often used to refer to the lack of a resemblance between the form of a word or sign and its meaning (e.g., Levelt et al., 1999; Newmeyer, 1992; Pinker & Bloom, 1990). For example, Saussure (1966) notes that the English word 'tree' is very different from the Latin word 'arbor', yet both words have the same meaning (see also Locke, 1689). The mappings between these words and their meanings are not iconic, and so must be established by language-specific convention. If a

word is not iconic, the assumption is that the choice to use the word is unmotivated and random, and so is described as ‘arbitrary’ (but see Paper 4, §6 for an argument that this term is not accurate). Hockett (1960) even proposes arbitrariness as a defining feature of human languages that sets them apart from animal communication systems. Indeed, the historical treatment of signed languages as pantomimes or elaborate systems of manual gestures, rather than as true languages, has been attributed partly to their evident iconicity (see Goldin-Meadow & Brentari, 2017). For instance, the American Sign Language sign ‘book’ depicts its meaning with the hands held open, facing upward and adjacent to one another, clearly representing the pages of an open book. Despite the lack of attention paid to iconicity historically, scholars have begun to argue that iconicity is another significant design feature of languages, including spoken languages (e.g., Dingemanse et al., 2015; Lockwood & Dingemanse, 2015; Perniss et al., 2010; Perniss & Vigliocco, 2014; Vinson et al., 2021).

Iconicity is prevalent in spoken languages. In English, onomatopoeic words like ‘splash’, ‘bang’ and ‘peep’ resemble the sounds they denote: something falling into or striking a liquid (e.g., a swimmer diving into a pool), a sudden loud, sharp noise (e.g. an explosion), and a brief, high-pitched noise (e.g., a whistle), respectively (see Bladon, 1977; Taylor, 2007). Some spoken languages also have a grammatical category called ideophones, alternatively known as mimetics or expressives (Akita & Pardeshi, 2019; Dingemanse, 2018; Voeltz & Kilian-Hatz, 2001), which feature words that iconically represent sound and other modalities. For example, in Japanese, the ideophone ‘nyaa’ mimics the sound made by a cat (Kita, 1997) and ‘dosadosa’ refers to a thudding sound (Akita, 2017), while other ideophones denote non-auditory events that may nonetheless

generate sound (e.g., ‘nurunur’: tactile sensation caused by a slimy object) or may not generate sound at all (e.g., ‘kutakuta’: being very tired) (Kita, 1997). Even in spoken languages without a dedicated class of ideophones, like English (Nuckolls, 1999), some sounds are used systematically for certain kinds of meanings. For example, Winter et al. (2022) show that /r/ systematically appears in words for ‘rough’ meanings in English and other languages in a sample of 332 spoken languages from 84 phyla. The authors link the ‘rough’ meaning of /r/ to the ‘rough’ sound of its trilled variant. This phoneme is no longer trilled in most English dialects, but the association between /r/ and roughness remains a feature of the lexicon. Sidhu et al. (2021) also report that phonemes associated with roundness (/b/, /u/, /m/) tend to appear in English words for round objects (e.g., ‘ball’ /bɔːl/, ‘hoop’ /huːp/, ‘moon’ /muːn/), while phonemes associated with spikiness (/t/, /k/, /ɪ/) tend to appear in English words for spiky objects (e.g., ‘cactus’ /kæktəs/, ‘spike’ /spaɪk/, ‘pitchfork’ /pɪtʃfɔːk/) (see also D’Onofrio, 2014; McCormick et al., 2015; Westbury et al., 2018).

Iconicity is also present in aspects of communication that are not conventionalised words, such as iconic gestures (e.g., McNeill, 1992, 2000, 2005). For example, Figure 6 shows TV personality Steve Higgins performing an iconic gesture with his left hand that depicts a “swinging punching bag”, with the arc of the movement back and forth depicting the “swinging” motion. To give another example, a speaker might depict ‘cooking’ by gesturing as if cooking with a frying pan. Speakers may also communicate iconically with vocal prosody, which encompasses modulations of pitch, speed, rhythm, and timbre (e.g., Ćwiek & Fuchs, 2019; Nygaard et al., 2009; Perlman, Clark, et al., 2015). For example, Winter et al. (2022) describe how the /r/ phoneme in US potato crisp brand Ruffles was trilled in its historical marketing slogan, emphasising the ‘roughness’ of the word to reflect

the ridged texture of the crisp. Therefore, although the trilled /r/ variant is not used in US English accents, the /r/ phoneme can be prosodically modulated to enhance its iconic connection to roughness. Moreover, English speakers sometimes lengthen the vowel in ‘long’, pronouncing it as ‘looooong’ to represent the length of time denoted (e.g., Guerrini, 2020). Finally, speakers use iconic, non-linguistic vocalizations that depict sound using imitative, non-verbal vocal noises (Edmiston et al., 2018; Lemaitre et al., 2016; Sarvasy, 2016). For instance, Lemaitre et al. (2014) report that a speaker of English depicted the unrolling of kitchen roll with the vocalization ‘fffff’, and then depicted detaching it from the cardboard tube with the vocalization ‘rrrrrrrr’. Perlman and Cain (2014) also discuss an example in which a speaker refers to an ‘elephant’ using a vocalization that imitates the roar of an elephant.



Figure 6. Steve Higgins performs an iconic gesture depicting the “swinging” motion of a punching bag. White arrows have been added to this image to illustrate the movement direction of the gesture. Source: NBC, The Tonight Show Starring Jimmy Fallon, accessed via the TV News Archive: https://tinyurl.com/pxd5fpkh.

Multiple iconic strategies – words, gesture, prosody, and vocalization – can be employed at the same time. For example, Figure 7 shows an example in which TV personality Beth Troutman talks about an “itty-bitty man” in a photograph who is standing on the top of a large rock, a distance away from the camera. ‘Itty-bitty’ /ɪtɪbɪti/ is an conventionalized expression containing the high front vowel /ɪ/, which tends to appear in words denoting ‘small’ meanings (e.g., Blasi et al., 2016; Ultan, 1978; Winter & Perlman, 2021). The high frequency of this vowel iconically depicts the high-pitched sounds that

smaller entities tend to make (Fitch, 1994; Ohala, 1994; Winter et al., 2021). When saying “itty-bitty”, Troutman performs an iconic gesture where her forefinger and thumb approach one another, as if she is holding the “itty bitty man” between her fingers, or otherwise representing the size of the man with the space between these fingers (see Hassemer & Winter, 2018; Winter et al., 2013). At the same time, Troutman uses iconic prosody by raising her vocal pitch, enhancing the iconic relation between pitch and physical size that was already expressed through the use of the phrase “itty-bitty”. In this example, then, iconic words, gesture, and prosody are all utilised at the same time.



Figure 7. Beth Troutman uses the iconic expression “itty-bitty” while raising her vocal pitch (iconic prosody) and pinching her thumb and forefinger together (iconic gesture). This example shows how multiple iconic strategies can be utilised at the same time to create a

depictive performance. Source: ABC, Right This Minute, accessed via the TV News Archive: https://tinyurl.com/y9kjpom9.

We can see, then, that speakers have different iconic strategies (words, prosody, vocalization, gesture) in different modalities (auditory, visuospatial) available to express sensorimotor information. Speakers sometimes utilise more than one outlet at the same time, creating an imagined experience (see Pascual, 2014) that allows their audience to picture "what it would be like to hear, see, or feel what the original speaker [or signer] did" (Wade & Clark, 1993, p. 188). In this way, speakers opportunistically use all the iconic strategies at their disposal for the purpose of depiction.

2.8. Sensorimotor Simulation in Language Comprehension

The use of iconicity to depict sensorimotor information leads us to a discussion of sensorimotor simulation and the part it plays in language comprehension in this chapter. The function of sensorimotor simulation in the comprehension and production of iconic communicative forms is central to the argument made in Paper 4, which connects this psychological mechanism to historical processes of language change and the evolution of iconic words over time.

Sensorimotor simulation is the mental reactivation of previously experienced perceptions and actions (e.g., Barsalou, 1999, 2007; Decety & Jeannerod, 1995). It is often seen as a more automatic, less conscious version of mental imagery (e.g., Bergen, 2012; Kosslyn, 1994; Pearson et al., 2015). Sensorimotor simulations allow people to imagine scenes they have not previously experienced directly by combining aspects of their

simulations creatively, allowing them to imagine an infinite number of experiences (Barsalou, 1999; Barsalou & Prinz, 1997). These sensorimotor simulations are proposed to involve similar but weaker patterns of neural activation as those involved during actual perception and action, such that, for example, simulating an action potentiates the performance of that action in experimental tasks, as we will see below.

Evidence for sensorimotor simulation comes from non-linguistic tasks in which participants view objects and have to make a judgment about these objects using a motor response. For example, Tucker and Ellis (1998) showed people images of a teapot and asked participants whether the object was upright or upside-down. Left-handed responses were facilitated when the teapot's handle was presented on its left side, and vice versa for right-handed responses. The authors argue that viewing the teapot caused participants to mentally simulate manipulating it with their left hand, which activated parts of their brain associated with performing that action, facilitating a left-handed response (see also Iani et al., 2019). Similarly, Tucker and Ellis (2001) found that participants were faster when responding to a visually presented object (e.g., a hammer versus a grape) with a hand grip similar to that which they would use to grasp the object (e.g., a more open versus more closed hand shape) (see also Heurley et al., 2020). The potentiation of manual actions by visually presented objects has been shown to persist even 700ms after the object has disappeared from view (Derbyshire et al., 2006).

Sensorimotor simulations may be recruited in language comprehension, such that reading words automatically activates parts of the brain associated with the meaning of those words. For example, when participants read action words like 'kick', 'lick', and 'pick', fMRI scans show increased blood flow (interpreted as increased activation) in the motor

cortex in a somatotopic fashion in which parts of the brain associated with the body parts used in the performance of the corresponding actions are activated (Hauk et al., 2004). Moreover, understanding nouns and verbs has been shown to involve activation of visual and action areas of the brain, respectively (Pulvermüller et al., 1999; see also Moseley et al., 2013), just as nouns tend to denote visual objects (e.g., Jones & Smith, 1993) and verbs tend to denote actions (e.g., Glenberg & Gallese, 2012). Event-related potentials also display different cortical activity for words denoting either sound or action concepts (Popp et al., 2016).

Some research suggests that comprehending iconic words may activate sensorimotor simulations more strongly than non-iconic words. In particular, neurological studies show that comprehending Japanese ideophones activates areas of the brain recruited in sensorimotor processing more strongly than non-ideophones (Kanero et al., 2014) and non-words (Osaka, 2009, 2011; Osaka et al., 2004; Osaka & Osaka, 2009). This finding was also corroborated with a multi-language stimulus set (Revill et al., 2014). Moreover, fMRI scans have shown that hearing iconic Japanese words for animal sounds activates brain areas implicated in the processing of environmental (versus linguistic) sounds more strongly than hearing non-iconic animal names (Hashimoto et al., 2006). Affective iconic words have also been found to generate a blood-oxygen-level-dependent (BOLD) response in the left amygdala, which is implicated in the multimodal representation of emotions (Aryani et al., 2019).

2.9. Sensorimotor Simulation in Iconic Language Production

Based on the link between sensorimotor simulation and the comprehension of iconic vocabulary, this chapter explores the idea that sensorimotor simulation is involved in not only the comprehension of iconic language, but also its production. This idea is central to Paper 4, which posits that sensorimotor simulation triggers speakers to produce iconic gestures, vocalizations, and prosodic modulations of existing words, which express the content of the speaker's simulation. Iconic forms of expression that are communicatively useful are conventionalized and gradually become part of the lexicon.

The notion that sensorimotor simulation is involved in the production of iconic utterances has been most directly explored in relation to gestures. Hostetter and Alibali's (2008, 2018) Gesture as Simulated Action (GSA) framework posits that sensorimotor simulation underlies the production of representational gestures, which includes iconic, metaphoric, and deictic (pointing) gestures (see McNeill, 1992). The main argument of the GSA framework is that simulating visuospatial or motoric information causes activation in the premotor cortex (Rizzolatti et al., 1996), which may then spread to the motor cortex, causing the performance of manual gestures. These gestures depict the sensorimotor simulation that gave rise to the gesture in the first place. For instance, in Figure 8, US talk show host Jimmy Fallon gestures as if revving the engine of a motorbike. According to the GSA framework, Fallon mentally simulated the action of revving a motorbike, triggering patterns of premotor activation similar to those involved in the physical action of revving a motorbike. This premotor activation was sufficiently strong that it spread to Fallon's motor cortex, causing him to produce a 'revving' gesture that expressed the motorbike-revving content of his simulation.



Figure 8. US talk show host Jimmy Fallon performs an iconic gesture depicting the revving of a motorbike while referring to the film name “Vroom”, which references the sound of a vehicle’s engine being revved. A white arrow has been added to this image to illustrate the movement direction of the gesture. Source: NBC, The Tonight Show Starring Jimmy Fallon, accessed via the TV News Archive: https://tinyurl.com/y8sd22ta.

Hostetter and Alibali (2008, 2018) present evidence for the predictions of their framework. One of these predictions is that if sensorimotor simulation motivates the production of representational gestures, we should observe more gestures when communication relies more on simulations than amodal, propositional (i.e., non-imagistic) codes. Supporting this prediction, one study showed that speakers were more likely to gesture when retelling a vignette after watching a cartoon of the vignette compared to

speakers who had only listened to a verbal description, presumably because the cartoon group had more vivid sensorimotor information available to access (Hostetter & Skirving, 2011). In another study, speakers gestured more often when their spatial skills were better than their verbal skills, perhaps because these speakers relied more on visuospatial simulations when communicating (Hostetter & Alibali, 2007). Speakers have also been shown to gesture more frequently during a visuospatial task (arranging shapes to make a picture) than a propositional task (arranging words to make a sentence) (Hostetter & Sullivan, 2011). Together, these studies are compatible with the notion that gesture production increases as communication relies more on simulation processes, providing support for the GSA framework.

As well as gesture, Perlman and Gibbs (2013) argue that simulation can influence speech prosody. Evidence indirectly supports this argument. For example, studies have shown that speakers talk faster when describing faster events (Perlman, 2010) and read stories about smaller entities such as grasshoppers with a higher pitch than stories about larger entities such as elephants (Perlman, Clark, et al., 2015). Participants also produce longer utterances when describing the movement of a dot they had previously seen move slowly across a screen compared to when the dot moved quickly (Shintel et al., 2006). Altogether, these studies suggest that people modify their speech prosody to reflect their sensorimotor imagery, just like people do with their gestures (Hostetter & Alibali, 2008, 2018). Moreover, as mentioned already in Chapter 2.6, speakers can produce imitative vocalizations that appear to depict the contents of their sensorimotor imagery (e.g., Blackwell et al., 2015; Hudson, 1985; Perlman & Lupyan, 2018), similar to the depictive function of iconic gestures. In fact, iconic vocalizations have been described as *vocal*

gestures due to their shared role with manual gestures (Clark et al., 2013; Okrent, 2002; Perlman, 2010). Thus, iconic gesture, prosody, and vocalization may be three communicative strategies through which sensorimotor simulations are expressed.

2.10. Sensorimotor Simulation in Conceptual Metaphor

In this final chapter of the literature review, I argue that sensorimotor simulation is the psychological basis not just of iconicity, but also of conceptual metaphor (see Gibbs & Matlock, 2008). Despite sensorimotor simulation being discussed explicitly in Paper 4 in relation to iconicity specifically, many instances of iconicity discussed in Paper 4 can be interpreted as a kind of metaphor (see Marks, 2014). In addition, the ideas in this chapter are relevant to contextualising the studies on conceptual metaphor and numerical communication presented in Papers 1, 2, and 3. Although these papers do not explicitly investigate these ideas, this chapter adds additional context that can inform the interpretation of their conclusions, which is discussed in more detail in Chapter 7.

In their GSA framework, Hostetter and Alibali (2008, 2018) propose sensorimotor simulation as the mechanism underlying representational gestures, which include metaphoric gestures, as well as iconic ones. The proposal of a common mechanism for iconic and metaphoric gestures is logically sound because, as discussed in Chapter 1, metaphoric gestures are a kind of iconic gesture (see Bouissac, 2008). When a person communicates iconically, they mentally simulate the literal meaning of what is being expressed. When a person communicates metaphorically, their simulation corresponds to the source domain of a conceptual metaphor, which stands metaphorically for a target domain. For example, a speaker might gesturally depict the size of a fish she caught by

moving her open palms away from each other, with the space between the palms representing the size of the fish. In comparison, if the gesture were not about a fish, but rather about 'huge numbers' (see Figure 1, §1), the speaker might still be simulating properties related to physical size, but size would be the source domain that is used to conceptualise the target domain of numerical quantity. Hence, sensorimotor simulation would be involved in the production of this metaphoric gesture.

Sensorimotor simulation may also be involved in the production of metaphoric prosody and vocalizations (see Perlman & Gibbs, 2013). For example, Chapter 2.8 discussed how English speakers lengthen the vowel in 'long', articulating this word as 'loooong' to depict the length of time denoted, which is iconic as the vocalization literally lasts for a 'long' time (e.g., Guerrini, 2020). Alternatively, speakers might opt for the elongated 'loooong' pronunciation to describe the length of a snake, which can be interpreted as a metaphor, as duration (source domain) is used to represent size (target domain) (Perlman & Gibbs, 2013). In both cases, the speaker may simulate the same temporal information, which elicits the same prosodic modulation (i.e., elongation of the vowel). However, in the first case, duration literally represents duration, whereas in the second case, duration metaphorically represents length. In a similar way, a speaker may simulate the loud sound of a car crash, and then express this simulation using a loud vocalization like 'bam!', with volume iconically representing the loudness of the crash. Speakers can also use volume metaphorically: Chapter 1 discussed how speakers use louder vocalizations to depict larger objects and quieter vocalizations to depict smaller objects (Perlman, Paul, et al., 2015). In both instances, the speaker's sensorimotor

simulation would correspond to volume, but in the second instance, the source domain volume would be used to metaphorically represent the target domain physical size.

Simulating a conceptual metaphor's source domain may be the cognitive basis of what it means for a linguistic metaphor to be conceptually active (see Woodin, 2019 for a similar argument). As discussed in Chapter 2.3, linguistic metaphors may be so conventionalized that people do not interpret them as metaphorical (e.g., Black, 1993; Hanks, 2006). Muller (2008) thus proposes metaphoric gestures as "activation indicators" (Müller, 2008, p. 197), showing that speakers are 'thinking metaphorically'. Metaphoric gestures may evidence conceptual activation because these gestures may reveal that the speaker is simulating the source domain of the linguistic metaphor, which cues them to perform the metaphoric gesture. As well as gesture, metaphoric prosody and metaphoric vocalizations may be activation indicators. For example, if the linguistic metaphor 'small number' is conceptually active, the speaker may be simulating physical size, which may cause them to perform a pinch gesture to depict holding the 'small' number between the forefinger and thumb (see Winter et al., 2013 for a discussion of size-based metaphoric gestures). While using this linguistic metaphor, the speaker may also be simulating pitch information (e.g., Perlman et al., 2015), which may cause them to raise the pitch of their voice. The use of both metaphoric gesture and metaphoric prosody may show that the linguistic metaphor 'small number' is conceptually active.

Müller (2008) defines conceptual activation in terms of Cameron's (1999, p. 19) notion of spreading activation through different neurological pathways. Because more or fewer pathways may be activated at one time, Müller (2008, pp. 197–198) argues that conceptual activation is a graded phenomenon, with the production of more (or fewer)

activation indicators demonstrating that the linguistic metaphor is more (or less) conceptually active. In other words, there are “degrees of metaphoricity” (Woodin, 2019). Hence, the use of a metaphoric gesture (e.g., a pinch gesture) together with the use of metaphoric prosody (e.g., a higher vocal pitch) may show that the linguistic metaphor is *more* conceptually active than if just one of these activation indicators were used. In this simulation-based account, Cameron’s (1999, p. 19) idea of spreading activation can be reinterpreted as the simulation of different kinds of information – in this case, visuospatial and auditory – which are expressed with different metaphoric strategies: gestural and prosodic.

Sensorimotor simulation might explain other findings I have discussed in this literature review – for example, cognitive fit theory (Vessey, 1991) and the interpretation of data visualizations described in Chapter 2.4. Cognitive fit theory postulates that people interpret graphs with a mental model of what they expect to see (see Padilla et al., 2018). Within the framework proposed here, this mental model can be reinterpreted as a sensorimotor simulation that is generated in anticipation of seeing a visualization. This simulation may be influenced by conceptual metaphors. If the visualization contradicts the simulation (i.e., if conceptual metaphors are subverted, for instance by reversing the direction of the y-axis), the graph user may have to transform their simulation or generate a new one to interpret the visualization. This transformation or regeneration may incur a processing cost, whereas visualizations that align with the graph user’s sensorimotor simulation may be quicker and easier to interpret. Also, when people design data visualizations, they may generate a sensorimotor simulation that they try to reproduce visually. As this simulation is influenced by conceptual metaphors, the visualization

designed will tend to conform with these metaphors. For these reasons, sensorimotor simulation, as a process involved in the production and comprehension of conceptual metaphors, may have been critically involved in the formation of the graphical conventions that have emerged in this field (see Cox, 2004; Fishwick, 2006; Tversky, 2011).

More speculatively, sensorimotor simulation may also be recruited for numerical cognition, particularly for the horizontal mental number line that is logarithmically scaled, involving compression at higher magnitudes (Dehaene, 1992, 2007), as is described in Chapter 2.6. Comparing the magnitudes of numbers may thus be at least partly a visuospatial task, involving the scanning of sensorimotor simulations and comparing the position of numbers. Because greater numbers are closer together on the mental number line, they are more difficult to visually discriminate, potentially explaining the size effect, which shows that magnitude comparison is harder for greater numbers (e.g., Merten & Nieder, 2009; Moyer & Landauer, 1967; Nieder & Miller, 2003). Simulation may also explain why people use more smaller digits than larger ones in random number generation tasks Simulation may also explain why people use more smaller digits than larger ones in random number generation tasks: people may select from a mentally simulated distribution of numbers on a logarithmically compressed mental number line, on which smaller numbers are further apart and so are represented more distinctly (see Dehaene, 1992). Finally, simulation may explain why people often look to the left when using smaller digits, as if looking at smaller, more leftward digits on a mentally simulated number line (Loetscher & Brugger, 2007). However, it should be noted that there are other explanations for some of these phenomena that do not necessarily involve simulation. For example, as discussed in Paper 3, smaller numbers may generally be used more frequently than larger numbers

because smaller quantities are encountered more often and are more easily countable, which may influence the numbers people use in random number generation tasks (Cummins, 2015, p. 32).

* * *

This literature review began with the idea that linguistic metaphors of numerical quantity, time, and emotional valence reflect conceptual metaphors that structure our thinking about each concept. Experimental, non-linguistic evidence for these conceptual metaphors was presented, which uncovered evidence for conceptual metaphors that are not instantiated in the English language. The conceptual metaphors discussed have the source domain space, including vertical, horizontal, and size-based mappings. The review then explored the expression of these conceptual metaphors via gestures and cultural artifacts, such as data visualizations, before discussing the influence of conceptual metaphor on graph interpretation and graphical conventions. The review then mooted Hebbian learning as the mechanism through which conceptual metaphors are formed, where associations are learnt by observing correlations in experience. Then, the review discussed the idea that the mental number line is logarithmically compressed at higher magnitudes, and explored the relevance of this idea to numerical communication (i.e., rounding). After this, the review investigated iconicity in the vocal (onomatopoeia, ideophones, sound symbolism, non-linguistic vocalization, prosody) and manual (gesture) channels, and described how people combine these iconic strategies. Evidence for the proposed role of sensorimotor simulation in the comprehension and production of iconic forms was then presented. Finally, the review made the argument that sensorimotor

simulation is implicated in conceptual metaphor, and the implications of this view were discussed for the other topics presented in this thesis.

The papers in this thesis investigate metaphor and iconicity in various kinds of multimodal communication, including gestures (Papers 1 and 4), data visualization (Paper 2), vocalization (Paper 4), and prosody (Paper 4). In its treatment of metaphor, this thesis explores metaphors of time (Paper 2), emotional valence (Paper 2), and numerical magnitude (Papers 1 and 2), before investigating numerical cognition and communication more generally (Paper 3). The thesis makes both empirical contributions (Papers 1, 2, and 3) and theoretical contributions (Paper 4).

Paper 1 (§3) investigates the conceptual basis of linguistic metaphors that refer to numerical quantity in terms of physical size (e.g., 'tiny number', 'huge number'). This multimodal corpus study shows that speakers tend to perform gestures that represent the metaphor used in their speech, hinting that speakers are not only talking about numbers using size-based language, but are also conceptualising numbers metaphorically in terms of physical size. This study provides the first large-scale empirical evidence that spatial-numerical associations that have been found in laboratory settings characterise numerical communication 'in the wild'.

Paper 2 (§4) investigates whether the interpretation of line graphs is influenced by conceptual metaphors of emotional valence. The experiment reported suggests that it is, showing that line graphs whose design aligns with metaphors of emotional valence, where 'more' of a positive quantity (e.g., vacation days) is depicted by higher y-axis positions, are more easily interpretable than line graphs that do not align with these metaphors (e.g., ones where the y-axis is reversed). We also consider how conceptualisations of numerical

quantity and time influence graphical conventions in the design of data visualizations. This paper has an important applied perspective in that we use our findings to make direct recommendations for data visualization practices.

Paper 3 (§5) explores the frequency with which people use different numbers in spoken and written English, and the main factors that influence these frequencies. This quantitative corpus study reveals the numbers people deem important to communicate about, in what contexts, and why, showing that number frequencies can be explained by four key factors: magnitude, roundness, cultural salience, and register. This paper is not directly connected to metaphor and iconicity but delves deeper into the topics of numerical communication and cognition that are the focus of Papers 1 and 2. This study is significant in that it is the most exhaustive, largest-scale analysis of number use in attested language to date.

Paper 4 (§6) puts forward a model of language evolution in which iconicity stands front and centre. In this theoretical paper, we contend that sensorimotor simulations drive the production of iconic forms in interactions. Via conventionalization, these iconic forms rejuvenate iconicity in the lexicon, which counteracts other system-wide pressures (e.g., toward efficiency) for iconicity to decline over time through 'deiconizing' processes of language change (e.g., phonetic reduction). This paper is notable for presenting a model of language evolution in which iconicity has primacy, and for relating processes of language evolution to cognition (i.e., sensorimotor simulation).

3. PAPER 1

Woodin, G., Winter, B., Perlman, M., Littlemore, J. & Matlock, T. (2020). 'Tiny numbers' are actually tiny: Evidence from gestures in the TV News Archive. *PLOS One*, 1–21. https://doi.org/10.1371/journal.pone.0242142

- **Conceptualization**: Bodo Winter, Marcus Perlman, Jeannette Littlemore, Teenie Matlock.
- **Data curation**: Greg Woodin, Bodo Winter, Teenie Matlock.
- **Formal analysis**: Greg Woodin.
- **Investigation**: Greg Woodin, Marcus Perlman, Jeannette Littlemore.
- **Methodology:** Greg Woodin, Bodo Winter, Marcus Perlman, Jeannette Littlemore.
- **Project administration:** Greg Woodin, Bodo Winter, Marcus Perlman.
- **Supervision:** Bodo Winter, Marcus Perlman, Jeannette Littlemore, Teenie Matlock.
- **Validation:** Bodo Winter, Marcus Perlman, Jeannette Littlemore, Teenie Matlock.
- **Visualization:** Greg Woodin.
- **Writing – original draft:** Greg Woodin.
- **Writing – review & editing:** Greg Woodin, Bodo Winter, Marcus Perlman, Jeannette Littlemore, Teenie Matlock.

Supplementary materials: https://osf.io/dncjg/

PLOS ONE

RESEARCH ARTICLE

'Tiny numbers' are actually tiny: Evidence from gestures in the TV News Archive

Greg Woodin[1]*, Bodo Winter[1], Marcus Perlman[1], Jeannette Littlemore[1], Teenie Matlock[2]

1 English Language & Linguistics, University of Birmingham, Birmingham, United Kingdom, **2** Cognitive and Information Sciences, University of California, Merced, CA, United States of America

* gawoodin@gmail.com

Check for updates

OPEN ACCESS

Citation: Woodin G, Winter B, Perlman M, Littlemore J, Matlock T (2020) 'Tiny numbers' are actually tiny: Evidence from gestures in the TV News Archive. PLoS ONE 15(11): e0242142. https://doi.org/10.1371/journal.pone.0242142

Editor: Dennis Tay, The Hong Kong Polytechnic University, HONG KONG

Received: March 16, 2019

Accepted: October 28, 2020

Published: November 17, 2020

Copyright: © 2020 Woodin et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data files, links to videos analysed, codebooks etc. are available under the following OSF repository: https://osf.io/dncjg/.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

We report a large-scale, quantitative investigation of manual gestures that speakers perform when speaking metaphorically about numerical quantities. We used the TV News Archive–an online database of over 2 million English language news broadcasts–to examine 681 videos in which 584 speakers used the phrase 'tiny number', 'small number', 'large number', or 'huge number', which metaphorically frame numerical quantity in terms of physical size. We found that the gestures speakers used reflect a number of different strategies to express the metaphoric size of quantities. When referring to greater versus lesser quantities, speakers were far more likely to gesture (1) with an open versus closed hand configuration, (2) with an outward versus inward movement, and (3) with a wider distance between the gesturing hands. These patterns were often more pronounced for the phrases containing more extreme adjectives ('tiny/huge number'). However, we did not find that speakers performed two-handed versus one-handed gestures. Nor did we find that speakers performed right-handed versus left-handed gestures, when referring to greater versus lesser quantities. Overall, this work supports the claim that metaphoric thought is involved in the production of verbal metaphors that describe numerical magnitudes. It demonstrates that size-based numerical associations observed in previous lab experiments are active in real-life communication outside the lab.

1. Introduction

English speakers often talk about quantities in terms of physical size [1, 2] For instance, numbers of different magnitudes are typically described using size terms such as 'tiny', 'small', 'large', and 'huge', and changing quantities can be characterised as 'shrinking' or 'growing'. Using words or phrases from one domain (e.g., physical size) to describe another (e.g., numerical quantity) reflects what is referred to in cognitive linguistics as *conceptual metaphor* [3–5]. Conceptual metaphor theory views everyday linguistic expressions such as 'tiny number' as surface-level manifestations of mental schemas that people use to conceptualise numerical quantities.

Behavioural experiments support the idea that there is a deep mental connection between the conceptualisation of physical size and numerical quantity, as instantiated by linguistic

metaphors. For instance, people are quicker to correctly judge which of two numbers is greater when the greater number is presented in larger typeface [6]. People are also quicker to correctly judge which of two dot displays contains more dots when the more numerous display covers more area [7]. Studies have linked these size-based spatial-numerical associations to manual actions [8]. For example, people are faster at initiating a precision grip (forefinger and thumb approaching one another, as if holding a small pellet) in response to lesser numbers, and initiating a power grip (firm grip involving the full hand, as if holding a pipe) in response to greater numbers [9]. Moreover, when people reach for blocks with numbers written on them, they spontaneously widen their grip aperture between index finger and thumb if the number is greater, regardless of the actual size of the blocks [10]. These studies show that thinking about greater numbers is mentally connected with actions used for interacting with larger objects, whereas thinking about lesser numbers is mentally connected with actions used for interacting with smaller objects.

Further evidence for the idea that linguistic metaphors such as 'tiny number' reflect a conceptualisation of quantity in terms of size comes from the gestures that speakers perform with their hands when talking about quantities. Winter and colleagues [11] discuss the gestures performed by a speaker on a TV news programme while making the following comment: 'There is a tiny number of people that are contributing a huge amount of money this election'. When saying 'tiny number', the speaker performed a gesture in which she drew her forefinger and thumb close together, as if holding the 'tiny number' between her fingers. When saying 'huge amount', she gestured outward from her body, with flat, open palms facing one another, as if representing the large physical size of the 'huge amount' with the space between her hands. Gestures such as these can be used as a window into the mind [12–14], and their potential to reveal metaphoric thought processes during language use has been shown for a number of abstract conceptual domains [15–18].

In this paper, we use gesture as a means of exploring mental number space. Much of the research on spatial-numerical association has focused on axial representations [2, 19, 20], but here we focus on the gestures that occur with linguistic expressions of size, specifically the metaphoric phrases 'tiny number', 'small number', 'large number', and 'huge number'. With gestures being a flexible way of expressing mental content, there is an array of possible strategies that speakers can use to express relative differences in size [21–24]. The experimental literature showing that precision grips are associated with small quantities [8–10] leads us to predict that speakers will use precision grips more often when talking about relatively smaller quantities. In addition, the fact that larger visually presented areas are associated with relatively larger quantities [6, 7] suggests that, when speakers demarcate a space between their hands, the distance between their hands should be wider when talking about greater quantities. Moreover, we may find that speakers will move their hands away from each other when talking about greater quantities, and toward each other when talking about smaller quantities.

The number of hands the speaker uses to gesture may also be associated with metaphoric size. For instance, while the distance between the thumb and index finger of a single hand can be used to depict smaller quantities, both hands may be needed to designate a wider space for representing greater quantities. In parallel to the use of precision-grip gestures–as if manipulating a small object–to represent lesser quantities, if greater quantities are conceptualised as physically larger, gestures may reflect two-handed manual actions associated with interacting with large physical objects. The use of both hands to represent greater numerical quantities may be comparable to the phenomenon of articulatory plurality in signed languages. Börstell and colleagues [25] showed in three historically unrelated signed languages (American Sign Language, Israeli Sign Language, Swedish Sign Language) that signs for plural concepts are more likely to be two-handed than one-handed. Here, we explore whether a similar effect can

be observed for the manual gestures English speakers use when talking about numerical quantities, specifically, whether speakers are more likely to use two-handed gestures when talking about relatively greater magnitudes.

When speakers gesture with only one hand, the hand they opt to use may also reflect the way they conceptualise quantities of different magnitudes. There is evidence that English speakers think of numbers in terms of a mental number line, with lesser numbers being associated with leftward space and greater numbers with rightward space (e.g., the SNARC effect) [19, 20]. In line with this evidence, Daar and Praat [26] found that, when participants are given a free choice about whether to respond to numbers with their left or right hand, they are more likely to use their right hand in response to greater numbers. Furthermore, in a similar task involving the manual selection of number blocks, participants preferentially selected lesser numbers using their left hand [27]. Based on these results, speaker may be more likely to gesture with their left hand when talking about lesser quantities, and with their right hand when talking about greater quantities.

In addition to their binary magnitude, the four size-based metaphoric descriptions we investigate differ with respect to their position on the scalar dimension of size. Specifically, 'tiny number' and 'huge number' are more extreme than 'small number' and 'large number'. Thus, we were also able to examine whether the patterns of gesturing (e.g., open hand configurations for greater quantities) are more pronounced with the extreme phrases.

Our investigation of size gestures used to represent numerical quantities is important for several reasons. First, in contrast to experiments on numerical cognition, which typically require participants to respond to stimuli under highly constrained, laboratory-controlled conditions, gesture allows us to test the association between physical size and numerical quantity in a more ecologically valid medium: verbal communication. Gesture is a highly flexible modality of expression, especially in comparison to lab experiments, where speakers are typically forced to respond by pressing fixed response keys [2, 28]. During verbal communication, there is no requirement that speakers gesture at all, and the form and movement of these gestures is limited only by the mechanical constraints of the human body. Thus, the study of co-speech gestures can elucidate whether mental associations between space and number are evident in a less constrained and more naturalistic setting.

A second motivation behind our study is methodological. We demonstrate the value of a procedure for studying a large number of gestures produced in relation to specific verbal expressions. Gesture research is time-consuming and finding naturally-occurring gestures that evidence a specific phenomenon is difficult. Many gesture researchers annotate multimodal discourse for features of interest, scanning primary [29] or secondary [30] data for gestures relevant to their research topic. Even in studies where conversation between participants is elicited on a specific topic [31], unavoidably there are parts of the conversation that are irrelevant to the study's aims. In this study, we were able to examine hundreds of gestures that speakers produced when using specific verbal expressions (e.g., 'tiny number') by using the TV News Archive (https://archive.org/details/tv), a huge online database of television news shows, as well as public lectures and governmental programming. The archive, searchable by closed-caption transcripts, allowed us to automatically identify clips from news broadcasts featuring speakers–typically politicians, pundits, newscasters, and authors–using size-based metaphors to refer to quantity. All our data are therefore immediately relevant to our research questions, facilitating a bottom-up approach that is especially suited to the analysis of gestures occurring with particular linguistic expressions.

Another methodological advantage of using the TV News Archive as a data source is that it allows the collection of a large sample of videos and speakers. In combination, the number of videos ($N = 681$) and unique speakers ($N = 584$) in our dataset exceeds most previous gesture

studies [18, 32, 33]. The quantitative focus of our study stands in contrast to observational gesture research, which has tended to focus on detailed qualitative analysis of particular examples [15, 28]. With this paper, we contribute to recent advances in large-scale, quantitative gesture research, particularly a recent study in this journal [34], which used the UCLA Red Hen Lab corpus to study hundreds of gestures related to time expressions [35]. In comparison to the Red Hen corpus, an advantage of the TV News Archive is that all its videos are immediately accessible to researchers, making our coding decisions fully transparent and reproducible [36]. Therefore, all videos used in this study and future studies conducted with the TV News Archive can be viewed and re-analysed by other scientists, without requiring registration or payment.

2. Methodology

2.1. The dataset

We began by downloading a list of URLs for 3200 videos selected at random from the TV News Archive that contained the phrases 'tiny number', 'small number', 'large number', and 'huge number' (800 videos per phrase), including plural phrases (e.g., 'huge numbers') and phrases interrupted by fillers (e.g., 'tiny erm number'). To do this, we used the statistical programming language R, version 3.5.1 [37], in the integrated development environment RStudio, version 1.1.456 [38], with the packages 'tidyverse', version 1.2.1 [47], 'rvest', version 0.3.3 [39], 'XML', version 3.98–1.20 [40], and 'jsonlite' [41]. The end result of this extraction was a spreadsheet with a list of video URLs that could be used to access the corresponding video in the TV News Archive. The scripts used to extract video URLs from the TV News Archive and the spreadsheet containing the URLs are publicly available at https://osf.io/dncjg/.

2.2. Video exclusion

Many videos did not lend themselves to informative gesture analysis. Videos were excluded from the final analyses for the following reasons:

- The video did not play or had no sound.
- The video was a duplicate copy of another video in the dataset.
- The speaker did not use the relevant phrase. These videos included false positives produced by the TV News Archive's search engine and closed captioning system, inflections (e.g., 'smaller number'), videos where the words in the target phrase occurred in separate clauses (e.g., 'number one: voter turnout was small; number two: the election was rigged') or not as part of its own noun phrase (e.g., 'how large numbers will be'), and videos where the target phrase was interrupted by a non-filler lexical item (e.g., 'large negative numbers').
- The target phrase was preceded by another size adjective (e.g., 'small tiny numbers').
- The target phrase was negated (e.g., 'it is not a tiny number').
- There was not a sufficiently clear view of the speaker's hands to determine whether or not they gestured, or, if the gesture was visible, what hand configuration they used.
- The audio and video were desynchronized to the extent that it was difficult to determine whether the speaker's gesture co-occurred with their use of the target phrase.
- The speaker spoke a language other than English that was translated into English via voiceover.

- It was not physically practical for the speaker to gesture or change their hand configuration (e.g., they were holding a large object such as a clipboard or microphone) or their hands were engaged in some other task (e.g., shuffling papers).

After these exclusions were made, a total of 681 videos including 584 unique speakers underwent statistical analyses. Only 52 speakers appeared in more than one video, which amounted to 97 videos in total (14.2% of the whole dataset). 36 repeat speakers appeared twice, and 8 appeared three times. Table 1 shows the number of appearances for speakers that appeared in our dataset more than three times. The contribution of multiple data points by repeat speakers was factored into the design of our statistical models (see §2.5).

In the final dataset, 'tiny number' contributed 167 videos and 146 unique speakers, 'small number' contributed 170 videos and 157 unique speakers, 'large number' contributed 151 videos and 144 unique speakers, and 'huge number' contributed 193 videos and 179 unique speakers. Speakers referred to a wide range of actual quantities using these expressions, including 'millions', 'two tenths of one percent', 'one hundred', 'forty percent', and so on.

2.3. Manual gesture annotation

The first author manually coded the data, first by indicating whether or not the speaker gestured (subsequently referred to as Gesture Co-occurrence). To be counted as a gesture, we used the criterion that the speaker must produce a seemingly communicative movement that occurred at least partly in time with their use of the target phrase. While conservative, the criterion of temporal co-occurrence avoided the problem of subjectivity that arises when a gesture occurs in close proximity to a verbal expression that seems semantically related but whose semantic relation cannot be determined objectively. Instances where the speaker's hands were returning to rest position, or where the target phrase was uttered during the post-stroke hold of a gesture, with no additional movement, were not counted as gestures.

The first author then determined the features of gestures using the following categories: Hand Configuration, Palm Orientation, Closed Handshape, Number of Hands, Hand Choice, Hand Distance, and Horizontal Movement. The first category, Hand Configuration, coded for whether the speaker gestured with an open or closed hand configuration. Open hand configurations involved the fingers being extended, such as when performing a palm-up open-hand gesture [42–44]. Open-hand gestures were further coded for Palm Orientation–whether the speaker's palms were facing predominantly upward, downward, inward (facing one another, toward the midline of the speaker's body), forward (away from the speaker), or backward (toward the speaker). The coding decisions for Palm Orientation were not connected to any specific hypothesis and so were not used in any inferential tests, but we report the overall results for Palm Orientation to provide a comprehensive description of the data. Closed hand

Table 1. Speakers that appeared in dataset more than three times and the number of times they appeared.

Speaker	Appearances
Jim Cramer	11
Donald Trump	8
Barack Obama	7
Bernie Sanders	6
Lawrence Lessig	6
Richard Wolff	6
Peter Lavelle	5
Joan Cashin	4

https://doi.org/10.1371/journal.pone.0242142.t001

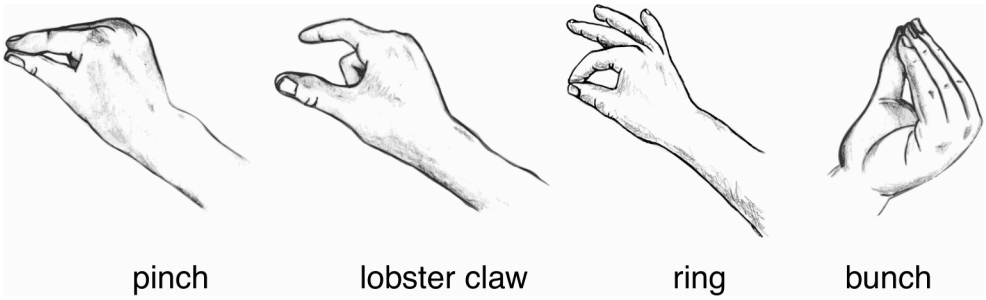


Fig 1. Illustrations of four different closed handshapes. The left three gestures involve precision grips (index finger and thumb approaching or touching each other). In what we call a 'pinch', the index finger and thumb are touching each other, or are held apart with a very narrow gap. In 'lobster claw' gestures, the index finger and thumb are held visibly apart. The 'ring' gesture is similar to the pinch, but with the middle, ring, and pinkie finger extended. For the 'bunch', speakers bring all fingers together, as if scraping together a heap of rice on a table.

https://doi.org/10.1371/journal.pone.0242142.g001

configurations included gestures where the speaker draws the fingers on a single hand close together. These close-hand gestures were also coded for their specific Closed Handshape. As shown in Fig 1, these handshapes included three variants of precision grip-type gestures: 'pinch' gestures (with the index finger and thumb touching or approaching each other), 'lobster claw' gestures (with the index finger and thumb held further apart), and 'ring'-type gestures (with the forefinger and thumb touching and the middle, ring, and pinkie fingers extended, similar to the 'OK' emblem) [42]. Closed handshapes also included 'bunch' gestures (what Kendon calls a 'grappolo' [42], which involves all fingers being held together), clenched fists, and pointing gestures.

While coding the data, it became apparent that a small but significant subset of gestures were more curved inward and tense than open-hand gestures but were not sufficiently closed to constitute closed-hand gestures. We coded these curved gestures as a separate, in-between category. Because they were an intermediate case, we decided to exclude them from the statistical model that tested Hand Configuration, but we still report the number of occurrences of curved gestures across the four phrases in our descriptive analyses.

We also coded for several other properties of the gestures. Number of Hands coded for whether gestures were performed with one hand or two hands. For one-handed gestures, Hand Choice coded for whether the gesture was performed with the left hand or right hand. For two-handed gestures, Hand Distance coded the distance between the speaker's hands using the following categories: 1) narrow: the distance between the speaker's hands was less than the width of their head, 2) medium-width: the distance between the speaker's hands was equal to or more than the width of their head but less than the width of their torso, and 3) wide: the distance between the speaker's hands was equal to or more than the width of their torso. We used this categorical coding distance (using the speaker's body as a frame of reference) due to the fact that the camera position in the TV News Archive is not constant, making it impossible to measure gesture in terms of a more continuous measurement, such as the number of pixels between both hands. Finally, for two-handed gestures, Horizontal Movement coded for whether the speaker gestured with an inward movement (hands moving toward one another), or with an outward movement (hands moving away from one another).

2.4. Inter-rater reliability

A second coder (the third author of this paper) independently coded a subset of the videos to test for inter-rater reliability. First, 156 videos (22.9% of dataset) were used as a training set.

During the training stage, disagreements between Coder 1 and Coder 2 were discussed and the coding scheme was updated (see OSF repository for coding scheme: https://osf.io/dncjg/). Following the training stage, 176 videos (25.8% of dataset) were coded by Coder 2. These 176 videos were coded for Gesture Co-occurrence, Hand Configuration, Palm Orientation, Closed Handshape, Number of Hands, Hand Choice, Hand Distance, and Horizontal Movement. We then calculated the inter-rater reliability (IRR) of coding decisions made by Coder 1 and Coder 2 with Cohen's kappa [45, 46] using the R package 'irr', version 0.84.1 [55]. The IRR results are shown in Table 2. Note that if a category was assigned a certain code for a particular video (e.g., Gesture Co-occurrence coded as 'no'), or if the coders could not agree on a code for this category (e.g., Coder 1 believed there was a gesture but Coder 2 did not), other categories were not coded (e.g., Hand Configuration). Because of this, the total number of videos coded for all columns to the right of Gesture Co-occurrence was fewer than 176.

We performed a weighted test for Hand Configuration to account for the fact that a disagreement over whether a hand is open or closed is a greater difference than a disagreement over whether a hand is curved or closed, or whether a hand is curved or open (curved gestures were in-between closed and open hand configurations). Similarly, we performed a weighted test for Hand Distance to account for the fact that a disagreement over whether a gesture is narrow or wide is a greater difference than a disagreement over whether a gesture is medium-width or narrow, or whether a gesture is medium-width or wide. We performed unweighted tests for the other categories because these coding decisions were not ordered.

For all columns except Closed Handshape, the IRR between Coder 1 and 2 was at least 'substantial', with 'almost perfect' agreement for Gesture Co-occurrence, Hand Configuration, Number of Hands, and Hand Choice. Given the lower coding reliability for Closed Handshape, we still report the descriptive statistics, but we do not perform any inferential tests.

Using a sample of videos from the current study and another investigation in progress comprising 69 videos in total, we addressed the concern that being able to hear the phrase used by the speaker may have affected gesture annotations. For example, if the coder knows that the phrase is 'huge number', this may bias them to analyse a hand configuration as open, rather than closed. To address this concern, we performed a round of blind coding. Blind coding was performed by a third coder (Samantha Ford) who viewed muted videos that were cut to the target phrase only, with the faces of speakers obscured to prevent lip reading. This round of coding focused on Gesture Co-occurrence and Hand Configuration, using a simplified version of the coding scheme that categorised gestures as either closed-hand or open-hand but not curved, as curved gestures were not included in our statistical models. For Gesture Co-occurrence, the IRR agreement was 92.8% (Cohen's $\kappa = 0.666$), indicating substantial agreement. For Hand Configuration, the IRR agreement was 94.6% (Cohen's $\kappa = 0.838$), indicating almost perfect agreement. Our main results are reported using the codes supplied by Coder 1.

Table 2. Inter-rater reliability between Coder 1 and Coder 2 of coding decisions relating to our hypotheses.

	Gesture Co-occurrence	Hand Configuration	Palm Orientation	Closed Handshape	Number of Hands	Hand Choice	Hand Distance	Horizontal Movement
No. of videos	176	127	88	31	129	68	53	54
Agreement	94.8%	96.1%	76.1%	67.7%	95.3%	95.6%	77.4%	87%
Cohen's kappa	0.855	0.94	0.649	0.547	0.906	0.912	0.737	0.799
Level of agreement	Almost perfect	Almost perfect	Substantial	Moderate	Almost perfect	Almost perfect	Substantial	Substantial

https://doi.org/10.1371/journal.pone.0242142.t002

2.5. Statistical analyses

All data were analysed with statistical programming software R, version 3.5.1 [37], in the integrated development environment RStudio, version 1.1.456 [38]. The package 'tidyverse', version 1.2.1 [47] was used for data processing and visualisation; 'ggmcmc' [48] and 'scales', version 1.0.0 [49] were used for data visualisation; 'brms' [50, 51] was used for Bayesian multilevel models; and 'irr', version 0.84.1 [52] was used to calculate inter-rater reliability. All data and analysis code are publicly available at https://osf.io/dncjg/. Statistical analyses were conducted to test the predictions summarised in Table 3.

We used Bayesian logistic regression to test the hypotheses with categorical and binary dependent variables: Hand Configuration (closed versus open; curved gestures were not included in this model), Horizontal Movement (inward versus outward), Number of Hands (one-handed versus two-handed), and Hand Choice (left hand versus right hand). We used Bayesian ordinal regression to test the Hand Distance hypothesis, because this dependent variable had three ordered levels (narrow versus medium versus wide).

In all models, the sole predictor was Phrase ('tiny number', 'small number', 'large number', 'huge number'). This predictor was Helmert coded. Helmert coding compares each level of a categorical predictor variable to the mean of subsequent levels of that variable, allowing us to make the following comparisons: (1) 'small number' versus 'tiny number', (2) 'large number' versus 'small number' and 'tiny number', and (3) 'huge number' versus 'large number', 'small number', and 'tiny number'. This coding scheme allowed us to observe whether certain types of gestures (e.g., open-hand) became more frequent as the phrases referred to increasingly large quantities.

Random intercepts were included in all models to account for by-speaker variation in gestures. We did not include random slopes because most speakers did not appear in our dataset more than once, so they did not have repeated measures for the main predictors of interest (i.e., the predictors were almost exclusively between speakers, rather than within speakers).

We used default priors from the package 'brms' [50, 51] for intercept and standard deviation. We set weakly informative priors on fixed slopes (normal distribution centred at 0 with a standard deviation of 1) and random slopes (half-normal distribution centred at 0 with a standard deviation of 1). As our inference criterion, we observed whether the 95% credible intervals of the posterior distributions for each predictor included zero. Credible intervals that did not contain zero were interpreted as providing strong evidence for the effect of each predictor on the dependent variable. We also report the probability of each effect being above zero.

3. Results

Fig 2 shows an example of a speaker performing a gesture in a video from our dataset. Fig 3 shows the proportions of videos for each phrase in which speakers did or did not gesture. The

Table 3. Summary of predictions for gestures we expected to observe alongside expressions referring to lesser and greater quantities; size of quantities increases from 'tiny number' to 'small number' to 'large number' to 'huge number'.

	Predictions	
	Smaller quantities	**Greater quantities**
Hand Configuration	closed	open
Hand Distance	smaller distance between hands	larger distance between hands
Horizontal Movement	inward movement	outward movement
Number of Hands	one	two
Hand Choice	left	right

https://doi.org/10.1371/journal.pone.0242142.t003

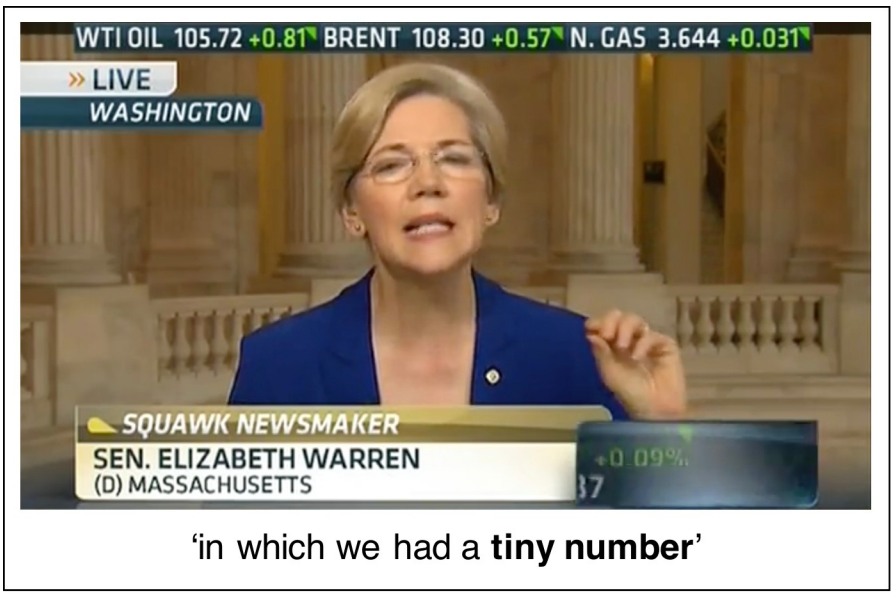


Fig 2. U.S. Senator Elizabeth Warren performs a pinch gesture with a closed hand configuration while using the target phrase 'tiny number', republished from CNBC (NBC Universal) under a CC BY license, with permission from CNBC (NBC Universal), original copyright 2013.

https://doi.org/10.1371/journal.pone.0242142.g002

proportions displayed in the figure are based on all videos, excluding two videos in which the speaker's hands were in different configurations (left hand open, right hand closed).

Overall, more videos contained gestures (78.4%, $N = 534$) than did not contain gestures (21.6%, $N = 147$), including the two videos excluded from the figure. The expression 'tiny number' had the highest rate of gesture co-occurrence (90.4%, $N = 151$), with comparatively few videos not containing gestures (9.6%, $N = 16$). For the other three phrases, gesture co-occurrence was lower but still relatively high: of these three phrases, 'huge number' had the

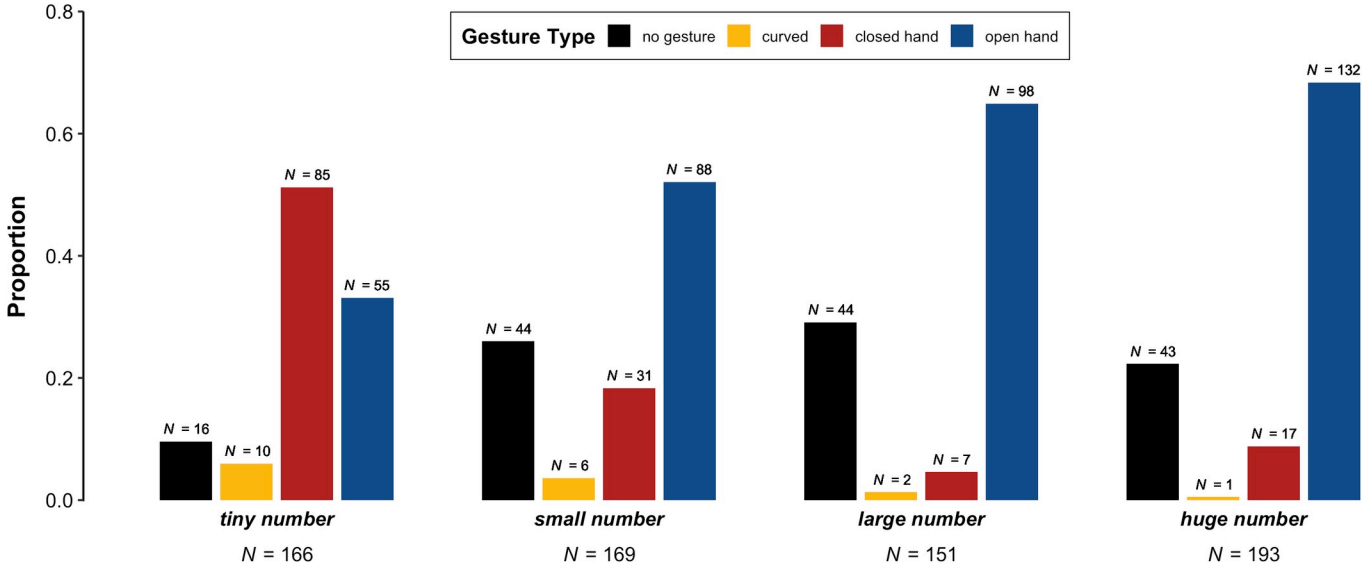


Fig 3. Proportion of videos with the phrases 'tiny number', 'small number', 'large number', and 'huge number' that occurred without a gesture or with a gesture in a curved, closed, or open hand configuration.

https://doi.org/10.1371/journal.pone.0242142.g003

highest rate of gesture co-occurrence (gesture: 77.7%, $N = 150$; no gesture: 22.3%, $N = 43$), followed by 'small number' (gesture: 74.1%, $N = 126$; no gesture: 25.9%, $N = 44$), and then 'large number' (gesture: 70.9%, $N = 107$; no gesture: 29.1%, $N = 44$).

3.1. Hand configuration

We now look at whether speakers performed gestures with closed or open hand configurations. In the following analysis, we exclude two gestures where the speaker's hands were in different configurations (left hand open, right hand closed). The proportions reported in the following paragraph are based solely on closed-hand and open-hand gestures.

As shown in Fig 3, 'tiny number' is the only phrase that was accompanied by closed-hand gestures in a majority of videos (60.7%, $N = 85$), with fewer open-hand gestures (39.3%, $N = 55$). In contrast, 'large number' was mostly accompanied by open-hand gestures (93.3%, $N = 98$), with comparatively few closed-hand gestures (6.7%, $N = 7$). For 'huge number', open-hand gestures were similarly dominant (88.6%, $N = 132$), with comparatively few closed-hand gestures (11.4%, $N = 17$). For 'small number', open-hand gestures still predominated (73.9%, $N = 88$), but closed-hand gestures were relatively more frequent (26.1%, $N = 31$) compared to the phrases referring to large quantities. In fact, there were more closed-hand gestures for 'small number' than for 'large number' and 'huge number' put together.

The posterior distributions of the Bayesian logistic regression model (see §2.3 for model specification) provide strong evidence for an effect of Phrase ('tiny number', 'small number', 'large number', 'huge number') on Hand Configuration (closed versus open). For all comparisons, credible intervals did not include zero. In addition, the percentage of posterior samples that were above zero for each comparison was exactly or near 100%. First, open-hand gestures were more frequent alongside 'small number' than 'tiny number' (Helmert-coded odds = 2.29 to 1, log odds = 0.83, Bayesian 95% credible interval = [0.51, 1.23], posterior samples $> 0 = 100\%$). Second, open-hand gestures were more frequent alongside 'large number' than 'tiny number' and 'small number' (odds = 2.34 to 1, log odds = 0.85, 95% *CI* = [0.55, 1.20], post $> 0 = 100\%$). Third, open-hand gestures were more frequent alongside 'huge number' than 'tiny number', 'small number', and 'large number' (odds = 1.3 to 1, log odds = 0.26, 95% *CI* = [0.11, 0.44], post $> 0 = 99.9\%$). Thus, the Helmert coded predictor shows that there were progressively more open hand configurations for progressively greater quantities.

Other gestures not included in the above analysis were coded as 'curved', being midway between closed and open hand configurations. The following proportions are based solely on curved gestures. Most curved gestures appeared alongside the phrase 'tiny number' (52.6%, $N = 10$), followed by 'small number' (31.6%, $N = 6$), 'large number' (10.5%, $N = 2$), and then 'huge number' (5.3%, $N = 1$). Thus, although these gestures were not very frequent overall, they seemed to be more commonly associated with lesser quantities.

For closed-hand gestures, we annotated the specific type of Closed Handshape. The following proportions are based solely on closed-hand gestures. For the expressions referring to smaller quantities, pinch gestures were the most prevalent gesture type ('tiny number': 45.8%, $N = 38$; 'small number': 48.4%, $N = 15$), followed by lobster claw gestures ('tiny number': 24.1%, $N = 20$; 'small number': 29%, $N = 9$), and then bunch gestures ('tiny number': 14.5%, $N = 12$; 'small number': 19.4%, $N = 6$). Less prevalent or absent were ring-type gestures ('tiny number': 7.2%, $N = 5$; 'small number': 0%, $N = 0$), pointing gestures ('tiny number': 4.8%, $N = 4$; 'small number': 0%, $N = 0$), and clenched fist gestures ('tiny number': 3.6%, $N = 3$; 'small number': 3.2%, $N = 1$). The counts for 'large number' are low and so do not present any interpretable pattern (7 gestures across 5 categories). However, the counts for 'huge number' reveal that pointing gestures predominated (41.2%, $N = 7$), followed by clenched fist gestures

(29.4%, $N = 5$). Gestures in the other categories were at similarly low levels (ring-type: $N = 2$; bunch: $N = 1$; lobster claw: $N = 1$; pinch: $N = 1$).

For open-hand gestures, we also coded for Palm Orientation. The following proportions are based solely on open-hand gestures, minus 9 videos we excluded for containing gestures where the speaker's palms were oriented differently from one another (e.g., left palm downward, right palm inward). For three of the four phrases, inward facing palms (facing each other, toward midline of speaker's body) were the most common orientation: 'tiny number' (59.3%, $N = 32$), 'small number' (55.8%, $N = 48$), and 'huge number' (51.9%, $N = 67$). For 'large number', inward facing palms were less common (40%, $N = 38$). The general dominance of inward facing palms makes sense given that this gesture type is a prototypical size gesture, where the distance between the palms represents the size of the referent [53]. Compared to inward facing palms, upward facing palms were less common across the four phrases, with 'large number' (20%, $N = 19$) having a slightly higher proportion of this palm orientation than 'huge number' (17.8%, $N = 23$), which had a higher proportion of upward facing palms than both 'tiny number' (14.8%, $N = 8$) and 'small number' (14%, $N = 12$). For the phrases that referred to greater quantities, backward facing palms were more common ('large number': 13.7%, $N = 13$; 'huge number': 10.1%, $N = 13$) than for the phrases that referred to lesser quantities ('tiny number': 5.6%, $N = 3$; 'small number': 4.7%, $N = 4$). Similarly, for the greater quantity phrases, downward facing palms were more common ('large number': 15.8%, $N = 13$; 'huge number': 15.5%, $N = 20$) than for the lesser quantity phrases ('tiny number': 11.1%, $N = 6$; 'small number': 11.6%, $N = 10$). Lastly, 'small number' had a higher proportion of frontward facing palms (14%, $N = 12$) than 'large number' (10.5%, $N = 10$) and 'tiny number' (9.3%, $N = 5$) with the lowest proportion of frontward facing palms accompanying 'huge number' (4.7%, $N = 6$).

3.2. Number of hands

We now look at whether speakers were more likely to gesture with one hand or two hands when using the target expressions. The following proportions are based on all videos in which both speakers' hands were visible and free to gesture. Fig 4 shows the results from this analysis. Descriptively, 'tiny number' had the highest proportion of one-handed gestures (71.4%, 45 gestures), with fewer two-handed gestures (28.6%, 18 gestures). The proportions of one-handed and two-handed gestures were similar for 'small number' (one hand: 17.3%, 9 gestures; two hands: 80.4%, 43 gestures), 'large number' (one hand: 19.6%, 10 gestures; two hands: 80.4%, 41), and 'huge number' (one hand: 19.2%, 14 gestures; two hands: 80.8%, 59 gestures).

The posterior distributions of the Bayesian logistic regression model do not provide evidence for an effect of Phrase ('tiny number', 'small number', 'large number', 'huge number') on Number of Hands (one versus two). First, the model did not reveal a reliable difference between 'small number' and 'tiny number', with a credible interval that included zero (odds = 1.48 to 1, log odds = 0.39, 95% $CI = [-0.11, 0.97]$, post $> 0 = 93.7\%$). Similarly, the model did not indicate that there was a reliable difference between 'large number' and the phrases 'tiny number' and 'small number' (odds = 1.07 to 1, log odds = 0.07, 95% $CI = [-0.24, 0.39]$, post $> 0 = 65.7\%$). Finally, the model did not indicate that there was a reliable difference between 'huge number' compared to 'tiny number', 'small number', and 'large number' (odds = 1.03 to 1, log odds = 0.03, 95% $CI = [-0.17, 0.24]$, post $> 0 = 62.3\%$).

3.3. Hand distance

For two-handed gestures, we now look at the distance between the speaker's hands at the end of the gesture stroke. The following proportions are based on all two-handed gestures, minus 6

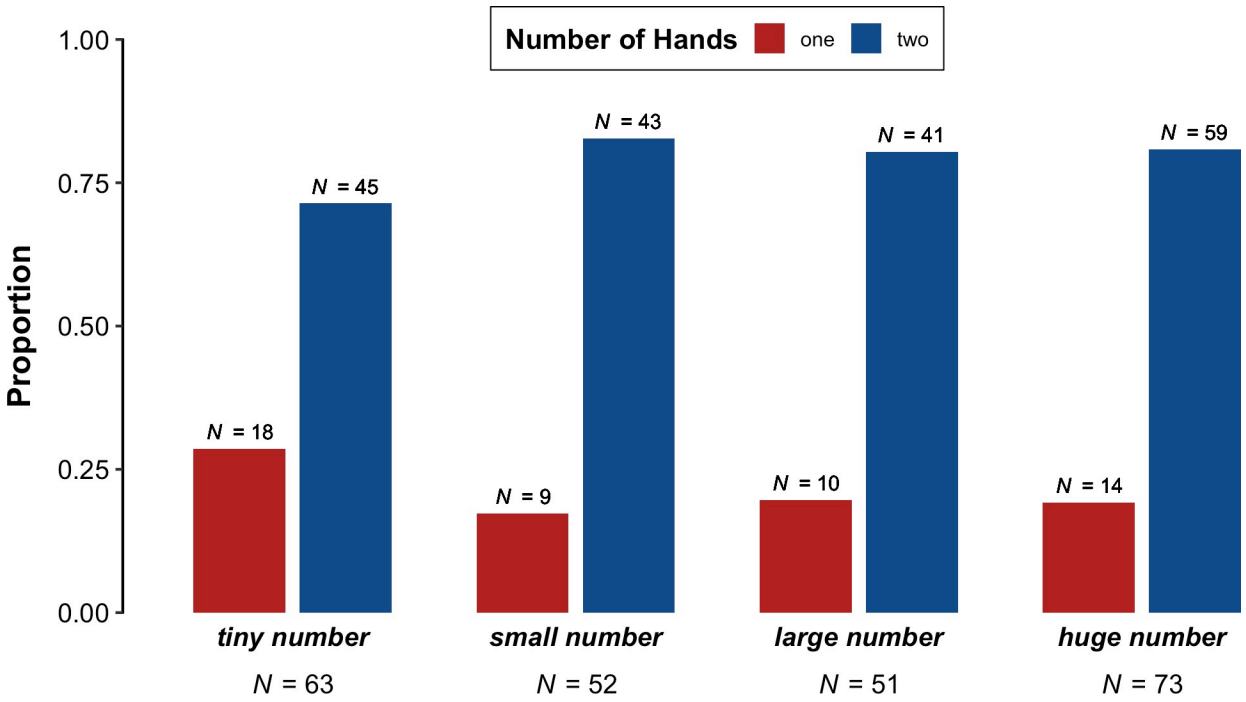


Fig 4. Proportion of videos in which both speakers' hands were visible and free to gesture per phrase ('tiny number', small number', 'large number', 'huge number') that contained one-handed versus two-handed gestures.

https://doi.org/10.1371/journal.pone.0242142.g004

gestures we excluded because it was not possible to determine the distance between the speaker's hands (e.g., the speaker's hands disappeared offscreen before the end of the gesture stroke). As shown in Fig 5, the phrases that referred to lesser quantities had a higher proportion of narrow gestures ('tiny number': 70.8%, 34 gestures; 'small number': 68.2%, 30 gestures) than 'large number' (41.5%, 17 gestures) and especially 'huge number' (30.8%, 20 gestures). In comparison, the lesser quantity expressions had a lower proportion of medium-width gestures ('small number': 22.7%, 10 gestures; 'tiny number': 18.8%, 9 gestures) compared to 'huge number' (29.2%, 19 gestures) and especially 'large number' (36.6%, 15 gestures). Similarly, the lesser quantity expressions had a lower proportion of wide gestures ('tiny number': 10.4%, 5 gestures; 'small number': 9.1%, 4 gestures) compared to 'large number' (22%, 9 gestures) and especially 'huge number', for which wide gestures was the most common hand distance (40%, 26 gestures).

The posterior distributions for the Bayesian ordinal regression model (outlined in §2.3) provide some evidence for an effect of Phrase ('tiny number', 'small number', 'large number', 'huge number') on Hand Distance (narrow versus medium versus wide). As a caveat, the model indicated that there was not a reliable difference between 'small number' and 'tiny number' (odds = 1.07 to 1, log odds = 0.07, 95% *CI* = [-0.51, 0.67], post > 0 = 59.6%). However, there was strong evidence that wider gestures were more frequent alongside 'large number' than alongside 'tiny number' and 'small number' (odds = 1.59 to 1, log odds = 0.5, 95% *CI* = [0.15, 0.85], post > 0 = 99.7%). There was also strong evidence that wider gestures were more frequent alongside 'huge number' than alongside 'tiny number', 'small number', and 'large number' (odds = 1.65 to 1, log odds = 0.5, 95% *CI* = [0.27, 0.82], post > 0 = 100%). In sum, there was a trend for speakers to gesture with a larger distance between their hands when using the phrases referring to greater quantities ('large number', 'huge number') than when using the phrases referring to lesser quantities ('tiny number', 'small number'). Speakers

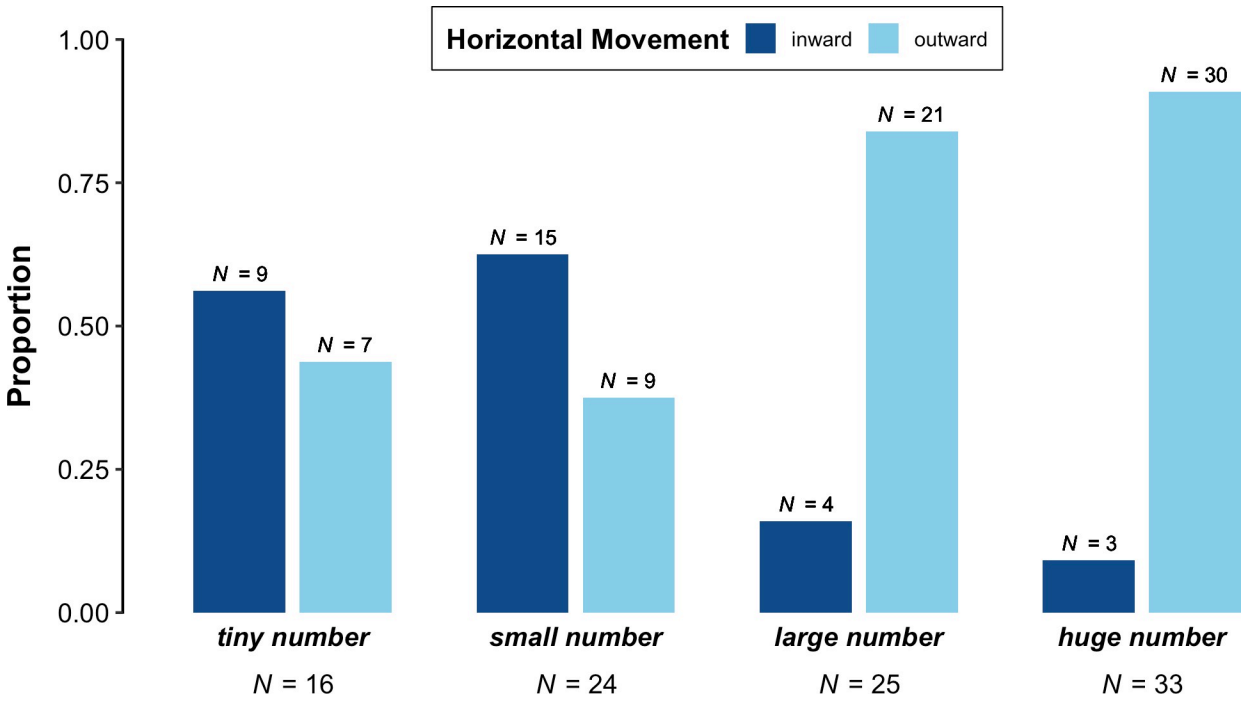


Fig 5. Proportion of two-handed gestures per phrase ('tiny number', small number', 'large number', 'huge number') that were inward moving versus outward moving.

https://doi.org/10.1371/journal.pone.0242142.g005

tended to perform gestures with a particularly large distance between the hands for 'huge number'.

3.4. Horizontal movement

We now examine whether speakers performed two-handed gestures with an inward movement (hands moving toward each other) or an outward movement (hands moving away from each other) when using the target expressions. The proportions reported in this section are based solely on two-handed gestures, minus 92 two-handed gestures in which the speaker's hands did not move in a horizontal direction, and 14 two-handed gestures where both hands moved horizontally in the same direction (see OSF repository for distribution of the excluded gestures across the four phrases: https://osf.io/dncjg/).

As shown in Fig 6, 'tiny number' (inward: 56.2%, 9 gestures; outward: 43.8%, 7 gestures) and 'small number' (inward: 62.5%, 15 gestures; outward: 37.5%, 9 gestures) were accompanied mostly by inward-moving gestures. In contrast, 'large number' (inward: 16%, 4 gestures; outward: 84%, 21 gestures) and 'huge number' (inward: 9.1%, 3 gestures; outward: 90.9%, 30 gestures) were accompanied mostly by outward-moving gestures.

Overall, the posterior distributions of the Bayesian logistic regression model provide strong evidence for an effect of Phrase ('tiny number', 'small number', 'large number', 'huge number') on Horizontal Movement (inward versus outward). As with Hand Distance, 'tiny number' and 'small number' did not appear to be reliably different (odds = 0.83 to 1, log odds = -0.18, 95% *CI* = [-1.02, 0.61], post > 0 = 32.5%). However, the model indicated that outward-moving gestures were more frequent alongside 'large number' than 'tiny number' and 'small number' (odds = 2.36 to 1, log odds = 0.86, 95% *CI* = [0.34, 1.53], post > 0 = 100%). Furthermore, the model indicated that outward-moving gestures were more frequent alongside 'huge number'

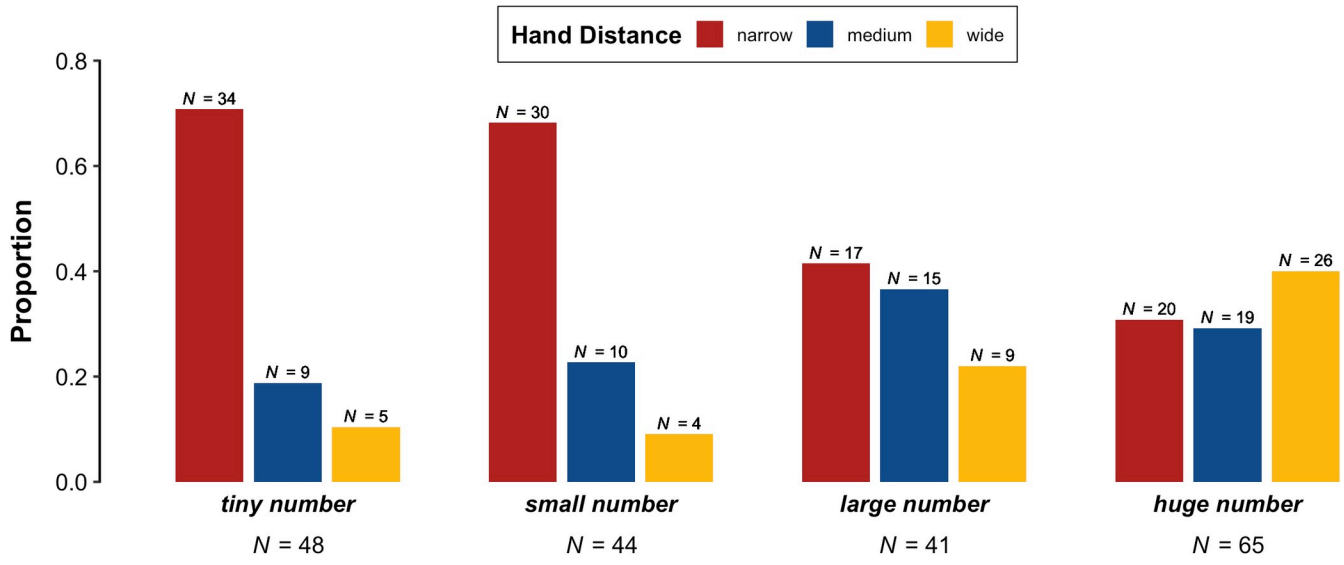


Fig 6. Proportion of two-handed gestures per phrase ('tiny number', small number', 'large number', 'huge number') in which the speaker's hands were separated by a narrow, medium, or wide distance.

https://doi.org/10.1371/journal.pone.0242142.g006

than 'tiny number', 'small number', and 'large number' (odds = 1.9 to 1, log odds = 0.64, 95% *CI* = [0.25, 1.16], post > 0 = 100%). The results of this model thus confirm the pattern depicted in Fig 6: the phrases referring to lesser quantities ('tiny number' and 'small number') were accompanied mostly by inward-moving gestures, whereas the phrases referring to greater quantities ('large number' and 'huge number') were accompanied mostly by outward-moving gestures. In addition, 'huge number' was accompanied by a higher proportion of outward-moving gestures than 'large number'.

3.5. Hand choice

Finally, we investigate whether speakers gestured with their left hand or right hand when using the target expressions. The following proportions are based on all one-handed gestures in which the speaker's hands were both visible and free to gesture. Because there were relatively few of this type of gesture, the counts reported here are low, but there was a clear trend for right-handed gestures to predominate, overall. As shown in Fig 7, descriptively, 'large number' had the highest proportion of right-handed gestures (80%, *N* = 8) and the lowest proportion of left-handed gestures (20%, *N* = 2). Similarly, 'tiny number' had a high proportion of right-handed gestures (77.8%, *N* = 14) and a low proportion of left-handed gestures (22.2%, *N* = 4). For 'huge number', right-handed gestures were slightly less frequent but still predominated (71.4%, *N* = 10), while left-handed gestures were slightly more frequent but still low in frequency (28.6%, *N* = 4). Finally, 'small number' had the lowest proportion of right-handed gestures (55.6%, *N* = 5) and the highest proportion of left-handed gestures (44.4%, *N* = 4).

The posterior distributions of the Bayesian logistic regression model do not provide strong evidence for an effect of Phrase ('tiny number', 'small number', 'large number', 'huge number') on Hand Choice (left versus right). The credible intervals for each comparison included zero: between 'small number' and 'tiny number' (odds = 0.61 to 1, log odds = -0.49, 95% *CI* = [-1.49, 0.48], post > 0 = 15.7%), between 'large number' and the phrases 'tiny number' and 'small number' (odds = 0.98 to 1, log odds = 0.26, 95% *CI* = [-0.43, 1.01], post > 0 = 76.1%), and between 'huge number' and the phrases 'tiny number', 'small number', and 'large number'

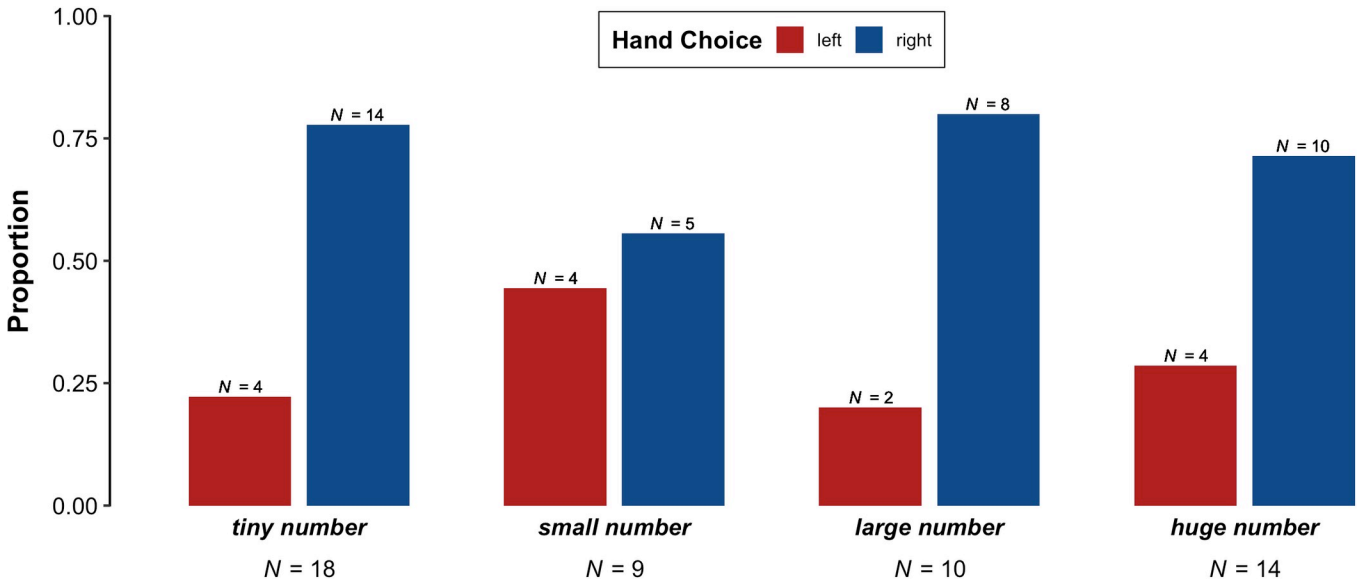


Fig 7. Proportion of videos in which both speakers' hands were visible and free to gesture per phrase ('tiny number', 'small number', 'large number', 'huge number') that contained left-handed versus right-handed gestures.

https://doi.org/10.1371/journal.pone.0242142.g007

(odds = 1.03 to 1, log odds = -0.01, 95% *CI* = [-0.46, 0.47], post > 0 = 47.1%). Therefore, we do not find any support for the hypothesis that right-handed gestures more frequently accompany phrases referring to greater quantities.

4. Discussion

We used the TV News Archive to conduct a large-scale, quantitative investigation of the gestures that speakers produce when talking about numerical quantity. Specifically, we examined whether their gestures reflect the size-based verbal metaphors they use when referring to quantities of different magnitudes. Our search included four metaphoric expressions: 'tiny number', 'small number', 'large number', and 'huge number'. This search returned 681 videos that met our inclusion criteria (speakers visible, hands free, etc.), with 534 gestures performed by 461 speakers coded in total.

Gesture co-occurrence rates ranged from 77.7% for 'large number' to 90.4% for 'tiny number', demonstrating that speakers were much more likely to gesture than not across the four phrases. Of the gestures that occurred, our results showed that speakers used a wide array of strategies for signalling relative differences in size via gesture: changing their hand configuration, changing the space spanned between their hands, and moving their hands inward or outward. These results indicate that the associations between precision grips and small quantities [8–10], and between larger visually presented areas and greater quantities [6, 7] observed in laboratory experiments, extend to communicative actions produced spontaneously during verbal communication.

The predominance of open-hand gestures for three of the four phrases in the dataset may indicate that open hand configurations are the gestural norm or default, perhaps because the hands may tend to be in an approximately open hand configuration while at rest, making it easier to perform an open-hand gesture than a closed-hand one. To override this default, a sufficiently strong association between a numerical quantity and closed-hand actions may be required [9, 10, 54], which appears to be triggered most reliably when speakers use the phrase 'tiny number', even more so than the less extreme expression 'small number'.

Of the precision grip-type gestures that occurred alongside the expressions referring to lesser quantities ('tiny number', 'small number'), pinches were most frequent, followed by lobster claw gestures. In contrast, bunch gestures occurred comparatively less frequently. Interestingly, the ring gesture (where the middle, ring, and pinkie finger are extended) was much less frequent than the pinch and the lobster claw, both of which involve curling in the middle, ring, and pinkie finger. These results are in line with Hassemer and Winter's [21, 22] analysis of precision grips, according to which curling in all fingers other than the profiled index finger and thumb is crucial for the expression of size information, particularly for small referents. Another potential explanation is that the ring gesture is similar to the 'OK' emblem [42], and so it may be avoided because of its association with this highly conventionalised meaning, which is unrelated to size.

When speakers referred to greater quantities ('large number', 'huge number'), they also tended to gesture with a wider distance between their hands, and were more likely to gesture with an outward movement, than when they referred to lesser quantities ('tiny number', 'small number'). While we treated the distance between speakers' hands and their movement direction separately in our analysis, they are often two facets of the same gesture: speakers tended to move their hands apart to a wide distance when talking about greater quantities, and move their hands to a narrow distance when talking about lesser quantities. Specifically, 77.4% ($N = 24$) of inward-moving gestures were narrow, whereas 12.9% ($N = 4$) were medium-width, and 9.7% ($N = 3$) were wide. In contrast, just 21% ($N = 13$) of outward-moving gestures were narrow, 32.3% ($N = 20$) were medium-width, and 46.8% ($N = 29$) were wide.

We did not find any overall two-handed bias for greater quantities, as we expected based on the observation that two-handed actions are more compatible with larger physical quantities, as well as the sign language phenomenon of articulatory plurality [25]. We also did not find any evidence that speakers were more likely to produce right-handed gestures for greater quantities, which we predicted based on the experimental finding that quantities are mentally represented with lesser quantities on the left to and greater quantities on the right [19, 20], in addition to previous research demonstrating patterns in the hand chosen to respond to different numerical magnitudes [26, 27]. Our results suggest that these experimental findings may not apply to gestures produced spontaneously when speakers use size-based expressions. Instead, consistent with the verbal metaphors that speakers used, speakers appeared to be conceptualising the numbers primarily in terms of physical size, rather than the horizontal axis.

Our results support the core claim of Conceptual Metaphor Theory that verbal metaphors reflect mental schemas that people use to conceptualise different aspects of the world [3–5]. At least with respect to the numerical quantity metaphors in our dataset, the patterns we observed in speakers' gestures suggested that these metaphors are not mere figures of speech, but rather represent a deeper, size-based conceptualisation of numerical quantities. The foregrounding of verbal metaphors through gesture can thus be seen as an additional source of evidence that metaphors are conceptually 'active' during the production of an utterance [18–21]. This is especially important given that expressions such as 'large number' are highly conventionalised. Conventionalisation is framed as a serial killer of metaphoricity–as conventionalisation increases, the likelihood that speakers will recognise an expression's metaphoric properties is thought to decrease [55–58]. Despite conventionalisation, the gestures observed in our dataset suggest that the particular metaphors we investigated were conceptually active for many speakers.

Because speakers used explicitly size-based language, we cannot rule out the possibility that language itself is the primary driver of the gestures we observed, without there being any deeper conceptual mapping. According to Bouissac [59], a gesture co-occurring with a metaphoric expression may redundantly imitate speech, which would make the gesture *iconically*

depictive of the concept expressed in speech, but not metaphoric. However, to the extent that size-based words such as 'tiny' and 'huge' are used to describe an abstract domain (i.e., numerical quantity) in expressions such as 'tiny number' and 'huge number', the whole phrase can be seen as metaphoric. In that case, the iconic gestures corresponding to these words would still occur within an overall abstract context and are metaphoric at least to this degree. Furthermore, while lexical priming of gestures is a possibility for the present dataset, other research has provided evidence for the existence of conceptual metaphors in language-independent tasks that do not use verbal metaphors as prompts [60–62]. Finally, gesture research has shown that speakers sometimes produce metaphoric gestures while not expressing these metaphors verbally [28, 63, 64]. These three points notwithstanding, future research with the TV News Archive can address this concern empirically by investigating the gestures that correlate with expressions that do not relate directly to size, such as 'more', 'less', or number words.

Taken together, our results show how cognitive associations previously observed in experiments extend to language use outside the laboratory–in a database of television news shows, public lectures, and governmental programming, featuring politicians, pundits, newscasters, and authors. This methodological contribution is important for several reasons. First, as experimental methods in research on spatial associations inherently constrain participants' behaviour and thus are subject to task demands. In contrast, gesture is a much freer form of expression. A further limitation of lab-based studies on spatial-numerical associations [20, 65] is that a small set of stimuli is typically explored–often for instance, the numbers 1 to 9. For the verbal metaphors investigated here, speakers referred to a wide range of numerical magnitudes; for instance, 'millions', 'two tenths of one percent', 'one hundred', 'forty percent', and so on. That we found congruent gestures across most of these different quantities shows that the large experimental body of spatial-numerical associations extends to vast numerical ranges spanning many orders of magnitude.

The TV News Archive also offers the advantage that it enables the collection of far more data, especially a higher number of speakers, than would usually be feasible in an experimental setting. Moreover, if time-consuming stages of the analysis procedure could be automated, such as the initial exclusion of unusable videos, still larger databases could be constructed (see [66]), and large portions of the analysis itself could be automated using the Python-operated OpenPose software [67, 68] (see [69, 70]). Finally, the data in the Archive are openly accessible, allowing completely reproducible analyses to be conducted. Therefore, we believe the TV News Archive to be a promising tool for gesture research, for use alongside the popular UCLA Red Hen Lab corpus [35].

However, while our data are arguably more ecologically valid than results obtained via experimentation, it is not clear to what extent gestures produced in televised contexts reflect natural communication. For instance, some politicians and newsreaders that appear in our dataset are likely to have received body language training [71], and may be engaged in non-standard communicative practices such as reading from a script or speaking to a camera, which may affect their gestures. We thus argue for the triangulation of results from different research paradigms when making broad claims about cognition based on gestures.

To conclude, in this study, we have provided large-scale, quantitative evidence to show how conceptual associations between size and numerical quantity cited in the research literature are commonly foregrounded through speakers' gestures when they use different metaphors in their speech. In particular, greater quantities are often conceptualised as physically larger than lesser quantities, and speakers can represent this conceptualisation through the different kinds of gestures that they produce, using their hand configuration, the movement of their hands, and the distance between them. The gestural patterns we observed support the idea that verbal metaphors such as 'tiny number' are psychologically real: 'tiny numbers' are actually tiny, at

least in the minds of speakers. These results show how lab-based research on spatial-numerical associations and other conceptual metaphors can benefit from investigating language and gesture, providing converging evidence for the same underlying conceptual mappings. More broadly, our analyses show how speakers' gestures can provide a window into their internal mental processes, making these processes visible to the researcher. Thus, by applying quantitative methods to large samples of different speakers across different contexts, we can use gesture to observe trends in the way the human mind works.

Acknowledgments

We thank Samantha Ford for blind-coding the data. We also thank three reviewers and Jennifer Hinnell for helpful comments and suggestions.

Author Contributions

Conceptualization: Bodo Winter, Marcus Perlman, Jeannette Littlemore, Teenie Matlock.

Data curation: Greg Woodin, Bodo Winter, Teenie Matlock.

Formal analysis: Greg Woodin.

Investigation: Greg Woodin, Marcus Perlman, Jeannette Littlemore.

Methodology: Greg Woodin, Bodo Winter, Marcus Perlman, Jeannette Littlemore.

Project administration: Greg Woodin, Bodo Winter, Marcus Perlman.

Supervision: Bodo Winter, Marcus Perlman, Jeannette Littlemore, Teenie Matlock.

Validation: Bodo Winter, Marcus Perlman, Jeannette Littlemore, Teenie Matlock.

Visualization: Greg Woodin.

Writing – original draft: Greg Woodin.

Writing – review & editing: Greg Woodin, Bodo Winter, Marcus Perlman, Jeannette Littlemore, Teenie Matlock.

References

1. Winter B, Marghetis T, Matlock T. Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number. Cortex. 2015; 64: 209–224. https://doi.org/10.1016/j.cortex.2014.10.015 PMID: 25437376
2. Winter B, Matlock T, Shaki S, Fischer MH. Mental number space in three dimensions. Neurosci Biobehav Rev. 2015; 57: 209–219. https://doi.org/10.1016/j.neubiorev.2015.09.005 PMID: 26365108
3. Lakoff G, Johnson M. Metaphors We Live By. Chicago, Illinois: University of Chicago Press; 1980. https://doi.org/10.1086/292206 PMID: 11661871
4. Gibbs RW Jr. The Poetics of Mind: Figurative Thought, Language, and Understanding. Cambridge: Cambridge University Press; 1994.
5. Kövecses Z. Metaphor: A Practical Introduction. New York City, New York: Oxford University Press; 2002. Available: https://www.questia.com/library/120086185/metaphor-a-practical-introduction
6. Henik A, Tzelgov J. Is three greater than five: The relation between physical and semantic size in comparison tasks. Mem Cogn. 1982; 10: 389–395. https://doi.org/10.3758/bf03202431 PMID: 7132716
7. Hurewitz F, Gelman R, Schnitzer B. Sometimes area counts more than number. PNAS. 2006; 103: 19599–19604. https://doi.org/10.1073/pnas.0609485103 PMID: 17159143
8. Badets A, Andres M, Di Luca S, Pesenti M. Number magnitude potentiates action judgements. Experimental Brain Research. 2007; 180: 525–534. https://doi.org/10.1007/s00221-007-0870-y PMID: 17279382

9. Lindemann O, Abolafia JM, Girardi G, Bekkering H. Getting a grip on numbers: numerical magnitude priming in object grasping. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33: 1400–1409. https://doi.org/10.1037/0096-1523.33.6.1400 PMID: 18085952

10. Andres M, Davare M, Pesenti M, Olivier E, Seron X. Number magnitude and grip aperture interaction. Neuroreport. 2004; 15: 2773–2777. PMID: 15597052

11. Winter B, Perlman M, Matlock T. Using space to talk and gesture about numbers: Evidence from the TV news archive. Gesture. 2013; 13: 377–408. https://doi.org/10.1075/gest.13.3.06win

12. McNeill D. Gesture and Thought. Chicago, Illinois: University of Chicago Press; 2005. https://doi.org/10.7208/chicago/9780226514642.001.0001

13. McNeill D, editor. Language and Gesture. Cambridge: Cambridge University Press; 2000. https://doi.org/10.1017/CBO9780511620850

14. McNeill D. Hand and Mind: What Gestures Reveal about Thought. Chicago, Illinois: University of Chicago Press; 1992.

15. Cienki AJ, Müller C. Metaphor, gesture, and thought. The Cambridge Handbook of Metaphor and Thought. 2008. pp. 483–502. https://doi.org/10.1016/j.annemergmed.2007.03.032 PMID: 17499391

16. Müller C. Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View. Chicago, Illinois: University of Chicago Press; 2008.

17. Müller C, Tag S. The dynamics of metaphor: Foregrounding and activating metaphoricity in conversational interaction. Cognitive Semiotics. 2013; 6: 85–120.

18. Casasanto D, Jasmin K. Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. Plos one. 2010; 5: e11805. https://doi.org/10.1371/journal.pone.0011805 PMID: 20676371

19. Dehaene S, Bossini S, Giraux P. The mental representation of parity and number magnitude. Journal of Experimental Psychology: General. 1993; 122: 371.

20. Wood G, Willmes K, Nuerk H-C, Fischer MH. On the cognitive link between space and number: a meta-analysis of the SNARC effect. Psychology Science. 2008; 489–525.

21. Hassemer J, Winter B. Decoding gestural iconicity. Cogn Sci. 2018; 42: 3034–3049. https://doi.org/10.1111/cogs.12680 PMID: 30471078

22. Hassemer J, Winter B. Producing and perceiving gestures conveying height or shape. Gesture. 2016; 15: 404–424.

23. Müller C. Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich. Berlin: Spitz; 1998.

24. Sowa T. Understanding coverbal iconic gestures in shape descriptions. Berlin: Akademische Verlagsgesellschaft Aka; 2006.

25. Börstell C, Lepic R, Belsitzman G. Articulatory plurality is a property of lexical plurals in sign language. Lingvisticæ Investigationes. 2016; 39: 391–407.

26. Daar M, Pratt J. Digits affect actions: The SNARC effect and response selection. Cortex. 2008; 44: 400–405. https://doi.org/10.1016/j.cortex.2007.12.003 PMID: 18387571

27. Mills KJ, Rousseau BR, Gonzalez CLR. A cross-sectional developmental examination of the SNARC effect in a visually-guided grasping task. Neuropsychologia. 2014; 58: 99–106. https://doi.org/10.1016/j.neuropsychologia.2014.04.001 PMID: 24732383

28. Walker E, Cooperrider K. The continuity of metaphor: Evidence from temporal gestures. Cogn Sci. 2016; 40: 481–495. https://doi.org/10.1111/cogs.12254 PMID: 26059310

29. Chui K. Spatial conceptualization of sequence time in language and gesture. Gesture. 2018; 17: 176–195. https://doi.org/10.1075/gest.00015.chu

30. Cooperrider K. Body-directed gestures: Pointing to the self and beyond. Journal of Pragmatics. 2014; 71: 1–16. https://doi.org/10.1016/j.pragma.2014.07.003

31. Parrill F, Bergen BK, Lichtenstein PV. Grammatical aspect, gesture, and conceptualization: Using co-speech gesture to reveal event representations. Cognitive Linguistics. 2013; 24: 135–158. https://doi.org/10.1515/cog-2013-0005

32. Hinnell J. The multimodal marking of aspect: The case of five periphrastic auxiliary constructions in North American English. Cognitive Linguistics. 2018; 29: 773–806.

33. Lempert M. Barack Obama, being sharp: Indexical order in the pragmatics of precision-grip gesture. Gesture. 2011; 11: 241–270. https://doi.org/10.1075/gest.11.3.01lem

34. Cánovas CP, Valenzuela J, Carrión DA, Olza I, Ramscar M. Quantifying the speech-gesture relation with massive multimodal datasets: Informativity in time expressions. PLOS ONE. 2020; 15: e0233892. https://doi.org/10.1371/journal.pone.0233892 PMID: 32484842

35. Steen FF, Turner M. Multimodal Construction Grammar. Language and the Creative Mind. Stanford, California: CSLI Publications; 2013. pp. 255–274.

36. Mesirov JP. Computer science. Accessible reproducible research. Science. 2010; 327: 415–416. https://doi.org/10.1126/science.1179653 PMID: 20093459

37. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019. Available: https://www.R-project.org/

38. RStudio Team. RStudio: Integrated development for R. Boston, Massachusetts: R Studio Inc.; 2016. Available: http://www.rstudio.com/

39. Wickham H. rvest: Easily harvest (scrape) web pages. 2020. Available: https://CRAN.R-project.org/package=rvest

40. Lang DT, Kalibera T. XML: Tools for parsing and generating XML within R and S-Plus. 2020. Available: https://CRAN.R-project.org/package=XML

41. Ooms J. The jsonlite package: A practical and consistent mapping between JSON data and R objects. arXiv:14032805 [cs, stat]. 2014 [cited 23 Sep 2020]. Available: http://arxiv.org/abs/1403.2805

42. Kendon A. Gesture: visible action as utterance. Cambridge; New York: Cambridge University Press; 2004.

43. Müller C. Forms and uses of the Palm Up Open Hand: A case of a gesture family. In: Müller C, Posner R, editors. The semantics and pragmatics of everyday gestures. Berlin: Weidler; 2004. pp. 233–256.

44. Cooperrider K, Abner N, Goldin-Meadow S. The palm-up puzzle: Meanings and origins of a widespread form in gesture and sign. Front Commun 3: 23 https://doi.org/103389/fcomm 2018.

45. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20: 37–46.

46. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012; 22: 276–282. PMID: 23092060

47. Wickham H. tidyverse: Easily install and load the "tidyverse". 2017. Available: https://CRAN.R-project.org/package=tidyverse

48. Fernández-i-Marín X. ggmcmc: Analysis of MCMC Samples and Bayesian Inference. J Stat Soft. 2016; 70. https://doi.org/10.18637/jss.v070.i09

49. Wickham H. scales: Scale functions for visualisation. 2018. Available: https://CRAN.R-project.org/package=scales

50. Bürkner P-C. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software. 2017; 80: 1–28. https://doi.org/10.18637/jss.v080.i01

51. Bürkner P-C. Advanced Bayesian multilevel modeling with the R package brms. The R Journal. 2018; 10: 395–411.

52. Gamer M, Lemon J, Fellows I, Singh P. irr: Various Coefficients of Interrater Reliability and Agreement. 2019.

53. Hassemer J. Towards a Theory of Gesture Form Analysis: Imaginary Forms as Part of Gesture Conceptualisation, with Empirical Support from Motion-Capture Data. PhD, RWTH Aachen University. 2016.

54. Badets A, Andres M, Di Luca S, Pesenti M. Number magnitude potentiates action judgements. Exp Brain Res. 2007; 180: 525–534. https://doi.org/10.1007/s00221-007-0870-y PMID: 17279382

55. Black M. More about metaphor. Metaphor and Thought. Cambridge: Cambridge University Press; 1993. pp. 19–41.

56. Murphy GL. On metaphoric representation. Cognition. 1996; 60: 173–204. https://doi.org/10.1016/0010-0277(96)00711-1 PMID: 8811744

57. Murphy GL. Reasons to doubt the present evidence for metaphoric representation. Cognition. 1997; 62: 99–108. https://doi.org/10.1016/s0010-0277(96)00725-1 PMID: 8997172

58. Holyoak KJ, Stamenković D. Metaphor comprehension: A critical review of theories and evidence. Psychological Bulletin. 2018; 144: 641. https://doi.org/10.1037/bul0000145 PMID: 29517263

59. Bouissac P. The study of metaphor and gesture: A critique from the perspective of semiotics. In: Cienki A, Müller C, editors. Metaphor and Gesture. 2008. pp. 277–282.

60. Winter B, Matlock T. Primary metaphors are both cultural and embodied. Metaphor: Embodied Cognition and Discourse. Cambridge: Cambridge University Press; 2017. pp. 99–116. https://doi.org/10.1017/9781108182324.007

61. Landau MJ, Meier BP, Keefer LA. A metaphor-enriched social cognition. Psychol Bull. 2010; 136: 1045–1067. https://doi.org/10.1037/a0020970 PMID: 20822208

62. Casasanto D, Dijkstra K. Motor action and emotional memory. Cognition. 2010; 115: 179–185. https://doi.org/10.1016/j.cognition.2009.11.002 PMID: 20096831

63. Calbris G. The Semiotics of French Gestures. Bloomington, Indiana: Indiana University Press; 1990.

64. Cienki A.J. Metaphoric gestures and some of their relations to verbal metaphoric expressions. In: Koenig Jean-Pierre, editor. Discourse and Cognition: Bridging the Gap. Amsterdam: CSLI Publications; 1998. pp. 189–204.

65. Dehaene S, Bossini S, Giraux P. The mental representation of parity and number magnitude. Journal of Experimental Psychology: General. 1993; 122: 371–396.

66. Östling R, Börstell C, Courtaux S. Visual Iconicity across sign languages: Large-scale automated video analysis of iconic articulators and locations. Front Psychol. 2018; 9. https://doi.org/10.3389/fpsyg.2018.00725 PMID: 29867684

67. Cao Z, Hidalgo G, Simon T, Wei S-E, Sheikh Y. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. arXiv:181208008 [cs]. 2018 [cited 16 Oct 2019]. Available: http://arxiv.org/abs/1812.08008

68. Cao Z, Simon T, Wei S-E, Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields. arXiv:161108050 [cs]. 2016 [cited 16 Oct 2019]. Available: http://arxiv.org/abs/1611.08050

69. Börstell C, Lepic R. Spatial metaphors in antonym pairs across sign languages. Lillo-Martin D, Sandler W, Coppola M, Stamp R, editors. Sign Language & Linguistics. 2020; 23: 111–140.

70. Pouw W, Trujillo JP, Dixon JA. The quantification of gesture–speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. Behav Res. 2020; 52: 723–740. https://doi.org/10.3758/s13428-019-01271-9 PMID: 31659689

71. Çatak EN, Açık A, Göksun T. The relationship between handedness and valence: A gesture study. Q J Exp Psychol. 2018; 71: 2615–2626. https://doi.org/10.1177/1747021817750110 PMID: 29355469

Additional comments

1. The two-handed gestures discussed on page 11 either had symmetrical or asymmetrical hand configurations. We did not make this distinction in our analysis of two-handed gestures, because the existing research literature led us to predict that two-handed gestures would be preferred for larger quantities, while one-handed gestures would be preferred for smaller quantities, irrespective of their hand configuration. However, our data (https://osf.io/dncjg/) can be explored to look at whether hand configurations for two-handed gestures were more likely to be symmetrical or asymmetrical, and whether this distinction has an effect on our results.
2. On page 15, we use the word 'default' to denote the default for gestures in general, for primarily mechanical reasons: "because the hands may tend to be in an approximately open hand configuration while at rest, making it easier to perform an open-hand gesture than a closed-hand one" (p. 15). Because open-hand gestures are arguably easier to perform, it may be used 'by default' unless this default is overridden. We suggest that the default may be overridden if there is a sufficiently strong activation between a numerical quantity and closed-hand actions, causing speakers to produce a different kind of gesture to the default (e.g., a closed-hand gesture).
3. On page 16, we talk about expressions like 'large number' being highly conventionalized, with gestures being treated as a way of deeming whether these expressions are conceptually active, as metaphoric gestures are seen as less conventionalized. However, we should acknowledge that some gestures are highly conventionalized, like the 'OK' gesture, and it is possible that metaphoric gestures

differ in the degree to which they are conventionalized. In fact, the systematicity with which we observe certain gestures in our dataset (e.g., pinch gestures for 'small' meanings) indicates a degree of conventionalization. There is an argument to be made, then, that these expressions are stored together with the corresponding gesture in a multimodal construction unit in the brain (e.g., Hinnell, 2018; Steen & Turner, 2013; Zima & Bergs, 2017), at least for some of the speakers, rather than necessarily expressing the conceptualisation processes underlying their language production.

4. On page 17, we talk about the possibility of speakers receiving body language training. The relevance of 'body language training' to gesture is that we present gesture as a means of examining the conceptualisation processes that underlie communication, because gesture is argued to be unconventionalized, produced 'on the fly' in a spontaneous manner. If speakers have been coached to gesture in a certain way, the gestures they produce may not be spontaneous. This argument is not as pertinent to speech itself, because, as we acknowledge, words may be so conventionalized that they may not reflect cognition, which partly motivated our decision to look at gestures in the first place. Of course, we do not know the effect of body language training on co-speech gesture for the speakers in our dataset, and we can only speculate as to its influence more generally. It is my personal belief that it probably does not exert too much effect, because attempting to control one's gestures while producing speech requires substantial effort. Moreover, most of the videos in the dataset appear to be fairly spontaneous speech rather than, for example, highly practised speeches.

4. PAPER 2

Woodin, G., Winter, B. & Padilla, L. (2022). Conceptual metaphor and graphical convention influence the interpretation of line graphs. *IEEE Transactions on Visualization and Computer Graphics*, 28(2): 1209–1221. http://doi.org/10.1109/TVCG.2021.3088343

- **Conceptualization:** Greg Woodin, Lace Padilla.
- **Data curation:** Greg Woodin, Lace Padilla.
- **Formal analysis:** Greg Woodin.
- **Investigation:** Greg Woodin, Lace Padilla.
- **Methodology:** Greg Woodin, Lace Padilla.
- **Project administration:** Greg Woodin, Lace Padilla.
- **Supervision:** Bodo Winter, Lace Padilla.
- **Validation:** Bodo Winter, Lace Padilla.
- **Visualization:** Greg Woodin, Lace Padilla.
- **Writing – original draft:** Greg Woodin.
- **Writing – review & editing:** Greg Woodin, Bodo Winter, Lace Padilla.

Supplementary materials: https://osf.io/5acjs/

Conceptual Metaphor and Graphical Convention Influence the Interpretation of Line Graphs

Greg Woodin, *Member, IEEE*, Bodo Winter, and Lace Padilla, *Member, IEEE*

Abstract—Many metaphors in language reflect conceptual metaphors that structure thought. In line with metaphorical expressions such as 'high number', experiments show that people associate larger numbers with upward space. Consistent with this metaphor, high numbers are conventionally depicted in high positions on the y-axis of line graphs. People also associate good and bad (emotional valence) with upward and downward locations, in line with metaphorical expressions such as 'uplifting' and 'down in the dumps'. Graphs depicting good quantities (e.g., vacation days) are consistent with graphical convention and the valence metaphor, because 'more' of the good quantity is represented by higher y-axis positions. In contrast, graphs depicting bad quantities (e.g., murders) are consistent with graphical convention, but not the valence metaphor, because more of the bad quantity is represented by higher (rather than lower) y-axis positions. We conducted two experiments (N = 300 per experiment) where participants answered questions about line graphs depicting good and bad quantities. For some graphs, we inverted the conventional axis ordering of numbers. Line graphs that aligned (versus misaligned) with valence metaphors (up = good) were easier to interpret, but this beneficial effect did not outweigh the adverse effect of inverting the axis numbering. Line graphs depicting good (versus bad) quantities were easier to interpret, as were graphs that depicted quantity using the x-axis (versus y-axis). Our results suggest that conceptual metaphors matter for the interpretation of line graphs. However, designers of line graphs are warned against subverting graphical convention to align with conceptual metaphors.

Index Terms—Conceptual metaphor theory, more is up, mental number line, cognition, linguistics, emotional valence, line graph, axis reversal, handedness, empirical evaluation

1 Introduction

Many data visualizations tap into existing cognitive associations. For example, it is conventional for numbers plotted on the y-axis of graphs to increase in magnitude from the bottom to the top of this axis [1], [2]. Consistent with this graphical convention, a large body of scientific research demonstrates that people tend to associate upward vertical locations with larger numerical quantities [3], [4], [5]. For instance, when participants are asked to spontaneously generate random sequences of numbers, they produce larger numbers when moved upward in a body-lifting device than when moved downward [6]. This empirical finding suggests that the upward ordering of numbers along the y-axis of graphs may not be arbitrary. Rather, this vertical orientation may have become established as conventional because it aligns with our vertical conceptualization of numerical quantities, making the resultant graphs easier to interpret [7], [8].

Cognitive associations where one concept (e.g., numerical quantity) is conceptualized in terms of another (e.g., vertical space) have been termed *conceptual metaphors* [9], [10], [11]. Conceptual metaphor theory is a framework based in cognitive science, linguistics, and psychology. The central tenet of conceptual metaphor theory is that metaphors in language are surface-level representations of deeper conceptual metaphors that structure the way we think about various concepts, including numerical quantity [3], [4], [12] and time [13], [14], [15]. For example, it is conventional for English speakers to describe numbers metaphorically as 'low', 'high', 'plummeting', or 'soaring' [11]. This linguistic pattern reflects the conceptual metaphor where larger quantities are associated with higher vertical positions. As well as language, the conventional design of data visualizations (e.g., the y-axis ordering of numbers) may reflect certain conceptual metaphors.

In the present study, we explored a subset of conceptual metaphors that may underlie graphical conventions in line graphs, focusing on metaphoric representations of numerical quantity and time. We investigated how the interpretation of line graphs is affected by the subversion of graphical conventions. We also explored whether line graph interpretation is influenced by conceptual metaphors of *emotional valence* [19], [20], [21], a term referring to whether concepts are perceived to be good or bad [22], [23], [24]. Emotional valence is relevant to the interpretation of data visualizations because, in many cases, the quantity depicted in a graph can be understood as good (e.g., vacation days) or bad (e.g., murders).

Our interest in emotional valence was inspired by a particular data visualization, widely discussed in the media [25], [26], blogs [27], [28], and scientific literature [1], [29], [30]. A graph designed in 2014 showed that gun deaths had increased since the introduction of Florida's Stand Your

- *Bodo Winter and Greg Woodin are with the Department of English Language and Linguistics, University of Birmingham, Birmingham B15 2SQ, U.K. E-mail: bodo@bodowinter.com, gawoodin@gmail.com.*
- *Greg Woodin is with the Spatial Perception, Applied Cognition & Education Lab, University of California Merced, Merced, CA 95343 USA. E-mail: gawoodin@gmail.com.*

Manuscript received 7 Sept. 2020; revised 20 May 2021; accepted 2 June 2021. Date of publication 10 June 2021; date of current version 31 Dec. 2021. (Corresponding author: Greg Woodin.) Recommended for acceptance by P. Isenberg. Digital Object Identifier no. 10.1109/TVCG.2021.3088343

1077-2626 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

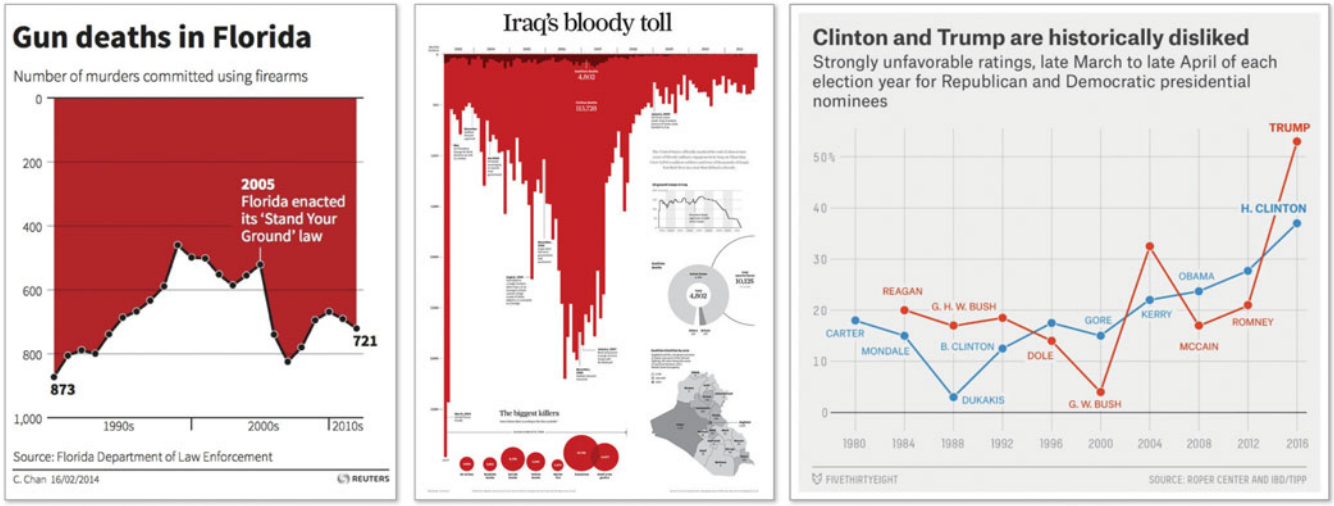


Fig. 1. Real-world data visualization examples that depict 'bad' (i.e., negatively valenced) quantities. A: Stand Your Ground visualization, designed by Christine Chan, published by Reuters [16]. B: Iraq's Bloody Toll visualization, designed by Simon Scarr, published by South China Morning Post [17]. C: Clinton and Trump Are Historically Disliked visualization, designed by Harry Enten, published by FiveThirtyEight [18].

Ground law in 2005, which permitted state residents to use deadly force in self-defence (see Fig. 1, A). The y-axis of this graph was inverted so that numbers increased from top to bottom rather than from bottom to top. While this inversion of death counts subverted graphical convention (the y-axis ordering of numbers), it aligned with experimental evidence showing that negatively valenced (bad) concepts (e.g., more deaths) are associated with lower vertical locations [19], [20], [21]. A previous study conducted by Pandey *et al.* [1] found that inverting the y-axis of graphs had an adverse effect on graph interpretability. However, this study did not consider the emotional valence of the quantities depicted in these graphs.

Motivated by the Stand Your Ground visualization (and other visualizations; see Fig. 1), we conducted a pair of experiments investigating the factors that may affect the interpretation of line graphs depicting time series. We determined whether conceptual metaphors of emotional valence affect the interpretation of line graphs, and, if so, whether the extent of this influence is sufficient to justify departing from graphical convention (i.e., by inverting the axes of graphs). In doing so, we offer recommendations for designers with regard to how to spatially organise line graphs. The key contributions of this work are to:

- Provide a summary of existing research into a subset of conceptual metaphors that are relevant to the data visualization community (numerical quantity, time, and emotional valence)
- Explore the effect that subverting graphical convention has on the interpretation of line graphs (preregistered, conceptual replication-extension of Pandey *et al.* [1])
- Investigate the influence of conceptual metaphors of emotional valence on the interpretation of line graphs
- Determine whether designing line graphs to align with conceptual metaphors of emotional valence is justified, given that subverting graphical convention may adversely affect graph interpretability [1]

In the following sections, we begin by providing an overview of research into certain conceptual metaphors and their potential relevance to the design of data visualizations. We then describe two experiments. The first experiment focused on conceptual metaphors of emotional valence relevant to the ordering of numbers along the y-axis. The second experiment broadened this scope to include emotional valence metaphors relevant to the ordering of numbers along the x-axis. Our findings provide evidence that, all else being equal, designing line graphs to align with conceptual metaphors of emotional valence does have a beneficial effect on graph interpretability. However, the beneficial effect of this valence alignment on graph interpretability does not outweigh the adverse effect of subverting graphical convention. Based on this result, we recommend that line graphs should be designed to align with conceptual metaphors of emotional valence only if doing so does not contradict other well-established graphical conventions.

2 Background

As argued by Parsons [7], data visualizations can be made easier to understand by designing them with human psychology in mind, particularly by using knowledge of conceptual metaphors to inform design choices. Research in cognitive science, linguistics, and psychology has produced a plethora of evidence for the effect of conceptual metaphors on perception and behaviour (e.g., [6], [20], [31]). Scholars have previously discussed conceptual metaphors in the context of visualization research [2], [32], [33]. However, Parsons [7] notes that these scholars typically have not used metaphor research to offer actionable suggestions for visualization design (but see [8] for an exception). In the present study, we addressed this lack with an empirical investigation of a subset of conceptual metaphors that may influence the interpretation of line graphs. We also explored the relationship between conceptual metaphor and graphical convention, and the implications of this relationship for the design of line graphs. The relevance of conceptual metaphors to the data visualization community in general can be understood through the lens of *cognitive fit theory*, which provides a framework for understanding how people interpret visualizations.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

2.1 Cognitive Fit Theory

Cognitive fit theory describes the mental effort required to understand data visualizations [34]. This theory proposes that people interpret visualizations with a mental model of what they expect to see, which informs the strategy used to interpret these visualizations [30]. For instance, the mental model used to interpret line graphs may include the expectation that numbers on the y-axis will increase from bottom to top. The Stand Your Ground visualization depicted in Fig. 1 (A) violates this expectation, reversing the direction of the y-axis so the trend line resembles blood dripping down a wall. This violation of expectation may force graph users to perform a mental transformation of the visualization, re-ordering the numbers on the y-axis to match their mental model. Such mental transformations are likely to increase cognitive processing costs, making the graph more difficult to interpret. For some graphs, the user may even arrive at an incorrect conclusion about the trend that the visualization depicts [29], [35], [36]. When viewing the Stand Your Ground visualization, for example, the graph user may mistakenly believe that gun deaths decreased rather than increased following the introduction of the Stand Your Ground law. Mental models may be informed by both graphical conventions and conceptual metaphors [7]. The first category of conceptual metaphor we discuss in this paper is relevant to the representation of numerical quantity and emotional valence along the y-axis of graphs – *vertical metaphors*.

2.2 Vertical Metaphors: Quantity and Valence

Research shows that people tend to associate upward vertical locations with larger numerical quantities and downward vertical locations with smaller numerical quantities [3], [6], [37]. Participants respond more quickly to larger numbers with a response button positioned higher in space and to smaller numbers with a response button positioned lower in space [3]. When participants are asked to spontaneously generate random sequences of numbers, they produce larger numbers when looking upward rather than downward [4], and when being moved upward rather than downward in a body-lifting device [6]. Gaze position also tends to shift upward when people count aloud in an ascending sequence [37]. This vertical conceptualization of numerical quantities is evident in the linguistic metaphors that English speakers use to talk about them. For example, speakers often describe numbers as 'low' or 'high', and may refer to 'plummeting shares' and 'soaring costs' [11].

Prior work suggests that the metaphoric association of larger numerical quantities with upward locations may form the basis for the conventional ordering of numbers upward along the y-axis of graphs [7], [8]. The benefit of graphical conventions aligning with conceptual metaphors may be that the resultant graphs fit a pre-existing mental model for thinking about numerical quantity and other concepts, decreasing cognitive processing costs and making these graphs intuitive and easy to understand [7].

There are, however, instances in which different conceptual metaphors make competing recommendations for how graphs should be designed. For example, the Stand Your Ground visualization (see Fig. 1, A) subverts the metaphoric association between larger numerical quantities and upward locations that is enshrined in graphical convention. However, this subversive design choice aligns with the vertical metaphor of emotional valence where positively valenced (good) concepts (e.g., fewer deaths) are associated with upward locations, whereas negatively valenced (bad) concepts (e.g., more deaths) are associated with downward locations. Evidence for this vertical metaphor of emotional valence shows that positively valenced words (e.g., 'pride') are evaluated more quickly when they appear in a higher position on a computer screen. Conversely, negatively valenced words (e.g., 'liar') elicit faster responses when they appear in a lower screen position [20]. Furthermore, participants recollect more positively valenced autobiographical memories when moving marbles upward from a lower box to a higher box than when performing the reverse vertical action [21], [38]. This conceptual metaphor is evident in linguistic expressions such as 'in high spirits', 'down in the dumps', 'uplifting', and 'downbeat' [11]. Conceptual metaphors of emotional valence are potentially important for graph comprehension because many quantities depicted in graphs have strong emotional connotations, such as income (increase = good) and infection rates (increase = bad) [39]. The fact that quantities depicted in graphs are often emotionally connotative raises the question of whether designers should factor vertical valence metaphors into the design of their visualizations.

The principal aim of the experiments reported in this paper was to establish whether conceptual metaphors of emotional valence influence the interpretation of line graphs. If they do, we investigated whether the influence of emotional valence is strong enough to justify inverting the y-axis of graphs, a design choice that subverts the conventional vertical ordering of numbers. A study conducted by Pandey *et al.* [1] has already shown that reversing the quantity axis of graphs – such that larger numerical quantities are represented by downward locations – has an adverse effect on the accuracy of responses to these graphs. Our experiments partly functioned as a preregistered, conceptual replication of this previous investigation, while also considering the effect of emotional valence. We showed participants line graphs with or without an inverted axis and asked these participants what the trend (i.e., the change in the quantity depicted in the graph over time) showed. We took the accuracy of responses to this question as a proxy for the interpretability of the graphs. Errors were taken to indicate that the graph in question contradicted the mental model that participants used to interpret the graph, in line with cognitive fit theory. We also extended this line of questioning to other conceptual metaphors, specifically ones that are relevant to the horizontal axis – *horizontal metaphors*.

2.3 Horizontal Metaphors: Quantity and Valence

Numerical quantities are not only conceptualized using the vertical axis, but also using the horizontal axis. The Spatial-Numerical Association of Response Codes (SNARC) effect shows that smaller numbers elicit faster left-hand responses, whereas larger numbers elicit faster right-hand responses [40], [41], [42]. Similarly, in random number generation tasks, leftward eye movements can be used to reliably predict that participants will generate a smaller number than the previous number in the sequence [5]. In addition, participants asked to place numbers anywhere on a piece of paper tend to orient the numbers horizontally from left to right [43]. These results

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

have been interpreted as evidence for the existence of a *mental number line* in which people imagine numbers as increasing in magnitude from left to right. This horizontal conceptualization of numerical quantity may be the reason why certain graphs, such as horizontal bar charts, conventionally label numbers from left to right across the x-axis [44].

Unlike vertical metaphors, horizontal associations are typically not evident in language (e.g., we do not describe numbers as being 'lefter' or 'righter' than one another) [4], [45] and so are not always described as conceptual metaphors. However, we treat them as such because they deal with the same kind of cognitive process as prototypical conceptual metaphors that do have linguistic representations [45], [46]. To investigate horizontal metaphors of numerical quantity, in the present study we used line graphs that plotted quantity along the x-axis. We examined whether reversing the conventional direction of this axis – so that the numbers increased in magnitude from right to left – affected graph interpretability.

Like numerical quantity, emotional valence may also be conceptualized horizontally. When participants are asked to sort different items into boxes positioned on the left and right of a piece of paper, right-handers tend to sort positively valenced (good) concepts into the right-side box and negatively valenced (bad) items into the left-side box. Conversely, left-handers tend to do the opposite [47], [48], [49]. This finding is indicative of a conceptual metaphor where the dominant side of the body (and therefore that side of space) is associated with positive valence and the non-dominant side with negative valence. Complicating matters, because most people are right-handed, an association of right-side space with positive valence is evident in cultural practices (e.g., raising one's right hand when taking an oath) and language (e.g., the 'correct' answer is the 'right' one). The representation of right-handed conceptual metaphors in cultural practices and language may result in even some left-handers learning the association between rightward space and positive valence [50], [51].

In the present study, we investigated whether horizontal metaphors of emotional valence matter for the interpretation of line graphs that plot quantity along the x-axis. If so, we explored whether this effect was relative to handedness, or whether it was absolute (i.e., where both left- and right-handers associate right-side space with positive valence). If we found an effect of horizontal valence metaphors on graph interpretability, we also wanted to know whether this effect was significant enough to justify reversing the conventional ordering of numbers along this axis (i.e., ordering numbers from right to left rather than left to right). Previous findings presented by Pandey *et al.* [1] led us to expect that subverting the conventional x-axis ordering of numbers would have an adverse effect on graph interpretability. However, Pandey *et al.* exclusively inverted the y-axis of graphs, so our additional focus on x-axis inversion constituted an extension of this previous study.

2.4 Time Metaphors

The graphs used in the present study were line graphs that depicted changes in quantity over time, based on those used by Pandey *et al.* [1]. In a time series graph, which is one of the most frequently used types of data visualization [52], time is usually depicted as passing from left to right across the x-axis – earlier times are to the left, whereas later times are to the right [53], [54], [55]. The spatial representation of time in line graphs is consistent with horizontal time metaphors. For example, the Spatial-Temporal Association of Response Codes (STEARC) effect indicates that Western participants tend to associate leftward space with earlier times and rightward space with later times. When participants are tasked with indicating whether the timing of a sound is earlier or later than expected based on the temporal regularity of preceding sounds, participants respond to earlier sounds more quickly with a left-side button and to later sounds more quickly with a right-side button [31], [56] (see also [57]).

There are also vertical metaphors associated with time, with limited evidence suggesting that time is conceptualized as passing upward through space. For instance, one study showed that participants responded more quickly to future-related words when they appeared higher on a computer screen, and vice versa for past-related words [58] (see also [59]). In the present study, our use of line graphs depicting time series allowed us to test whether inverting the time axis of these graphs adversely affected graph interpretability. Our decision to invert the time axes of graphs extended the methodology used by Pandey *et al.* [1], who exclusively inverted the quantity axis of graphs.

2.5 Quantity Mapping

Our experimental stimuli included graphs in which quantity was 'mapped' onto either the vertical or horizontal axis, which we refer to as the graph's *quantity mapping*. We were therefore interested to see which quantity mapping would elicit more accurate responses. Previous research suggests that numbers are more readily conceptualized using the horizontal axis [43], [60]. However, results reported by Fischer *et al.* [44] suggest that vertically oriented bar charts are easier to interpret than horizontally oriented ones. Because line graphs conventionally depict quantity using the vertical axis, we predicted that graphs observing this convention would be easier to interpret. We predicted that graphs that subverted this convention by depicting quantity using the horizontal axis would be more difficult to interpret. Our choice to map quantity onto either the y-axis or x-axis also allowed us to test both horizontal and vertical representations of time.

2.6 Negativity Bias

A final possibility explored in our experiments was that the emotional valence of the quantity depicted in a graph matters by itself, irrespective of whether the graph aligns with valence metaphors. This possibility is suggested by the fact that people are quicker to recognise pleasant and safe stimuli than unpleasant or threatening stimuli [61], [62], [63]. Researchers have argued that negatively valenced stimuli contain more information than positively valenced stimuli, requiring greater attention and cognitive processing. This phenomenon has been termed the *negativity bias* [64], [65]. Based on the fact that positively valenced stimuli are processed more quickly and easily than negatively valenced stimuli, graphs depicting positively valenced quantities (e.g.,

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

TABLE 1
Hypothesis 4: Vertical Valence Alignment

Stimuli	Valence Aligning (y-axis)
non-inverted y-axis, positive valence	yes
non-inverted y-axis, negative valence	no
inverted y-axis, positive valence	no
inverted y-axis, negative valence	yes

vacation days) may have a cognitive processing advantage over graphs depicting negatively valenced quantities (e.g., murders).

3 Empirical Evaluation

We now report two experiments in which participants were asked questions about line graphs depicting a change in quantity over time. Experiment 1 focused on vertical valence metaphors. Experiment 2 served as a replication of Experiment 1, but also collected data from an equal number of left- and right-handers, allowing us to test horizontal valence metaphors.

3.1 Experiment 1

In this experiment, we tested the following hypotheses:

H1. Graphs without inverted axes will be easier to interpret than graphs with an inverted axis (preregistered, conceptual replication-extension of Pandey *et al.* [1]).

H2. Graphs that map quantity information onto the y-axis will be easier to interpret than graphs that map quantity information onto the x-axis (preregistered, conceptual replication of Fischer *et al.* [44] with line graphs rather than bar charts).

H3. Graphs depicting a positively valenced quantity (vacation days) will be easier to interpret than graphs depicting a negatively valenced quantity (murders).[1]

H4. Graphs aligning with vertical valence metaphors will be easier to interpret than graphs that do not align with these metaphors (see Table 1).

It should be stated from the outset that the experiments we present in this paper cannot disentangle the relative effects of graphical convention and conceptual metaphor where both make identical recommendations for how graphs should be designed. For example, both graphical convention and conceptual metaphor dictate that numbers should increase upward along the y-axis of graphs, and that time should pass from left to right along the x-axis. If we were to find that inverting these axes had an adverse effect on graph interpretability, we would not be able to say whether this adverse effect were due to graphical convention, conceptual metaphor, or both. Therefore, although previous research has led us (and others [2], [32], [33]) to predict that conceptual metaphors shaped the formation of graphical conventions relating to numerical quantity and time, the present investigation is not intended to test this prediction. The only conceptual metaphors we investigate directly in this paper relate to emotional valence.

1. Hypothesis 3 was not preregistered prior to Experiment 1 but we include it here for consistency with Experiment 2 and as an additional test of this hypothesis.

3.1.1 Procedure

The study was implemented using survey software Qualtrics [65] and distributed via Amazon Mechanical Turk [67], a crowd-sourcing marketplace where workers receive payment for completing tasks virtually. Our experimental procedure was inspired by Pandey *et al.* [1], particularly the aspect of their study that concerned the inversion of graphical axes. Participants answered multiple-choice questions about line graphs depicting made-up trends over time in fictional cities. The names of these cities were randomly generated using an online town name generator (https://www.namegenerator2.com/town-name-generator.php), with the criterion that each name should be six characters long. For example, one graph depicted the number of murders in Torley between 2015 and 2018 (see Fig. 2, left). Each graph showed a trend for a different city. Participants were asked: 'What can you say about the situation in this city?' Participants then indicated whether they believed the situation to be improving, declining, or neither improving nor declining (three multiple-choice options). In the Torley graph (Fig. 2, left), the number of murders in this city increased over time, indicating that the situation was declining. In contrast, if the graph showed a rising trend for the average number of vacation days (see trial types in Fig. 2, middle), the correct answer would be that the situation in the city was improving. The correct answer was never that the situation was neither improving nor declining.

Of primary relevance to our hypotheses was whether participants responded correctly or incorrectly. A higher proportion of correct responses to a graph was taken to indicate that this graph was easier to interpret than a graph that elicited a lower proportion of correct responses. The latencies of participants' responses were also recorded and investigated exploratorily.

The experiment proceeded as follows. In accordance with approved IRB protocol, after giving consent, participants were shown the instructions to the study (available at https://osf.io/5acjs/). These instructions included an exemplar graph that depicted the number of cars over time in a fictional city called Murell. Participants then answered an attention check question, being asked to accurately recall a particular word in the instructions they had just been shown. Participants who responded incorrectly to the attention check question were disqualified from completing the remainder of the study. Participants then completed four main experimental trials (see Fig. 2, middle), after which they supplied their demographic information.

3.1.2 Stimuli

We used a mixed between- and within-participants design. As shown in Fig. 2 (right), the type of graph shown to each participant was varied in a 3×2 design according to the following two factors:

- *Axis Orientation*: whether the graph (i) does not have inverted axes (right, top row in Fig. 2), (ii) has an inverted y-axis (middle row), or (iii) has an inverted x-axis (bottom row);
- *Quantity Mapping*: whether the graph maps quantity information onto its (i) y-axis (left column) or (ii) x-axis (right column).

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

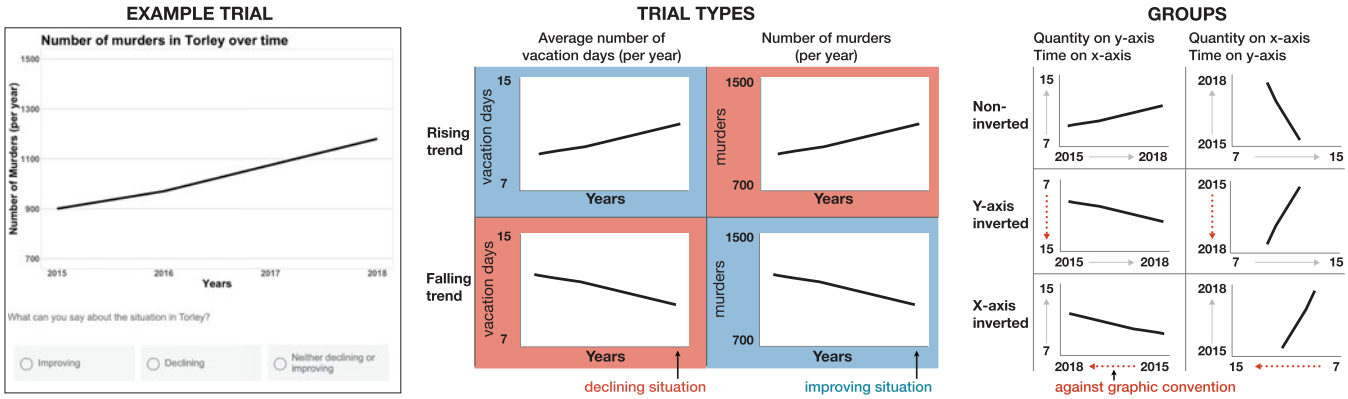


Fig. 2. Left: Example of a trial completed by participants. All graphs (four per group) can be viewed at https://osf.io/5acjs. Middle: Trend (rising versus falling) and valence (positive versus negative) were varied within participants to create the four trial types shown to participants. Right: Each participant took part in one of six groups, each of which contained graphs that had been manipulated in a particular way. Axis Orientation (non-inverted versus y-axis inverted versus x-axis inverted) and Quantity Mapping (y-axis versus x-axis) were varied between participants to create these groups.

These manipulations were conducted between participants to prevent participants from being cued to focus on the graph's axes more than they would naturally. This 3×2 design resulted in there being six groups, which were run as separate experiments on Amazon Mechanical Turk. Participants self-selected which group to participate in, rather than there being a strict randomization procedure. Participants were unaware of the group they were in as each person was only able to view one experiment. Responses from each group were collected until responses in each group reached the criterion for our preregistered stopping rule ($N = 50$).

Within each group, each participant completed four trial types, as shown in Fig. 2 (middle). This meant that each participant contributed four data points to the final data set. To generate the four trials within each group, we manipulated the following two factors within participants in a 2×2 design:

- *Trend*: whether the quantity depicted in the graph was (i) rising (middle, top row in Fig. 2) or (ii) falling (bottom row) over time;
- *Valence*: whether the quantity depicted in the graph was (i) positively valenced (left column) or (ii) negatively valenced (right column).

We chose 'vacation days' as a positively valenced quantity and 'murders' as a negatively valenced quantity. These two concepts were selected from Warriner *et al.*'s [68] list of 13,195 English words normed for emotional valence. 'Vacation' was the most positively valenced word in the data set, according to the ratings. 'Murders' was the seventh most negatively valenced word. The six words assigned more negatively valenced ratings were avoided on the basis that they were emotionally sensitive or offensive ('pedophile', 'rapist', 'AIDS', 'torture', 'leukemia', 'molester'). The simple trend lines shown in each graph all resembled the trend lines depicted in Fig. 2 and were based on the stimuli used by Pandey *et al.* [1]. The full range of graph stimuli shown to participants can be viewed at https://osf.io/5acjs/.

In sum, each participant saw the four trial types show in Fig. 2 (middle), presented in a randomized order. Each participant saw two graphs in which the situation in the displayed city was improving, and two graphs in which the situation was declining. Each of these graphs was labelled with a different fictional city name.

3.1.3 Participants

Access to Amazon Mechanical Turk requires users to be 18 years or older, restricting our sample to adult participants. Participation was restricted to US users of the platform, and payment was $1 per participant based on the US state of Illinois's minimum wage laws. Three hundred MTurk users volunteered to participate. Of this total, seven participants were excluded for answering an attention check question incorrectly prior to the study. Furthermore, three participants were excluded because their response time to one of the four trials was more than two standard deviations above the mean response time to all trials across all participants (>55.9 seconds). In total, data from 290 participants were included in the final analyses. For these participants, the age range was 24-72 years old ($M = 39$, $SD = 11$). In terms of gender, 130 participants were female (44.8 percent), 159 participants were male (54.8 percent), and one participant (0.3 percent) was non-binary/third gender.

3.1.4 Statistical Analysis

All statistical analyses were conducted using statistical programming language R, version 4.0.3 [69], inside the integrated development environment RStudio 1.1.456 [70]. The packages 'plyr' 1.8.6 [70] and 'tidyverse' 1.3.0 [72] were used for data processing and visualization. The packages 'ggmcmc' 1.5.0 [73], and 'tidybayes' 2.3.1 [74] were used for data visualization. The package 'brms' 2.14.4 [75], [76] was used for Bayesian modelling.

The data and analysis scripts for this study are available at the publicly accessible Open Science Framework repository: https://osf.io/5acjs/. The hypotheses and planned analyses for this study were preregistered and can be found at the aforementioned URL. We used Bayesian multi-level logistic regression to assess our hypotheses. The dependent variable in all statistical models was Accuracy (incorrect versus correct). Accuracy corresponded to whether the participant correctly judged the situation in each city to be either improving or declining by selecting the appropriate multiple choice option. Valence was the only variable relevant to our hypotheses that varied within participants (each participant answered questions about graphs depicting both positively and negatively valenced quantities). To account for by-participant variation in

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

TABLE 2
Accuracy: Experiment 1

Graph Stimuli	Accuracy (Experiment 1)
non-inverted	77.8% *308 vs 87*
inverted	56.1% *431 vs 337*
quantity on y-axis	55.0% *317 vs 259*
quantity on x-axis	71.7% *419 vs 165*
positive valence	72.9% *423 vs 157*
negative valence	54.0% *313 vs 267*
valence aligning: y-axis	66.7% *192 vs 96*
valence misaligning: y-axis	43.4% *125 vs 163*

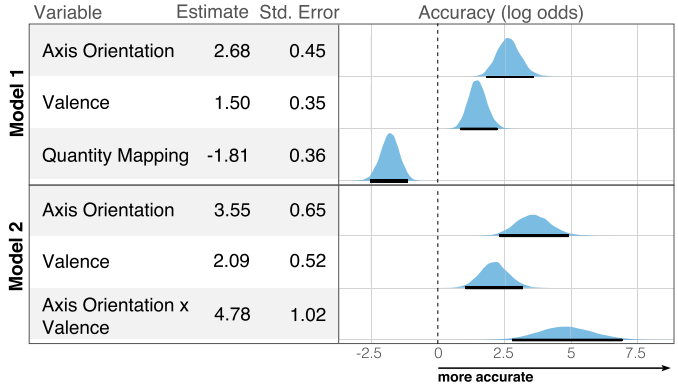


Fig. 3. Estimates, standard errors, and posterior distributions (blue) with 95 percent credible intervals (black lines) for coefficients in Bayesian logistic regression model 1 (all graphs) and 2 (graphs with quantity on y-axis) in Experiment 1.

Accuracy for Valence, random slopes and intercepts for Valence were included in all models. We set weakly informative priors centred at 0 with a standard deviation of 2.

The data from Experiment 1 were analysed using two statistical models. Model 1 assessed hypotheses 1–3 and included the variables Axis Orientation (inverted versus non-inverted), Quantity Mapping (x-axis versus y-axis), and Valence (negative versus positive). These variables were treatment-coded (0, 1). Model 2 assessed hypothesis 4 and included the variables Axis Orientation (inverted versus non-inverted) and Valence (negative versus positive), as well as the interaction between these variables. The variables in model 2 were deviation-coded (-0.5, $+0.5$) to facilitate interpretation of the main effects. Hypothesis 4 was assessed in a separate model because this hypothesis was relevant to only those graphs that plotted quantity along the y-axis.

For each of these models, we observed whether the 95 percent credible intervals for each variable included zero. Credible intervals that did not contain zero were interpreted as showing strong evidence for the effect of the variable on accuracy.

3.1.5 Results

Table 2 shows the accuracy of responses to different graph stimuli in Experiment 1. Fig. 3 shows the coefficients for model 1 (top) and 2 (bottom). The blue posterior distributions show the range of log odds that represent credible values of the coefficient for each variable in the model. Values with a higher probability density, represented by denser areas of the distribution, are more credible than values with a lower probability density. The black lines beneath the distributions represent the 95 percent credible intervals for each variable.

Hypothesis 1 predicted that graphs without inverted axes would be easier to interpret than graphs with either the x- or y-axis inverted [1]. Consistent with this hypothesis, participants were more likely to respond correctly to graphs without inverted axes than to graphs with an inverted axis (see Table 2, rows 1-2). As shown in Fig. 3 (model 1), the credible interval for Axis Orientation did not include zero, indicating that there was strong evidence for the effect of this variable on the accuracy of responses.

Hypothesis 2 predicted that graphs plotting quantity information along the y-axis would be easier to interpret than graphs plotting quantity information along the x-axis [44]. Contradicting this hypothesis, participants were more likely to respond correctly to graphs that represented quantity information along the x-axis rather than the y-axis (see Table 2, rows 3-4). As shown in Fig. 3 (model 1), the 95 percent credible interval for Quantity Mapping did not include zero, indicating that this effect was strongly supported by the data.

Hypothesis 3 predicted that graphs depicting a positively valenced quantity (vacation days) would be easier to interpret than graphs depicting a negatively valenced quantity (murders). Consistent with this hypothesis, participants were more likely to respond correctly to graphs depicting a positively valenced quantity than a negatively valenced quantity (see Table 2, rows 5-6). As shown in Fig. 3 (model 1), the 95 percent credible interval for Valence did not include zero, indicating that this effect was strongly supported by the data.

Hypothesis 4 predicted that graphs that aligned with vertical valence metaphors would be easier to interpret than graphs that did not align with these metaphors. Consistent with this hypothesis, participants were more likely to respond correctly to graphs that aligned with vertical valence metaphors compared to graphs that did not (see Table 2, rows 7-8). As shown in Fig. 3 (model 2), the 95 percent credible interval for Axis Orientation $\times$ Valence did not include zero, indicating that this effect was strongly supported by the data.

The main effects for model 2 predicted responses to be most accurate for non-inverted graphs depicting a positively valenced quantity (95 percent CI = [3.77, 7.08]). Non-inverted graphs depicting a negatively valenced quantity were predicted to elicit the second most accurate responses, although accuracy was substantially lower (95 percent CI = [-0.29, 2.08]). Predicted accuracy for inverted graphs depicting a positively valenced quantity (95 percent CI = [-1.39, 0.06]) and inverted graphs depicting a negatively valenced quantity (95 percent CI = [-1.16, 0.48]) were similar. These main effects suggest that Axis Orientation (inverted versus non-inverted) was the primary determinant of accuracy for graphs that mapped quantity information onto the y-axis, rather than Valence (negative versus positive) or Valence Alignment (misaligning versus aligning).

3.1.6 Discussion

Our results showed that participants were more likely to respond accurately to graphs without inverted axes compared

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

to graphs with an inverted axis. This finding conceptually replicates the results reported by Pandey *et al.* [1] and extends these results to graphs with an inverted x-axis (rather than an inverted y-axis) and an inverted time axis (rather than an inverted quantity axis). Participants were also more likely to respond accurately to graphs that aligned with vertical valence metaphors compared to graphs that did not align with these metaphors. This result indicates that valence alignment matters for the interpretation of graphs that map quantity information onto the y-axis. Because our results show that inverting the axes of graphs generally has an adverse effect on response accuracy, a follow-up question is whether valence alignment matters enough to justify inverting the y-axis of graphs when these graphs depict negatively valenced quantities. Our results suggest not: responses to inverted graphs depicting negatively valenced quantities were less likely to be accurate than responses to non-inverted graphs depicting negatively valenced quantities. Ultimately, axis orientation with respect to the ordering of quantity along axes was a more important determinant of accuracy than valence alignment.

Our results also showed that participants were more likely to respond accurately to graphs depicting a positively valenced quantity than graphs depicting a negatively valenced quantity. This result is consistent with the cognitive processing advantage reported for positively valenced stimuli compared to negatively valenced stimuli [61], [62], [63]. In fact, for graphs that mapped quantity onto the y-axis, the experimental condition most likely to elicit accurate responses – by a substantial margin – was a non-inverted graph that depicted a positively valenced quantity (as shown by the main effects). The success of this experimental condition may be attributable to the additive effect of being non-inverted and depicting a positively valenced quantity, as well as the multiplicative effect of this experimental condition aligning with valence metaphors.

Participants were more likely to respond accurately to graphs mapping quantity information onto the x-axis compared to graphs mapping quantity information onto the y-axis. This result is surprising because line graphs conventionally represent quantity information using the y-axis, so the subversion of this convention may be expected to result in lower accuracy. However, the plotting of quantity onto the x-axis is a salient design choice that participants are likely to have spotted precisely due to its subversion of convention. The novelty of this design choice may have prompted participants to deliberate for longer when responding to this type of graph to verify that their answer was correct [77], [78]. In contrast, when the graph presented to participants mapped quantity onto its y-axis, consistent with graphical convention, participants may have been less likely to initiate a lengthy verification procedure, leaving room for errors to creep in. The latencies of participants' responses lend credence to this interpretation: participants spent less time responding to graphs that plotted quantity information using the y-axis (mean = 3.7 seconds) compared to graphs that plotted quantity information using the x-axis (mean = 5.5 seconds). However, it should be noted that there was little evidence for a general speed-accuracy trade-off in either experiment (see additional analyses at https://osf.io/5acjs/).

3.2 Experiment 2

Experiment 2 served as a replication-extension of Experiment 1. As a replication, we reproduced the experimental procedure from Experiment 1 and observed whether the new data exhibited the same patterns. As an extension, we introduced a test of horizontal valence metaphors. We investigated handedness-relative valence metaphors [47], which suggest that people associate the dominant side of their body (and therefore that side of space) with more positive valence, and vice versa. We also investigated the competing hypothesis that the representation of right-handed metaphors in cultural practices may lead even left-handers to associate positive valence with right-side space, a non-relative, 'absolute' metaphor [50], [51].

TABLE 3
Hypothesis 5 & 6: Horizontal Valence Alignment

Stimuli	Valence Aligning (x-axis)	
	Left-handed	**Right-handed**
non-inverted axis, positive valence	no	yes
non-inverted axis, negative valence	yes	no
inverted axis, positive valence	yes	no
inverted axis, negative valence	no	yes

Experiment 2 tested the following two hypotheses in addition to Hypotheses 1–4 tested in Experiment 1:

H5. Graphs that align with handedness-relative horizontal valence metaphors will be easier to interpret than graphs that do not align with these metaphors (see Table 3).

H6. Graphs that align with absolute horizontal valence metaphors where 'good' is associated with rightward locations, irrespective of handedness, will be easier to interpret than graphs that do not align with these metaphors.

3.2.1 Methodology

The procedure for Experiment 2 followed a similar method to that detailed for Experiment 1. To assess our hypothesis regarding handedness-relative metaphors of emotional valence, eligibility criteria for participation specified that 50 percent of participants in each experimental condition (N = 25) should self-identify as left- and 50 percent as right-handed based on demographic data that participants provided to MTurk.

Three hundred MTurk users volunteered to participate. Of this total, six participants were excluded for answering an attention check question incorrectly prior to the study. Furthermore, seven participants were excluded because their response time to one of the trials was more than two standard deviations above the mean response time to all trials across all participants (25.9 seconds). Thus, data from 287 participants were included in the final analyses. The reported age range of participants was 23-72 years old (M = 39, SD = 10). In terms of gender, 175 participants were male (61 percent), 108 participants were female (37.6 percent), 2 participants were non-binary/third gender (0.7 percent), and 2 participants preferred not to say (0.7 percent). In terms of handedness, 145 participants were right-handed (50.5 percent) and 142 participants were left-handed (49.5 percent).

Three statistical models were used to analyse the data. Model 1 and 2 used in Experiment 1 were used again in Experiment 2. Model 3 assessed hypotheses 5–6 and contained the variables Axis Orientation (inverted versus non-

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

TABLE 4
Accuracy: Experiment 2

Graph Stimuli	Accuracy (Experiment 2)
non-inverted	82.6% *317 vs 67*
inverted	61.5% *470 vs 294*
quantity on y-axis	59.7% *346 vs 234*
quantity on x-axis	77.6% *441 vs 127*
positive valence	71.6% *411 vs 163*
negative valence	65.5% *376 vs 198*
valence aligning: y-axis	64.5% *187 vs 103*
valence misaligning: y-axis	54.8% *159 vs 131*
valence aligning: x-axis (relative)	75.4% *214 vs 70*
valence misaligning: x-axis (relative)	79.9% *227 vs 57*
valence aligning: x-axis (absolute)	84.9% *214 vs 43*
valence misaligning: x-axis (absolute)	70.4% *200 vs 84*

inverted), Valence (negative versus positive), and Handedness (left versus right). This model also assessed the interaction between Axis Orientation and Valence (hypothesis 6), as well as the three-way interaction between Axis Orientation, Valence, and Handedness (hypothesis 5). The variables in model 3 were deviation-coded (−0.5, +0.5) to facilitate interpretation of the main effects. Hypotheses 5–6 were assessed in a separate model because these hypotheses applied only to graphs that mapped quantity information onto the x-axis.

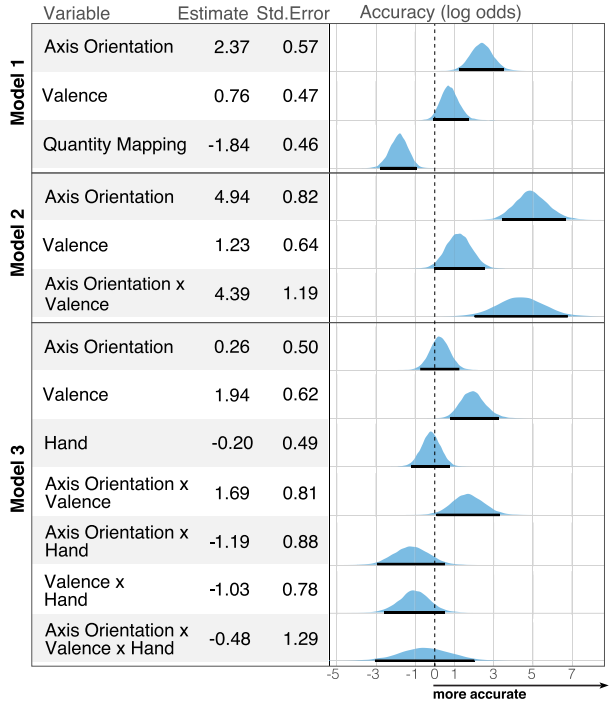


Fig. 4. Estimates, standard errors, and posterior distributions (blue) with 95 percent credible intervals (black lines) for coefficients in Bayesian logistic regression model 1 (all graphs), 2 (graphs with quantity on y-axis), and 3 (graphs with quantity on x-axis) in Experiment 2.

3.2.2 Results

Table 4 shows the accuracy of responses to different graph stimuli in Experiment 1. Fig. 4 shows the coefficients for model 1 (top), 2 (middle), and 3 (bottom). The blue posterior distributions represent the credible values of the coefficients for each variable and the probability density of these values. The black line beneath these distributions represents the 95 percent credible intervals.

Hypothesis 1 predicted that graphs without inverted axes would be easier to interpret than graphs with either the x- or y-axis inverted [1]. Consistent with this hypothesis, participants were more likely to respond correctly to graphs without inverted axes than to graphs with an inverted axis (see Table 4, rows 1-2). As shown in Fig. 4 (model 1), the 95 percent credible interval for Axis Orientation did not include zero, indicating that this effect was strongly supported by the data.

Hypothesis 2 predicted that graphs mapping quantity onto the y-axis would be easier to interpret than graphs mapping quantity onto the x-axis [44]. Contradicting this hypothesis, and in line with the results from Experiment 1, participants were more likely to respond correctly to graphs that represented quantity information along the x-axis than the y-axis (see Table 4, rows 3-4). As shown in Fig. 4 (model 1), the 95 percent credible interval for Quantity Mapping did not include zero, indicating that this effect was strongly supported by the data.

Hypothesis 3 predicted that graphs depicting a positively valenced quantity (vacation days) would be easier to interpret than graphs depicting a negatively valenced quantity (murders). There was a descriptive difference between graphs that depicted a positively valenced quantity and graphs that depicted a negatively valenced quantity (see Table 4, rows 5-6). However, as shown in Fig. 4 (model 1), the 95 percent credible interval for Valence marginally included zero, albeit with the bulk of the credible interval being above zero (lower bound = −0.10).

Hypothesis 4 predicted that graphs aligning with vertical valence metaphors would be easier to interpret than graphs misaligning with these metaphors. Consistent with this hypothesis, participants were more likely to respond correctly to graphs that aligned with vertical valence metaphors than to graphs that did not (see Table 4, rows 7-8). As shown in Fig. 4 (model 2), the 95 percent credible interval for Axis Orientation × Valence did not include zero, indicating that this effect was strongly supported by the data.

The main effects for model 2 predicted accuracy to be highest for non-inverted graphs depicting a positively valenced quantity (95 percent *CI* = [4.53, 8.44]), followed by non-inverted graphs depicting a negatively valenced quantity (95 percent *CI* = [1.47, 4.71]). Accuracy for inverted graphs depicting a negatively valenced quantity (95 percent *CI* = [−0.83, 1.27]) and inverted graphs depicting a positively valenced quantity (95 percent *CI* = [−1.73, 0.22]) were predicted to be substantially lower. These main effects indicate that Axis Orientation (inverted versus normal) was the primary determinant of accuracy for graphs that plotted quantity using the y-axis, rather than Valence (negative versus positive) or Valence Alignment (misaligning versus aligning).

Hypothesis 5 predicted that graphs that align with handedness-relative horizontal valence metaphors would be easier to interpret than graphs that do not align with these metaphors. There was a descriptive difference in accuracy between graphs that did or did not align with handedness-

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

relative horizontal valence metaphors (see Table 4, rows 9-10). However, as shown in Fig. 4 (model 3), the 95 percent credible interval for Axis Orientation × Valence × Hand included zero, indicating that this effect was not strongly supported by the data.

Hypothesis 6 predicted that graphs that align with absolute horizontal valence metaphors, irrespective of participants' handedness, would be easier to interpret than graphs that do not align with these metaphors. Consistent with this hypothesis, graphs that aligned with the absolute vertical valence metaphor were more likely to elicit correct responses than graphs that did not align with this metaphor (see Table 4, rows 11-12). As shown in Fig. 4 (model 3), the 95 percent credible interval for Axis Orientation × Valence did not include zero, indicating that this effect was strongly supported by the data.

The main effects for model 3 predicted accuracy to be highest for non-inverted graphs depicting a positively valenced quantity (95 percent *CI* = [2.64, 5.68]). The second highest predicted accuracy was elicited by inverted graphs depicting a positively valenced quantity (95 percent *CI* = [1.85, 4.23]). Accuracy for inverted graphs depicting a negatively valenced quantity (95 percent *CI* = [1.11, 2.61]) and non-inverted graphs depicting a negatively valenced quantity (95 percent *CI* = [0.39, 2.15]) were similar. These main effects indicate that Valence (negative versus positive) was the primary determinant of accuracy for graphs that plotted quantity using the x-axis, rather than Axis Orientation (inverted versus non-inverted) or Valence Alignment (misaligning versus aligning).

3.2.3 Discussion

Consistent with Experiment 1 and with the results of Pandey *et al.* [1], our results show that participants were more likely to respond accurately to graphs without inverted axes than to graphs with inverted axes. Participants were also more likely to respond accurately to graphs that aligned with vertical valence metaphors than to graphs that did not align with these metaphors. For graphs that depicted quantity information using the y-axis, axis orientation mattered more for graph interpretability than alignment with vertical valence metaphors. For instance, accuracy was higher for non-inverted graphs depicting a negatively valenced quantity than for inverted graphs depicting a negatively valenced quantity, despite the fact that the latter graph aligned with vertical valence metaphors. Altogether, these results indicate that alignment with valence metaphors matters for the interpretation of line graphs, but not enough to justify inverting the axis of these graphs, which is likely to confuse graph users [1], [35].

Unlike Experiment 1, the credible interval for emotional valence (negative versus positive) in Experiment 2 included zero. However, the majority of this credible interval was above zero, with the lower bound of the interval being -0.10. Furthermore, while it was not the primary test of our hypothesis, the model assessing only horizontal valence metaphors indicated that there was strong evidence for an effect of emotional valence for graphs that depicted quantity using the x-axis (95 percent *CI* = [0.81, 3.25]). Therefore, data across both experiments could be interpreted together as showing an effect of emotional valence on response accuracy, consistent with research showing that positively valenced stimuli are recognised more quickly than negatively valenced stimuli [61], [62], [63].

As in Experiment 1, participants in Experiment 2 were more likely to respond accurately to graphs that mapped quantity information onto the x-axis rather than the y-axis (in contrast to [44]). Participants also spent more time responding to these graphs (mean = 5.11 seconds) than to graphs that depicted quantity information using the y-axis (mean = 3.79 seconds). This result lends further support to the notion that participants made fewer mistakes when responding to x-axis graphs because they spent longer verifying that their answers to these graphs were correct [77], [78].

For graphs mapping quantity onto the x-axis, participants were more likely to respond accurately to graphs that aligned with absolute valence metaphors (where 'good' is associated with rightward space) rather than handedness-relative valence metaphors (where 'good' is associated with the side of space corresponding to the dominant side of one's body). This finding supports the idea that the representation of the right-handed valence metaphor in cultural practices and language causes an absolute association between right-side space and positive valence that extends even to left-handers [50], [51].

Our results also showed that emotional valence (negative versus positive) mattered more than axis orientation for the interpretability of graphs mapping quantity onto the x-axis. In fact, accuracy was similar for x-axis graphs with non-inverted axes (78.2 percent, 147 versus 41) compared to x-axis graphs with an inverted axis (77.4 percent, 294 versus 86). A possible reason for the lack of an axis orientation effect for x-axis graphs may be that the novel presentation of quantity on the x-axis caused participants to pay more attention to the direction of this axis. The lack of an axis orientation effect may then have left more room for emotional valence to influence responses.

4 General Discussion

Our main results can be summarised as follows:

- Non-inverted line graphs are easier to interpret than line graphs with an inverted axis, confirming and extending results reported by Pandey *et al.* [1] to graphs with an inverted x-axis, and graphs with an inverted time axis (either the y-axis or x-axis).
- Line graphs that adhere to conceptual metaphors of emotional valence [19], [20], [21] are easier to interpret than line graphs that do not adhere to these conceptual metaphors. However, the beneficial effect of valence alignment on graph interpretability does not outweigh the adverse effect of reversing graphical axes.
- Line graphs depicting a positively valenced quantity are easier to interpret than line graphs depicting a negatively valenced quantity, both when quantity was mapped onto the y-axis and the x-axis, consistent with the negativity bias [61], [63], [65].
- Line graphs that plot quantity onto the x-axis are easier to interpret than line graphs that plot quantity onto the y-axis.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

Our results support the idea that conceptual metaphor theory can be relevant to the design of data visualizations [7], [8] – in this case, line graphs. However, our results lead us to recommend that line graphs should be designed to align with conceptual metaphors of emotional valence only if doing so does not subvert other well-established design graphical conventions. Our results also suggest that the inverted y-axis of the Stand Your Ground visualization, which depicted gun deaths in Florida [26], did indeed have the potential to mislead viewers of the visualization. Contrary to comments made by graphic designer Christine Chan, while axis direction *can* "be shown either way" [26], we recommend that designers hoping to elicit the correct interpretation of the data being visualized should not invert the y-axis.

There are, however, some visualizations in which y-axis inversion seems to have been employed successfully [36]. For example, the infographic entitled Iraq's Bloody Toll (see Fig. 1, B), designed by Simon Scarr, was awarded the silver prize at annual infographics conference Malofiej [17], [78]. This infographic depicts the number of deaths per month in Iraq between 2003 and 2011, inverting its y-axis and filling its bars in red to resemble blood running down the page. In contrast to the Stand Your Ground visualization, the Iraq infographic's title and its use of thin red bars with rounded end caps emphasise the metaphor of blood. In addition, the use of bars in the Iraq infographic allows the viewer to see that the bars originate at the top of the page, whereas the Stand Your Ground graph is a line chart and so does not provide an easy cue for determining axis orientation [36]. More research is needed, but y-axis inversion may be acceptable if the axis inversion is made highly salient by other design features of the visualization. However, doing so necessarily involves risk as some viewers might not notice the axis inversion.

In some cases, graph designers may prefer to avoid plotting quantity on the y-axis entirely if using the y-axis for quantity conflicts with metaphors of emotional valence, and instead may prefer to use the x-axis. In support of this recommendation, Tversky [2] has argued that the horizontal axis is more neutral than the vertical one. Our results corroborate the relative neutrality of the horizontal axis with respect to line graphs: while we found an effect of valence alignment for the x-axis in Experiment 2, this effect was not as strong as the effect of valence alignment for the vertical axis in both Experiment 1 and 2 (as indicated by the distance of the 95 percent credible intervals from zero, see Figs. 3 and 4).

Our results have potential applications to other graphs where designers select the ordering of categorical variables, such as stacked bar charts. When there is no strict convention for the ordering of y-axis categories, these categories could be listed from smallest to largest from the bottom to the top of the y-axis, or from the top to the bottom, presumably without causing confusion. The interpretability of these graphs might be enhanced when the category with the 'most' is positioned at the top of the graph, and the category with the 'least' is positioned at the bottom, in line with vertical quantity metaphors [3], [4], [6]. Alternatively, the optimal ordering might depend on whether the quantity is positively or negatively valenced [20], [21], [38]. This is an intriguing avenue of research for future studies to explore.

In general, more research needs to be conducted looking at graphs showing different trends and quantities to determine the extent to which our results can be generalised. For example, the graphs we showed to participants were based on those used by Pandey *et al*. [1] and thus were highly schematic, showing a simple, relatively linear line without any other fluctuations. Real-life data are often more noisy and complex than in the current work. Future research is needed to determine whether our results generalise to different types of data, specifically data with more variability. It also remains to be seen how non-linearities in the trend line might affect conceptualisations of emotional valence.

In addition to the vertical axis, our data revealed an effect of absolute horizontal metaphors of emotional valence that was not relative to handedness. Interpretation of the line graphs in this study thus did not appear to be influenced by body-specificity [47], [48], [49]. This finding supports the notion that even left-handers may come to acquire an absolute association between right-side space and positive valence due to the representation of the right-handed valence metaphor in cultural practices and language [50], [51].

Based on the support our results provided for the negativity bias, an additional recommendation we can make for the design of line graphs is that, if it is possible to frame a quantity in positively or negatively valenced terms (e.g., '10 percent less sad' versus '10 percent happier', or '30 percent uncertain' versus '70 percent certain'), the visualization designer should use the positively valenced framing. For example, it may be easier to interpret the graph shown in Fig. 1 (C) if this graph displayed ratings of favourability rather than unfavourability.

A possible criticism of our assessment of emotional valence is that participants may not have universally agreed that 'vacation days' is a positively valenced concept – the judgment of a stimulus as positively or negatively valenced is subjective. For example, employers might view 'vacation days' as negatively valenced due to its association with diminished workplace productivity. Nonetheless, the fact that 'vacation' was rated as the most positively valenced word in the 13,915 words normed by Warriner *et al.* [67] leads us to assume that the overwhelming majority of our participants assessed this concept favourably. Because Warriner *et al.* used a sample of US participants recruited via Amazon Mechanical Turk, the majority of whom had English as their native language, similar to our sample, there is good reason to believe that our participants shared similar views about the valence of 'vacation days'. Furthermore, the standard deviation of the ratings for 'vacation' (0.77) was substantially below the average standard deviation for words in the Warriner *et al.* dataset as a whole (1.68) (more extreme ratings tend to be less variable, see [79]), indicating that attitudes toward 'vacation' were highly consistent. The large samples in both our experiments ensured that anomalous data from participants who felt differently did not inordinately skew our results. Finally, because participants answered questions about both vacation days and murders, it seems likely that vacation days would have been perceived as more positively valenced than murders in relative terms. This contrast in itself might be expected to affect results in the direction we observed.

Our finding that participants were more likely to respond accurately to graphs that plotted quantity using the x-axis

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

was contrary to our predictions. We speculated that the subversion of convention in these graphs was likely to have been salient to participants, who then spent longer interpreting these graphs to make sure their response was correct. This finding differs from the y-axis advantage for bar charts reported by Fischer *et al.* [44]. The discrepancy between these two sets of results may be attributable to the fact that horizontal bar charts are relatively 'normal', conventional visualizations, unlike line graphs that map quantity onto the x-axis, which are unusual. Our results may also speak to the notion of *desirable difficulties*, which posits that making learning more challenging can be beneficial [77], [78].

This paper demonstrates the role that the field of conceptual metaphor research may have to play in the design of data visualizations. Our results suggest that conceptual metaphor theory matters for the interpretation of line graphs. However, visualization designers are warned against subverting graphical convention in order to align with conceptual metaphors that are not conventionally instantiated graphically. Otherwise, the resultant visualization may be more liable to mislead than to inform.

Acknowledgments

The work of Greg Woodin was supported by the Economic and Social Research Council under Grant ES/P000711/1. The work of Bodo Winter was supported by the UKRI Future Leaders Fellowship under Grant MR/T040505/1.

References

[1] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini, "How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 1469–1478.

[2] B. Tversky, "Visualizing thought," *Top. Cogn. Sci.*, vol. 3, no. 3, pp. 499–535, 2011.

[3] M. Hartmann, V. Gashaj, A. Stahnke, and F. W. Mast, "There is more than "more is up": Hand and foot responses reverse the vertical association of number magnitudes," *J. Exp. Psychol. Hum. Perception Perform.*, vol. 40, no. 4, pp. 1401–1414, 2014.

[4] B. Winter and T. Matlock, "More is up. . . and right: Random number generation along two axes," in *Proc. Annu. Meeting Cogn. Sci. Soc.*, 2013, pp. 3789–3974.

[5] T. Loetscher, C. J. Bockisch, M. E. R. Nicholls, and P. Brugger, "Eye position predicts what number you have in mind," *Curr. Biol.*, vol. 20, no. 6, pp. 264–265, 2010.

[6] M. Hartmann, L. Grabherr, and F. W. Mast, "Moving along the mental number line: Interactions between whole-body motion and numerical cognition," *J. Exp. Psychol. Hum. Perception Perform.*, vol. 38, no. 6, pp. 1416–1427, 2012.

[7] P. Parsons, "Conceptual metaphor theory as a foundation for communicative visualization design," in *Proc. IEEE VIS Workshop Vis. Commun.*, 2018, pp. 1–6. [Online]. Available: https://par.nsf.gov/biblio/10087033

[8] J. S. Risch, "On the role of metaphor in information visualization," 2008, *arXiv:0809.0884*

[9] R. W. Gibbs, Jr., *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge, UK: Cambridge Univ. Press, 1994.

[10] Z. Kövecses, *Metaphor: A Practical Introduction*. Oxford, U.K.: Oxford Univ. Press, 2002.

[11] G. Lakoff and M. Johnson, *Metaphors We Live By*. Chicago, IL, USA: Chicago Univ. Press, 1980.

[12] T. Loetscher, C. J. Bockisch, and P. Brugger, "Looking for the answer: The mind's eye in number space," *Neuroscience*, vol. 151, no. 3, pp. 725–729, 2008.

[13] L. Boroditsky, "Metaphoric structuring: Understanding time through spatial metaphors," *Cognition*, vol. 75, no. 1, pp. 1–28, 2000.

[14] L. Boroditsky and M. Ramscar, "The roles of body and mind in abstract thought," *Psychol. Sci.*, vol. 13, no. 2, pp. 185–189, 2002.

[15] M. S. McGlone and J. L. Harding, "Back (or forward?) to the future: The role of perspective in temporal language comprehension," *J. Exp. Psychol. Learn., Memory, Cogn.*, vol. 24, no. 5, pp. 1211–1223, 1998.

[16] Reuters, 2014. [Online]. Available: https://www.reuters.com/news/world

[17] S. Scarr, "Iraq's bloody toll," 2011. [Online]. Available: https://www.scmp.com/infographics/article/1284683/iraqs-bloody-toll

[18] H. Enten, "Americans' distaste for both Trump and Clinton is record-breaking," 2016. [Online]. Available: https://fivethirtyeight.com/features/americans-distaste-for-both-trump-and-clinton-is-record-breaking/

[19] B. P. Meier and M. D. Robinson, "Does "feeling down" mean seeing down? Depressive symptoms and vertical selective attention," *J. Res. Pers.*, vol. 40, no. 4, pp. 451–461, 2006.

[20] B. P. Meier and M. D. Robinson, "Why the sunny side is up: Association between affect and vertical position," *Psychol. Sci.*, vol. 15, no. 4, pp. 243–247, 2004.

[21] D. Casasanto and K. Dijkstra, "Motor action and emotional memory," *Cognition*, vol. 115, no. 1, pp. 179–185, 2010.

[22] L. F. Barrett and J. A. Russell, "The structure of current affect: Controversies and emerging consensus," *Curr. Directions Psychol. Sci.*, vol. 8, no. 1, pp. 10–14, 2016.

[23] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Motivated attention: Affect, activation and action," in *Proc. Attention Orienting Sensory Motivational Processes*, 1997, pp. 97–135.

[24] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychol. Rev.*, vol. 110, no. 1, pp. 145–172, 2003.

[25] P. Engel, "Gun deaths in Florida increased with 'Stand Your Ground'," 2014. [Online]. Available: https://www.businessinsider.com/gun-deaths-in-florida-increased-with-stand-your-ground-2014–2?r=US&IR=T

[26] M. Lallanilla, "Misleading gun-death chart draws fire," 2014. [Online]. Available: https://www.livescience.com/45083-misleading-gun-death-chart.html

[27] J. deVilla, "Lies, damned lies, and Reuters' Graph of Florida firearm deaths before and after "stand your ground"," 2014. [Online]. Available: http://www.joeydevilla.com/2014/04/14/lies-damned-lies-and-reuters-graph-of-florida-firearm-deaths-before-and-after-stand-your-ground

[28] L. Wade, "How to lie with statistics: Stand your ground and gun deaths," 2014. [Online]. Available: https://thesocietypages.org/socimages/2014/12/28/how-to-lie-with-statistics-stand-your-ground-and-gun-deaths/

[29] L. M. Padilla, "A case for cognitive models in visualization research," in *Proc. IEEE Eval. Beyond Methodol. Approaches Vis.*, 2018, pp. 69–77.

[30] L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci, "Decision making with visualizations: A cognitive framework across disciplines," *Cogn. Res. Princ. Implic.*, vol. 3, no. 29, pp. 1–25, 2018.

[31] M. Ishihara, P. E. Keller, Y. Rossetti, and W. Prinz, "Horizontal spatial representations of time: Evidence for the STEARC effect," *Cortex*, vol. 44, no. 4, pp. 454–61, 2008.

[32] D. J. Cox, "The art and science of visualization: Metaphorical maps and cultural models," *Technoetic Arts*, vol. 2, no. 2, pp. 71–80, 2004.

[33] P. A. Fishwick, "Metaphoric mappings: The art of visualization," in *Proc. Aesthetic Comput.*, 2006, pp. 89–114.

[34] I. Vessey, "Cognitive fit: A theory-based analysis of the graphs versus tables literature," *Decis. Sci.*, vol. 22, no. 2, pp. 219–240, 1991.

[35] M. Correll and J. Heer, "Black hat visualization," in *Proc. IEEE DECISIVe Workshop Dealing Cogn. Biases Vis.*, 2017, pp. 1–4. [Online]. Available: https://decisive-workshop.dbvis.de/index.html%3Fp=555.html#115

[36] R. Kosara, "When bars point down," 2014. [Online]. Available: https://eagereyes.org/journalism/when-bars-point-down

[37] M. Hartmann, F. W. Mast, and M. H. Fischer, "Counting is a spatial process: Evidence from eye movements," *Psychol. Res.*, vol. 80, no. 3, pp. 399–409, 2016.

[38] T. Seno, T. Kawabe, H. Ito, and S. Sunaga, "Vection modulates emotional valence of autobiographical episodic memories," *Cognition*, vol. 126, no. 1, pp. 115–120, 2013.

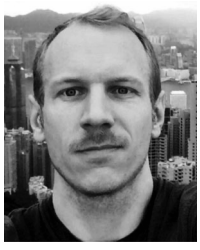
[39] B. Winter and T. Matlock, "Primary metaphors are both cultural and embodied," in *Proc. Metaphor Embodied Cogn. Discourse*, 2017, pp. 99–116.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

[40] A. Chinello, M. D. de Hevia, C. Geraci, and L. Girelli, "Finding the spatial-numerical association of response codes (SNARC) in signed numbers: Notational effects in accessing number representation," *Funct. Neurol.*, vol. 27, no. 3, pp. 177–185, 2013.
[41] S. Dehaene, S. Bossini, and P. Giraux, "The mental representation of parity and number magnitude," *J. Exp. Psychol. Gen.*, vol. 122, no. 3, pp. 371–396, 1993.
[42] G. Wood, K. Willmes, H.-C. Nuerk, and M. H. Fischer, "On the cognitive link between space and number: A meta-analysis of the SNARC effect," *Psychol. Sci.*, vol. 50, no. 4, pp. 489–525, 2008.
[43] G. Woodin and B. Winter, " Placing abstract concepts in space: Quantity, time and emotional valence," *Front. Psychol.*, vol. 9, pp. 1–14, 2018.
[44] M. H. Fischer, N. Dewulf, and R. L. Hill, "Designing bar graphs: Orientation matters," *Appl. Cogn. Psychol.*, vol. 19, no. 7, pp. 953–962, 2005.
[45] B. Winter, T. Matlock, S. Shaki, and M. H. Fischer, "Mental number space in three dimensions," *Neurosci. Biobehav. Rev.*, vol. 57, pp. 209–219, 2015.
[46] B. Winter, T. Marghetis, and T. Matlock, "Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number," *Cortex*, vol. 64, pp. 209–224, 2015.
[47] D. Casasanto, "Embodiment of abstract concepts: Good and bad in right- and left-handers," *J. Exp. Psychol. Gen.*, vol. 138, no. 3, pp. 351–367, 2009.
[48] D. Casasanto and T. Henetz, "Handedness shapes children's abstract concepts," *Cogn. Sci.*, vol. 36, no. 2, pp. 359–372, 2012.
[49] J. W. Van Strien and S. Van Beek, "Ratings of emotion in laterally presented faces: Sex and handedness effects," *Brain Cogn.*, vol. 44, no. 3, pp. 645–652, 2000.
[50] D. Casasanto, "The hierarchical structure of mental metaphors," in *Proc. Metaphor Embodied Cogn. Discourse*, 2017, pp. 46–61.
[51] D. Casasanto and R. Bottini, "Spatial language and abstract concepts," *Wiley Interdiscipl. Rev. Cogn. Sci.*, vol. 5, no. 2, pp. 139–149, 2014.
[52] S. Y. Kim, J. W. Lee, and I. S. Sohn, "Comparison of various statistical methods for identifying differential gene expression in replicated microarray data," *Stat. Methods Med. Res.*, vol. 15, no. 1, pp. 3–20, 2016.
[53] A. Abdul-Rahman *et al.*, "Re-spatialization of time series plots," *Inf. Vis.*, vol. 16, no. 4, pp. 275–290, 2017.
[54] S. Havre, E. Hetzler, P. Whitney, and L. Nowell, "ThemeRiver: Visualizing thematic changes in large document collections," *IEEE Trans. Vis. Comput. Graph.*, vol. 8, no. 1, pp. 9–20, Jan.–Mar. 2002.
[55] M. Wattenberg, "Baby names, visualization, and social data analysis," in *Proc. IEEE Symp. Inf. Vis.*, 2005, pp. 1–7.
[56] A. Vallesi, M. A. Binns, and T. Shallice, "An effect of spatial-temporal association of response codes: Understanding the cognitive representations of time," *Cognition*, vol. 107, no. 2, pp. 501–527, 2008.
[57] M. Conson, F. Cinque, A. M. Barbarulo, and L. Trojano, "A common processing system for duration, order and spatial information: Evidence from a time estimation task," *Exp. Brain Res.*, vol. 187, no. 2, pp. 267–274, 2008.
[58] S. Ruiz Fernandéz, M. Lachmair, and J. J. Rahona, "Human mental representation of time in the vertical space," in *Proc. 6th Int. Congr. Med. Space Extreme Environ.*, 2014.
[59] M. J. Leone, A. Salles, A. Pulver, D. A. Golombek, and M. Sigman, "Time drawings: Spatial representation of temporal concepts," *Consciousness Cogn.*, vol. 59, pp. 10–25, 2018.
[60] M. H. Fischer and H. Campens, "Pointing to numbers and grasping magnitudes," *Exp. Brain Res.*, vol. 192, no. 1, pp. 149–153, 2009.
[61] M. W. Matlin and D. J. Stang, *The Pollyanna Principle: Selectivity in Language, Memory, and Thought*. Cambridge, MA, USA: Schenkman Pub. Co., 1978.
[62] G. Peeters, "Evaluative positive-negative asymmetry in adjective-noun compounds," *Polish Psychol. Bull.*, vol. 20, no. 4, pp. 255–266, 1989.
[63] G. Peeters, "The positive-negative asymmetry: On cognitive consistency and positivity bias," *Eur. J. Social Psychol.*, vol. 1, pp. 455–474, 1971.
[64] G. Peeters and J. Czapinski, "Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects," *Eur. Rev. Social Psychol.*, vol. 1, no. 1, pp. 33–60, 1990.
[65] A. Vaish, T. Grossmann, and A. Woodward, "Not all emotions are created equal: The negativity bias in social-emotional development," *Psychol. Bull.*, vol. 134, no. 3, pp. 383–403, 2008.
[66] Qualtrics, "Four core experiences of business: Four applications to get them right," 2019. [Online]. Available: https://www.qualtrics.com/
[67] Amazon Mechanical Turk, 2021. [Online]. Available: https://www.mturk.com/
[68] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behav. Res. Methods*, vol. 45, no. 4, pp. 1191–1207, 2013.
[69] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org/
[70] RStudio Team, "RStudio: Integrated Development Environment for R," RStudio, Inc., Boston, MA, USA: 2016. [Online]. Available: http://www.rstudio.com/
[71] H. Wickham, "The split-apply-combine strategy for data analysis," *J. Stat. Softw.*, vol. 40, no. 1, pp. 1–29, 2011.
[72] H. Wickham *et al.*, "Welcome to the tidyverse," *J. Open Source Softw.*, vol. 4, no. 43, 2019, Art. no. 1686.
[73] X. F. i Marín, "GGMCMC: Analysis of MCMC samples and Bayesian inference," *J. Stat. Softw.*, vol. 70, no. 9, pp. 1–20, 2016.
[74] M. Kay, "Tidybayes: Tidy data and Geoms for Bayesian models," R package version 2.3.1, 2020. [Online]. Available: http://mjskay.github.io/tidybayes/
[75] P.-C. Bürkner, "brms: An R package for Bayesian multilevel models using Stan," *J. Stat. Softw.*, vol. 80, no. 1, pp. 1–28, 2017.
[76] P.-C. Bürkner, "Advanced Bayesian multilevel modeling with the R package BRMS," *R J.*, vol. 10, no. 1, pp. 395–411, 2018.
[77] R. A. Bjork, "Memory and metamemory considerations in the training of human beings," in *Metacognition: Knowing About Knowing*. Cambridge, MA, USA: MIT Press, 1994, pp. 185–205.
[78] J. Hullman, E. Adar, and P. Shah, "Benefitting InfoVis with visual difficulties," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2213–2222, Dec. 2011.
[79] A. Cotgreave, "Lies, damn lies, and statistics: How to take something positive from the UK's EU referendum campaign," 2016. [Online]. Available: https://www.infoworld.com/article/3088166/why-how-to-lie-with-statistics-did-us-a-disservice.html
[80] L. Pollock, "Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study," *Behav. Res. Methods*, vol. 50, no. 3, pp. 1198–1216, 2018.



Greg Woodin (Member, IEEE) is currently working toward the PhD degree in English Language and Applied Linguistics with the University of Birmingham, U.K. His research interests include metaphor, iconicity, gesture, spatial cognition, and data visualization.



Bodo Winter is currently a senior lecturer in cognitive linguistics with the University of Birmingham, U.K. His research interests include metaphor, gesture, language evolution, and statistical methodology. He is a UKRI Future Leaders Fellow.



Lace Padilla (Member, IEEE) is currently an assistant professor with the Department of Cognitive and Information Sciences, UC Merced, USA. Her research interests include visual decision-making and spatial cognition.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on September 17,2022 at 07:50:17 UTC from IEEE Xplore. Restrictions apply.

5. PAPER 3

Woodin, G., Winter, B., Littlemore, J., Perlman, M. & Grieve, J. (under review). Large-scale patterns of number use in spoken and written English.

- **Conceptualization:** Greg Woodin, Bodo Winter, Jeannette Littlemore, Marcus Perlman, Jack Grieve.
- **Data curation:** Greg Woodin.
- **Formal analysis:** Greg Woodin.
- **Investigation:** Greg Woodin, Bodo Winter, Jeannette Littlemore, Marcus Perlman, Jack Grieve.
- **Methodology:** Greg Woodin, Bodo Winter, Jeannette Littlemore, Marcus Perlman, Jack Grieve.
- **Project administration:** Greg Woodin, Bodo Winter, Jeannette Littlemore, Marcus Perlman, Jack Grieve.
- **Supervision:** Bodo Winter, Jeannette Littlemore, Marcus Perlman, Jack Grieve.
- **Validation:** Bodo Winter, Jeannette Littlemore, Marcus Perlman, Jack Grieve.
- **Visualization:** Greg Woodin.
- **Writing – original draft:** Greg Woodin.
- **Writing – review & editing:** Greg Woodin, Bodo Winter, Jeannette Littlemore, Marcus Perlman, Jack Grieve.

Supplementary materials: https://osf.io/ze9vk/

Large-scale patterns of number use in spoken and written English

Greg Woodin, Bodo Winter, Jeannette Littlemore, Marcus Perlman, and Jack Grieve

Greg Woodin is an ESRC-funded PhD student in English Language and Applied Linguistics in the Department of English Language and Linguistics at the University of Birmingham.

Bodo Winter is an Associate Professor in Cognitive Linguistics in the Department of English Language and Linguistics at the University of Birmingham and a UKRI Future Leaders Fellow.

Jeannette Littlemore is a Professor of Applied Linguistics in the Department of English Language and Linguistics at the University of Birmingham.

Marcus Perlman is a Lecturer in English Language and Linguistics in the Department of English Language and Linguistics at the University of Birmingham.

Jack Grieve is a Professor of Corpus Linguistics in the Department of English Language and Linguistics at the University of Birmingham.

Abstract

This paper describes patterns of number use in spoken and written English and the main factors that contribute to these patterns. We analysed more than 1.7 million occurrences of numbers between 0 and a billion in the British National Corpus, including conversational speech, presentational speech (e.g., lectures, interviews), imaginative writing (e.g., fiction), and informative writing (e.g., academic books). We find that four main factors affect number frequency: (1) Magnitude – smaller numbers are more frequent than larger numbers; (2) Roundness – round numbers are more frequent than unround numbers of a comparable magnitude, and some round numbers are more frequent than others; (3) Cultural salience – culturally salient numbers (e.g., recent years) are more frequent than non-salient numbers; and (4) Register – more informational texts contain more numbers (in writing), types of numbers, decimals, and larger numbers than less informational texts. In writing, we find that the numbers 1–9 are mostly represented by number words (e.g., 'three'), 10–999,999 are mostly represented by numerals (e.g., '14'), and 1 million–1 billion are mostly represented by a mix of numerals and number words (e.g., '8 million'). Altogether, this study builds a detailed profile of number use in spoken and written English.

Keywords: number frequencies, numerical cognition**,** rounding, big data, register studies

1. Introduction

People use some numbers more often than others. For example, English speakers tend to refer to numbers like 3, 100, and 2022 more often than to numbers like 43, 104, and 3022. More generally, smaller numbers may be used more often than larger numbers, round numbers (e.g., multiples of ten) more often than unround numbers, and culturally salient numbers (e.g., recent years) more often than non-salient numbers. People may also use numbers more often when communicating precise information, or in situations where presenting facts in an informative manner is important. Exploring the frequency with which different numbers are used can reveal the numbers people deem important to use to communicate about which quantities, in what contexts, and why. The present study investigates patterns of number use in spoken and written English and explores the primary factors that contribute to these patterns. Using the 100 million word British National Corpus (BNC Consortium, 2007), we identified over 1.7 million occurrences of numbers from 0 to a billion and analysed their frequency in relation to four factors: magnitude, roundness, cultural salience, and register. In written texts, we explored the format in which numbers of different magnitudes are expressed, i.e., as numerals (e.g., '140'), number words (e.g., 'three'), or a mix of both (e.g., '1 million'). Overall, this paper reports the largest, most comprehensive corpus analysis of number use to date.

2. Background

Several factors are believed to influence the frequency with which people use different numbers. One is *magnitude*: studies have shown that people use smaller numbers more frequently than larger numbers (Coupland, 2011; Dehaene & Mehler, 1992; Dorogovtsev et al., 2005; Jansen & Pollmann, 2001). A possible explanation for why people discuss smaller numbers more frequently is that smaller quantities are encountered more often, and are more easily countable, making them more relevant to discuss (Cummins, 2015, p. 32). For example,

we are more likely to encounter a group of three people than a group of 27, and even if we do encounter the larger group, the exact number may be unknown, or irrelevant to the point that a simple non-numerical description such as “a large group” or “several people” may suffice. The bias toward using small numbers may create a feedback loop in in which smaller numbers are encountered more often and learned earlier, which may cause them to be used more frequently still (Rath, 1966). Dehaene and Mehler (1992) also consider the possibility that people discuss small numbers more often because these numbers are easier to mentally process, citing magnitude comparison tasks in which smaller numbers are identified more quickly than larger ones (Buckley & Gillman, 1974; Dehaene, 1989). Another aspect to consider is that we conventionally use scales to keep numbers small (Coupland, 2011, pp. 34–35) – for example, we can avoid having to talk about 600 seconds by using a larger temporal unit, like ten minutes.

The second factor that influences number frequency is *roundness:* research has shown that round numbers are used more frequently than unround numbers of a similar magnitude (Coupland, 2011; Dehaene & Mehler, 1992; Dorogovtsev et al., 2005; Jansen & Pollmann, 2001). In decimal number systems — those based on the number ten (including tenths and powers of ten), such as the English system — round numbers are typically considered to be multiples of ten and sometimes five (e.g., Dorogovtsev et al., 2005; Sigurd, 1988). One possible reason why round numbers are used more frequently is that they can be used approximately in a practice known as *rounding*, where the nearest round number is used in place of the real value (Cummins, 2015, p. 20; Jansen & Pollmann, 2001; Krifka, 2009; Sigurd, 1988). For example, instead of saying that a lecture has 99 attendees, a speaker may round 99 up to 100. Hence, round numbers can represent a range of values – they have a larger *pragmatic halo* (see Lasersohn, 1999) than unround numbers, which usually only represent a single precise value. For example, it is generally acceptable to use ‘100’ to denote the number 103, while it is generally unacceptable to use ‘103’ to denote the number 100. The

increased flexibility in the use of round numbers may help explain why they have been found to be used more frequently than unround numbers. In other contexts, the nearest round number can be used as a benchmark indicating some level of completion (Dehaene, 1997, Ch. 4; Gunasti, 2016). For instance, if there are 99 people at a lecture, someone may write that there are “nearly 100 people”, whereas if there are 103 people, they may instead write that there are “over 100 people”.

The reason that multiples of ten and five are treated as round numbers may be that they are psychologically salient (Van der Henst & Sperber, 2004), making these numbers more cognitively accessible, simpler, and less cognitively costly compared to other numbers (Dehaene, 1997; Krifka, 2009; see Cummins, 2015, p. 32). The psychological salience of these numbers may be due to their structural prominence within the decimal system, which may have deeper roots in finger counting practices (Butterworth, 1999; Dehaene, 1997; see Cummins, 2015, p. 33): the number of fingers on each hand may be the basis of decimal systems (e.g., Bender & Beller, 2012; Ifrah, 1987; Wiese, 2004). Due to this special status, cultural artifacts and customs are often based on these values. For example, packs of ten items are more common than packs of 11 items (Cummins, 2015, p. 36), and we celebrate jubilees on the 5th, 10th, 25th, and 50th (etc.) anniversaries, rather than the 6th, 12th, 18th, and so on (Jansen & Pollmann, 2001). These cultural artifacts and customs may create a feedback loop in which multiples of ten and five are encountered even more often (Chrisomalis, 2020, p. 35), further enhancing the psychological salience of these numbers (Cummins, 2015, p. 36).[1]

[1] The idea that roundness is determined by the base of the mathematical system means that the numbers considered round differ between different systems. We can see evidence of this in sexagesimal systems (base 60) used, for example, in the measurement of time. In this time-telling context, 15 and 30 may be round numbers because they are, respectively, a quarter and half of 60 seconds or minutes (Dehaene & Mehler, 1992; Jansen & Pollmann, 2001). Similarly, 90, 180, and 360

People use round numbers in an approximate way for several different reasons. First, the precise number is often deemed unnecessary to mention. For example, while it might be helpful to have a ballpark figure of the number of people that will be attending your lecture, the precise number may be practically irrelevant. By using round numbers in such contexts, language users conform to Grice's (1975) maxims, being only as precise as is necessary to achieve the aims of the interaction (Gibbs Jr. & Bryant, 2008; Van der Henst et al., 2002). Unnecessary precision may also be seen as pedantic (Lin, 2013; McCarthy, 2006, p. 22; see also Cotterill, 2007; Beltrama et al., 2022), exerting additional social pressure toward rounding in certain contexts. Second, people often use round numbers imprecisely when the exact number is unknown, such as when estimating the size of a crowd (e.g., 'a thousand people'), using a round number to reduce commitment to a precise number (Ruud et al., 2014). This strategy works because people usually interpret round numbers approximately, unless they are modified by words such as 'exactly' (Krifka, 2009; Lasersohn, 1999). Finally, another reason that can motivate the use of round numbers is strategic manipulation, such as when stating that a university is a 'top ten university' (a round number) as opposed to a 'top six university' (an unround number in context), because the latter expression yields the disadvantageous implicature that the university is exactly 6th place (Cummins & Franke, 2021).

Corpus studies have shed light on the two roles of magnitude and roundness in numerical communication. Dehaene and Mehler (1992) investigated number words (e.g., 'three', 'sixteen', 'eighty', 'thousand') from 0 to a billion in a corpus of over 1 million American English words (Francis & Kučera, 1982). The results showed that number frequencies decline with numerical magnitude, but that round numbers 10, 20, 50, and 100 are used more frequently than other numbers of a similar magnitude. Jansen and Pollman (2001) reported similar results for numerals and number words in the range 2–1000 in a 40 million word corpus

---

may be round numbers in the measurement of angles. However, in this paper we focus on round numbers in decimal systems, the predominant system in Western cultures.

of 1994 volumes of English newspaper *The Times.* Coupland (2011) uncovered a similar pattern for number words from 1 to 20 in the 100 million word British National Corpus (BNC Consortium, 2007), and for both numerals and number words from 1 to 100 on the internet. Lastly, though not framed explicitly as a corpus study, Dorogovtsev et al. (2005) found that frequencies of web pages containing positive numerals declined with numerical magnitude, but that frequencies were relatively higher for pages containing powers of ten (i.e., round numbers in the decimal system).

Some authors have argued that roundness is a matter of degree, rather than being absolute. For example, Sigurd (1988) proposed a model of roundness where powers of ten (e.g., 10, 100, 1000) and halves (e.g., 5, 50, 500) and quarters (e.g., 2.5, 25, 250) of these powers are 'rounder' than other numbers. Jansen and Pollman (2001) developed Sigurd's model, showing empirically that numbers are used more frequently if they have one of the properties of 10-ness (10, 20, 30, ... 100, 200, 300, ...), 2-ness (20, 40, 60, ... 200, 400, 600, ...), 2.5-ness (25, 50, 75, ... 250, 500, 750, ...), and 5-ness (50, 100, 150, ... 500, 1000, ...). Notably, 10-ness, 2-ness, 2.5-ness, and 5-ness do not simply refer to numbers that are divisible by these respective factors; rather, they refer to numbers that equal an integer no greater than 9 when divided by 1, 2, 2.5, or 5 multiplied by a power of 10. For example, 200 has 10-ness because it equals 2 when divided by 100 (1×10^2), and 750 has 2.5-ness because it equals 3 when divided by 250 (2.5×10^2). Jansen and Pollman (2001) argue that these numbers are used more often due to humans having a "natural propensity" (p. 201) for doubling and halving the base of the mathematical system (e.g., powers of ten in a decimal system), and, in the case of 2.5-ness, halving again.

Two questions regarding roundness remain unanswered. First, it is not clear if multiples of ten and five *without* 10-ness, 2-ness, 2.5-ness, or 5-ness (e.g., 1010, 70,515, 944,500) are used more often than non-multiples, or if it is only those multiples of ten and five with 10-ness, 2-ness, 2.5-ness, or 5-ness (e.g., 70, 250, 600) that are used more often.

Second, Jansen and Pollman (2001) report that numbers with more roundness properties (e.g., 40 has 10-ness, 2-ness, and 5-ness; 300 has 10-ness and 5-ness) are used more frequently than numbers with fewer. However, as Cummins (2015, p. 34) points out, we do not know whether certain of these properties are more important predictors of number frequency than others. For instance, 10-ness may be a more important predictor of number frequency than 2-ness, or vice versa. Table 1 shows examples of numbers with each of the roundness properties discussed in this paper.

TAB. 1

The two factors of magnitude and roundness may interact: people may round larger numbers to a greater extent than smaller numbers (see Coupland, 2011). If so, this aspect of numerical communication may reflect numerical cognition, which becomes less precise for larger quantities. We can precisely quantify sets of three or four objects or fewer, which is known as *subitization* (Butterworth et al., 2008; Gordon, 2004; Pica et al., 2004). However, above this limit, this quantification ability is approximate; we cannot, for example, quantify a crowd at a glance (Cheyette & Piantadosi, 2019; McCrink & Wynn, 2007; Xu & Spelke, 2000). This quantification ability is increasingly imprecise for larger sets, which is described by the Weber-Fechner law: the just-noticeable difference between quantities is linearly related to their ratio. According to this law, sets of 110 and 120 (ratio = 11:12) are harder to discriminate than sets of 10 and 20 (ratio = 1:2), despite each pair differing by an equal number (10) (DeWind et al., 2015; Shepard et al., 1975). If we find it difficult to discriminate larger sets, we may be unlikely to communicate the differences between these sets precisely, leading to a more approximate use of round numbers at higher magnitudes.

In addition to magnitude and roundness, a third factor that influences the frequency with which numbers are used is *cultural salience*. For example, studies have reported that

recent years (e.g., 2021) are discussed more often than less recent years (e.g., 1788), probably because more recent years tend to be more relevant to present discussion (Pollmann, 1998; Pollmann & Baayen, 2001). It is also possible that numbers referring to significant dates and periods in human history, such as 1066 (the Battle of Hastings), 2019 and 2020 (the beginning of the Covid-19 pandemic), and 1939–1945 (World War II) are used more frequently today than other numbers. Moreover, Coupland (2011) finds that numbers with repeated numerals, like 99, 11 and 44, are often used in product names, like the Xbox 360 game 'Ninety-Nine Nights', perhaps for their phonological alliteration, aesthetic appeal, and 'coolness'. Other numbers are imbued with numerological significance and so may be expected to be used more or less often on this basis. For example, in Christian cultures, the number seven is associated with perfection and the Christian God (and is also the number of days in a week), while 666 is associated with the Christian devil (Ayonrinde et al., 2021). Dehaene and Mehler (1992) also report that the number 13 is used less frequently than 12 or 14, perhaps because 13 is deemed unlucky in Western cultures, to the extent that floor numbering systems in many buildings skip the number 13 (Pokryshevskaya & Antipov, 2015).

Finally, a fourth factor that may affect number frequencies is *register.* Registers are varieties of language use that are linked to a communicative context or goal (e.g., Trudgill, 1983, p. 101). Whereas the other three factors influence the frequency of certain numbers, this factor relates to how number use differs in different registers. One dimension of register variation is modality – speech versus writing – and studies have revealed differences in language use across more specific spoken and written registers, such as conversational speech (Biber et al., 1999, Ch. 13; Biber, 2009b; Conrad & Biber, 2009) and informational writing (Biber, 2009a, 2009b; Biber et al., 1999; Biber & Clark, 2002; Biber & Gray, 2010; Conrad & Biber, 2009). Neumann (2014) also reports on linguistic differences between instruction manuals, novels, letters to shareholders, and other registers, while Egbert and Mahlberg (2020) show differences between narration and speech in novels, and Love et al.

(2019) find differences in language use (e.g., turn length, word length, modal verbs) across different activity types and locations (e.g., a family playing a board game at home vs colleagues discussing a project at work). There is reason to believe that number use may also vary across registers according to whether the primary communicative goal is to inform. For example, Zillman and Brosius (2000, p. 42) report that 44% of the news articles they investigated included numbers, like percentages, amounts, and proportions. Koetsenruijter (2011) finds that the use of numbers in news texts enhances the perceived credibility of the information being communicated. Porter (1995) also argues that statistics are used in academic and professional practice due to their perceived objectivity (see Barchas-Lichtenstein et al., 2022 for a similar argument pertaining to journalism). Moreover, quantification may be necessary to convey precise numerical information; for example, in factual reports or financial transactions (Coupland, 2011). For these reasons, there may be more of an emphasis on numbers in more informational registers, such as news broadcasts, academic writing, and journalism, compared with less informational registers, like casual conversation and fiction writing.

Investigating number use in written registers raises the additional question of how these numbers are expressed – as numerals (e.g., '1,000,000'), number words (e.g., 'one million'), or a mix of both (e.g., '1 million'). Throughout history, there have always been multiple ways of writing a number in all literary traditions (Chrisomalis 2020, Ch. 6). Many style guides, like the Office for National Statistics' (2022), prescribe number words for numbers 0–9, numerals for numbers 10–999,999, and a mix of numerals and number words for numbers above this range. We might wonder, then, whether the representational formats chosen by authors reflect these style guides, and hence whether numbers of different magnitudes tend to be represented in different formats. On the internet, Coupland (2011) finds that numerals outnumber number words by a ratio of about 4:1, and that this ratio increases rapidly to approximately 200:1 by the number 99. However, the number words explored in this study

were exclusive to English, while the numerals 1, 2, 3, and so on are used in many different languages (Chrisomalis, 2020). Thus, we may observe a different pattern with proportionally more 1 on number words in a monolingual corpus.

In this study, we examine the frequency with which different numbers are used in a large corpus of spoken and written English. We observe how these frequencies are affected by magnitude, roundness, cultural salience, and register, and within written registers, we investigate how different formats (numerals, number words, mixed numbers) are used to represent numbers of different magnitudes. Our study, based on over 1.7 million numbers from 0 to a billion used in English in the 100 million word British National Corpus (BNC Consortium, 2007), makes the following contributions. First, it replicates previous corpus studies of number use in regard to magnitude and rounding (Coupland, 2011; Dehaene & Mehler, 1992; Dorogovtsev et al., 2005; Jansen & Pollmann, 2001). Second, it extends work by Sigurd (1988) and Jansen and Pollman (2001) by presenting an updated model of round numbers, exploring which roundness properties matter more than others in determining a number's frequency, controlling for its magnitude. This insight suggests that round numbers are not created equal – some are 'rounder' than others. Third, it investigates cultural salience as a general factor that may influence number frequencies, extending previous results pertaining to recent years (Pollmann, 1998; Pollmann & Baayen, 2001), and numbers with numerological significance (e.g., 13) (Dehaene & Mehler, 1992) or aesthetic appeal (e.g., 99) (Coupland, 2011). Fourth, it introduces register as a novel factor that influences number frequencies, investigating whether number use differs across more informational and less informational texts. Finally, it shows that, in writing, numbers of different magnitudes tend to be represented in different formats (numerals, number words, mixed numbers). In doing so, this study builds a detailed profile of number use in spoken and written English.

3. Methodology

3.1. The corpus

The British National Corpus (BNC; BNC Consortium, 2007), collated from 1991–1994, is a 100 million word corpus of spoken and written British English from the late 20th century. The corpus was slightly revised prior to the 2001 and 2007 releases (hence the citation date above) but no new texts were added. The corpus includes samples of 45,000 words from longer texts, whereas texts under the 45,000 word limit are included in full. Samples were taken to avoid overrepresenting idiosyncratic texts and obtain a wide cross-section of British English. The corpus contains both single-author texts (e.g., monographs) and multi-author texts (e.g., magazine articles).

The spoken (10%) and written (90%) subcorpora are divided into four categories, including two spoken categories (conversational speech and presentational speech) and two written categories (imaginative writing and informative writing). The conversational speech subcorpus was collected via demographic sampling, aiming to capture a representative spread of language users according to age, gender, social group, and region. In contrast, what we call presentational speech – 'context-governed speech' according to the BNC nomenclature – encompasses presentations to an audience, such as broadcast interviews, lectures, and legal proceedings. Regarding the written subcorpora, imaginative writing includes literary and creative fictional works, while informative writing includes non-fiction works in domains like applied science, finance, and world affairs. Broadly, presentational speech and informative writing are focused on communicating information, whereas conversational speech and imaginative writing are less informational. For more information about the BNC, see http://www.natcorp.ox.ac.uk/corpus/index.xml.

3.2. Software

The programming language Python (version 3.7) (Python Software Foundation, 2021) was used inside the integrated development environment PyCharm (version 2021.1.1) (JetBrains, 2021) to extract numbers from the British National Corpus (BNC Consortium, 2007). The following built-in Python libraries were used: Re (version 2.2.1), Time, OS, and IterTools (all: Van Rossum, 2020). The following external Python libraries were used: NumPy (version 1.20.2) (Harris et al., 2020), Pandas (version 1.2.4) (McKinney, 2010), NLTK (version 3.6.1) (Bird et al., 2009), BeautifulSoup (version 4.9.3) (Richardson, 2007), Requests (version 2.25.1) (Chandra & Varanasi, 2015), Word2Number (Batorsky et al., 2021), and Num2Words (Dupras et al., 2021).

The statistical programming language R (version 4.0.3) (R Core Team, 2020) was used inside integrated development environment RStudio (version 2022.7.1.554) (RStudio Team, 2022) to assist with the number identification and perform the main statistical analyses. The following R packages were used: tidyverse (version 1.3.0) (Wickham et al., 2019), brms (Bürkner, 2017, 2018), ggpubr (version 0.4.0) (Kassambara, 2020), scales (Wickham & Seidel, 2020), car (version 3.0.11) (Fox & Weisberg, 2019), ggmcmc (version 1.5.1.1) (Fernández-i-Marín, 2016), tidybayes (version 3.02) (Kay, 2022), and ggrepel (version 0.9.1). (Slowikowski, 2021). All data, analysis scripts, and information about the procedure detailed in this section are stored in an OSF repository (https://tinyurl.com/mr4dcc23).

3.3. Number identification and data processing

We identified integers and decimals ≥ 0 (i.e., no negative numbers) in the 4054 texts that comprise the BNC. Unlike previous corpus studies (Coupland, 2011; Dehaene & Mehler, 1992; Dorogovtsev et al., 2005; Jansen & Pollmann, 2001), which searched for specific numbers in their respective corpora, we used natural language processing to parse the BNC texts and capture all words identified as numbers. As a result, we identify a greater range of

numerical language than the aforementioned studies, but we have to perform extra steps to process the data, which we detail in this section. All decisions made in regard to number identification and data processing are fully documented with reproducible scripts at the OSF repository associated with this paper (https://tinyurl.com/mr4dcc23).

To understand our data processing decisions, we must differentiate between three senses in which numbers are used: cardinal, ordinal, and nominal (Nieder, 2005; Wiese, 2003). Cardinal numbers refer to quantity or numerosity, such as the number of apples in a basket. Ordinal numbers refer to numerical rank in an ordered list, such as the top ten universities in the world. Finally, nominal numbers are used as names or identifiers, as in telephone numbers or bus numbers, with no quantity or order necessarily being implied. Cardinal numbers are arguably the most prototypical number sense, and are what we focus on in this paper, motivating our analysis of numerical magnitude and roundness. For this reason, we do not analyse numbers that are explicitly marked as being ordinal (e.g., 'sixth', '6th'), in line with previous studies (Coupland, 2011; Jansen & Pollmann, 2001; but see Dehaene & Mehler, 1992). Thus, we focus here on a *specific form* of numbers: the most unmarked forms of number words ('one', 'two', 'three thousand', etc.) and numerals ('1', '2', '3000', etc.). We also exclude some instances of nominal numbers, as discussed below. However, it is impossible to make sure we only identify cardinal numbers in such a large dataset, as doing so would require checking each number in its context of use. The concern that our dataset is 'contaminated' by non-cardinal numbers is, however, shared with all previous analyses of number word and numeral frequencies (Coupland, 2011; Dehaene & Mehler, 1992; Dorogovtsev et al., 2005; Jansen & Pollmann, 2001). A benefit of our study is that our analysis explicitly acknowledges non-cardinal uses of numbers by investigating numbers whose frequency cannot be predicted accurately by focusing solely on magnitude and roundness (e.g., some culturally salient numbers).

We used the NLTK Python library to tag words in each text for whether they were numbers. The NLTK library tags each word individually, meaning that numbers comprising multiple words (e.g., 'twenty five' = two words) were initially tagged as separate numbers (e.g., 'twenty' and 'five'). To capture the full number, we grouped together words tagged as numbers that were adjacent in the same text. We included the words 'and' and 'point' in this grouping to capture number words containing a conjunction (e.g., 'one hundred and one') or decimal values (e.g., 'two point one'). Hyphenated number words, like 'twenty-five', were captured as a whole. NLTK was able to differentiate pronominal uses of number word 'one' (e.g., 'as one does') from numerical uses (e.g., 'one person'), which is a novel aspect of our methodology: these uses of 'one' are confounded in Dehaene and Mehler's (1992) and Coupland's (2011) corpus analyses of number words, and Jansen and Pollman (2001) chose not to look at the number word 'one' for this reason.

This number identification procedure identified 1,918,146 items (called as such because some items were not numbers, or included more than one number; see below). We then removed items that did not meet our definition of non-negative numbers that were not explicitly marked as ordinals. Also removed were items that were incorrectly tagged as numbers, did not refer to a specific number, or were not relevant for other reasons. In particular, we removed 146,797 items containing a mixture of both numerals and letters or other characters, including the aforementioned ordinals (e.g., '6th'), year ranges ('1990s'), file names (e.g., '011207.tmp'), digital times (e.g., '06:00'), and ratios (e.g., '2:1'). Furthermore, we removed number words like 'twelve fifty' (6016 items) that were ambiguous between different types of numbers, such as prices (e.g., £12.50), times (e.g., '12:50') and years (e.g., '1250'); non-numerical words that were erroneously tagged as numbers (e.g., 'year', 'ze': 3855 items); and vague numbers that did not refer to a precise value (e.g., 'twenty odd': 9 items). Finally, we excluded 5 items that included numerals preceded or followed by a full stop and then followed by a multiplier word (e.g., '1. million'), and 9407 items beginning with '0' that

resembled binary codes (e.g., '00010'), digital times (e.g., 07.15), or telephone numbers (e.g., 0207521133).[2] Following these exclusions, 1,750,086 items remained.

For the purposes of the data analysis, the number words we identified had to be translated into numerals so they could be recognised by R as numbers rather than as character strings. The Word2Num library in Python was used to translate the number words we identified into numerals. To verify each translation, we used the Num2Words library to back-translate the numerals into number words. We deemed the translation accurate if the back-translated number word matched the original number word. We manually translated some numbers that Word2Num was not able to automatically translate, including mixed numbers comprising a numeral followed by a multiplier word (e.g., '1.5 million' = 1,500,000) and numbers written in a different format (e.g., 'seventy one hundred' rather than 'seven thousand one hundred' = 7100). Furthermore, number words with an empty decimal place (e.g., 'one point zero') were manually translated into integers (e.g., '1').

Even after these manual translations, we were unable to translate some items into numerals. Consulting these untranslatable items, we saw that many were not individual numbers, but were concatenations of multiple numbers (e.g., 'one hundred one thousand' = 'one hundred' and 'one thousand'). These multi-number items were a by-product of our grouping together words tagged as numbers that were adjacent in the same text. To capture the individual numbers in these multi-number items separately, we classified the multi-number items into structural types, so, for example, items such as 'one and two' and 'eight and six' were both categorised as 'NUMBER and NUMBER'. Then, we updated the Python script to capture the numbers in these multi-number items separately, and translated them into numerals. We did this until 20 or fewer cases of each structural type remained (an arbitrary

[2] We have no systematic way of identifying binary codes, digital times, and telephone numbers that begin with a number other than 0, as these look the same as other numbers (e.g., 1001, 20.55, 3669734), so there may be some of these numbers in the final dataset.

threshold as it would have been impractical to account for every structural type due to the large size of the dataset). We excluded the remaining 2078 multi-number items from the dataset. To finish, we constrained the numbers in our dataset to the range 0 to a billion, which led to the exclusion of 6425 items.

This part of the procedure highlights a key limitation of an automatic approach to identifying numbers: when number words are written in nonstandard ways, or are not separated by commas, our automatic procedure may identify some of these numbers incorrectly (as blanket rules were established for how multi-number strings were dealt with, which may not be accurate for all cases), or these numbers may be excluded. Only a manual analysis would allow us to identify the numbers in these strings correctly in all instances. Ultimately, we believe that any numbers that are inaccurately identified will constitute noise in the data that should not affect the overall pattern of results we report, and that must be tolerated to facilitate the collection of such a large dataset.

Overall, out of 97,476,231 words in the BNC, the number identification procedure identified 1,739,343 numbers that met our search criteria. This figure constitutes 1.8% of the total number of words in the BNC mentioned above, but recall that some number word expressions comprised multiple words (e.g., ‘eight hundred’).

4. Results

4.1. Overall results

First, we analysed the influence of magnitude, roundness, and cultural salience overall by looking at the number frequencies across the whole of the BNC. Figure 1 displays the frequencies for numbers from 0 to a billion, excluding decimals. Number and frequency are visualised on base-10 logarithmic scales ($\log_{10}$) – that is, x or $y = 1$ represents 10, x or $y = 2$ represents 100, x or $y = 3$ represents 1000, x or $y = 4$ represents 10,000, and so on.

FIG. 1

The downward trend shows that frequency decreases as numerical magnitude increases: smaller numbers are used more often than larger numbers. In fact, the number 1 is used more often than any other number, occurring 302,676 times (18.1% of all integer tokens). The second most frequent number is 2, which occurs 180,079 times (10.8% of all integer tokens), and the third most frequent number is 3, which occurs 98,035 times (5.9% of all integer tokens). As an exception to this trend, 0 is the 57th most frequent number in the dataset, occurring only 4216 times (0.3% of all integer tokens). Zero is different from other integers because it usually represents the absence of a quantity, rather than the presence. Consequently, this number is not typically used in counting practices. It should, however, be noted that only the numeral '0' and the number word 'zero' were considered in this study – words such as 'none' and 'nought' were not included.

The figure also shows that round numbers are used more often than numbers of a comparable magnitude. For example, focusing on the round numbers circled and labelled, 100 is used 14,819 times, while 99 is used 881 times and 101 is used 429 times; 1000 is used 6975 times, while 999 is used 182 times and 1001 is used 54 times; 1 million is used 4778 times, while 999,999 is used 13 times and 1,000,001 is used 9 times; and 1 billion is used 1075 times, while 999,999,999 is not used at all.

The colour of the data points denotes whether the numbers are non-multiples or multiples of round numbers of different magnitudes: 10, 50, 100, 500, 1000, 10,000, and 100,000. The transition from dark blue to light turquoise rightward across the data points highlights the fact that, at higher magnitudes, the numbers that appear in the BNC are increasingly round numbers, and multiples of a larger round number (e.g., 110,000 is a multiple of 1000 and 10,000, while 111,000 is a multiple of 1000 but not 10,000). This result hints that people are more likely to round numbers at higher magnitudes, and to round to a

greater extent – for instance, rounding 110,789 down to 110,000 (the nearest multiple of 10,000) instead of rounding it up to 111,000 (the nearest multiple of 1000).

On top of the influence of magnitude and roundness, there are some other notable features of the figure. First, there is a spike in frequency starting at about $x = 3$ that reflects the high frequency of numerals from 1000–2000. Concordance lines extracted via the English Corpora interface (https://www.english-corpora.org/bnc/), which allows the user to conduct online searches of the BNC, showed that many of these numerals are years (e.g., 'the year 1990'), and are used with increasing frequency as they approach the year the BNC's collation was completed (1994). Following Pollman and Baayen (2001), we might infer that these numbers are used more frequently because recent dates are more relevant to discuss and so are more culturally salient. The second notable feature of the figure is the spike in frequency just before $x = 5$, which roughly represents the range 36,000–40,000. Concordance lines for the most frequent of these numbers (e.g., 38,166) show that many of these numbers are citations of page numbers mentioned frequently in multiple political and economic records from the Keesings Contemporary Archives in the informative writing subcorpus. The pages to which these numbers refer contain information about widely significant, relevant topics (e.g., the Kuwait Democratic Forum). The high frequency of these numbers appears to be an artifact of how the BNC's compilers sampled texts to include in the corpus, rather than these numbers being culturally salient.

4.2. Statistical models

To understand the relative influence of magnitude, roundness, and cultural salience on number frequencies, we statistically modelled number frequency as a function of magnitude and roundness. We then identified the numbers used more frequently than predicted when controlling for magnitude and roundness, whose frequency may be attributable to cultural salience.

Bayesian negative binomial regression was used to model the number frequencies (see Winter & Bürkner, 2021). We included numbers that do not appear in the BNC in this regression, which were assigned a frequency of 0, in addition to the numbers that were identified (i.e., ones that appear at least once). This step is justified because the frequency of the 'missing' numbers is genuinely 0 – it is not just that we did not identify them. Number frequencies are positive integers, and so are amenable to Poisson regression, which can be used to model unbounded count data. However, because there is more variance in these count data than what is expected under the Poisson distribution (known as overdispersion; see OSF repository for evidence of overdispersion in the data: https://tinyurl.com/mr4dcc23) negative binomial regression was used, which modelled both the mean and the variance in counts. Like Poisson regression, negative binomial regression uses the log link function (natural logarithm), which means that all coefficients reported below model the $\log_e$ number frequencies.

We modelled number frequencies for the range 1 to 1 million. We did not include 0 because logarithmically transforming 0 is not possible, and we imposed an upper limit of 1 million to mitigate the computational difficulties and time demands associated with analysing larger datasets in this manner (all numbers from 1 to 1 million were analysed, including numbers with a frequency of 0). We used default priors from the R package brms (Bürkner, 2017, 2018) for the intercept and standard deviation. We set weakly informative priors on the independent variable slopes (normal distribution centred at 0, standard deviation of 0.5) to build "mild skepticism" (McElreath, 2016, p.186) into our analyses. Weakly informative priors bias slope estimates slightly towards zero, making our results more conservative when compared to a corresponding frequentist model.

The dependent variable in the model was Frequency and the independent variables were Log_{10} Number Magnitude and the roundness properties Multiple of 5, Multiple of 10, 10-

ness, 2-ness, 2.5-ness, and 5-ness.[3] Variance Inflation Factors (VIFs) for the roundness properties in a linear regression model were all below 3, indicating that collinearity was not an issue (Winter, 2019; see OSF repository for VIFs: https://tinyurl.com/mr4dcc23).[4]

The coefficients from this model confirm that Log_e Frequency declines with Log_{10} Number Magnitude (β = –3.62, 95% *Bayesian credible interval* = [–3.65, –3.59]). From highest to lowest, the credible intervals for the roundness properties are: 10-ness (β = 4.46, 95% *CI* = [4.06, 4.88]), 2.5-ness (β = 3.84, 95% *CI* = [3.42, 4.29]), 5-ness (β = 3.39, 95% *CI* = [2.95, 3.87]), 2-ness (β = 2.74, 95% *CI* = [2.29, 3.20]), Multiple of 10 (β = 2.45, 95% *CI* = [2.38, 2.53]), and Multiple of 5 (β = 0.06, 95% *CI* = [–0.01, 0.13]). All the credible intervals are well above zero, except for Multiple of 5, which slightly overlaps with, but is mostly above, zero. All roundness properties thus predict that a number will be used more frequently, but Multiple of 5 much less so than the other roundness properties. Furthermore, numbers with more roundness properties (e.g., 200, which is a Multiple of 10, a Multiple of 5, and also has 10-ness, 2.5-ness, 5-ness, and 2-ness) are predicted to be used more frequently than numbers with fewer roundness properties (e.g., 15,000, which is a Multiple of 10, a Multiple of 5, and also has 2.5-ness and 5-ness). Furthermore, some credible intervals do not overlap, indicating that some roundness properties are more predictive of number frequency than others, with 10-ness being the most predictive (more predictive than 5-ness, 2-ness, 2.5-ness, Multiple of 10, and Multiple of 5), followed by 2.5-ness (more predictive than 2-ness, 2.5-ness, Multiple

[3] Technically, the mathematical definitions of 10-ness, 2-ness, 2.5-ness, and 5-ness proposed by Jansen and Pollman (2001) mean that all numbers from 1 to 9 have 10-ness, because dividing these numbers by 1 ($1 \times 10^0 = 1$) equals an integer no greater than 9. Similarly, all multiples of 2 from 2 to 8 have 2-ness, because dividing these numbers by 2 ($2 \times 10^0 = 2$) equals an integer no greater than 9. As our treatment of roundness assumes that all round numbers are multiples of ten or five (although not all multiples of ten or five are necessarily round), we avoided this outcome by only considering the first power of ten and above for all roundness properties (rather than the zeroth).

[4] There is currently no implementation of VIFs in the brms R package, and in any case, whether the model is Bayesian or frequentist does not affect assessments of collinearity.

of 10, and Multiple of 5), 5-ness (more predictive than 2-ness, Multiple of 10, and Multiple of 5), 2-ness and Multiple of 10 (both more predictive than Multiple of 5), and Multiple of 5.

We then identified numbers whose high frequency cannot be explained by either magnitude or roundness by examining the model residuals, which show how different the actual number frequencies are from the model's predictions (see Winter, 2019, Ch. 4). High residuals reflect numbers that are used more often than expected based on magnitude and roundness alone, factoring out the influence of the model's predictors. To better understand why these numbers are used so frequently, we consulted concordance lines for them using the English Corpora interface (https://www.english-corpora.org/bnc/). A full list of the residuals is provided in the OSF repository for this paper (https://tinyurl.com/mr4dcc23).

The ten numbers with the largest residuals are all between 1984 and 1993, reflecting the prevalence of dates in the BNC, especially those recent to the completion of the BNC's collation in 1994. If we exclude numbers between 1000 and 2000, the ten largest residuals are for numbers with at least one roundness property (1st: 250,000, 2nd: 300,000, 3rd: 12,000). If we exclude numbers with any roundness properties, the largest residual is for the number 80486, which refers to a microprocessor. The second largest residual is for the number 2001, which refers to the year and a savings plan called SaverPlus 2001. The third largest residual is for the number 999, which is one of the UK's emergency services numbers. The numbers at ranks 8–10 in this list (38,211, 37,838, 37,914) are page numbers mentioned in texts from the Keesings Contemporary Archives.

4.3. Analysis of subcorpora

We now turn our attention to comparisons across and within the BNC subcorpora. To investigate the effect of register on number use, we compared more informational (presentational speech, e.g., lectures, educational demonstrations, classroom interactions; informative writing, e.g., non-fiction books about applied science, the arts, world affairs, etc.)

and less informational contexts (conversational speech; imaginative writing, i.e. fiction books) within speech and writing respectively. Across these registers, we compared: (1) Numerical density – how often numbers are used in general; (2) Numerical diversity – how varied the set of numbers used is; (3) Decimals – how frequently decimals are used; and (4) Magnitude – how frequently smaller versus larger numbers are used. Then, in the written subcorpora, we investigated the format (numerals, number words, mixed numbers) used to represent numbers of different magnitudes.

4.3.1. Numerical density

Relative to the size of each spoken subcorpus, people use numbers about the same amount in presentational speech (1.6%, 97,589 number tokens out of 5,987,379 words) and conversational speech (1.6%, 63,196 number tokens out of 3,976,158 words). Relative to the size of each written subcorpus, people use numbers more frequently in informative writing (2.1%, 1,473,425 number tokens out of 71,355,964 words) than in imaginative writing (0.7%, 105,133 number tokens out of 16,156,730 words). Thus, in writing, the more informational subcorpus (informational writing) is more numerically dense than the less informational subcorpus (imaginative writing), whereas in speech, there is not a substantial difference in numerical density between the more informational (presentational speech) and less informational (conversational speech) subcorpora.

4.3.2. Numerical diversity

When people speak about numbers, they use more different numbers in presentational speech (1633 types, type-token ratio (TTR) = 1.7%) than in conversational speech (634 types, TTR = 1.0%). When people write about numbers, they use more different numbers in informative writing (27,016 types, TTR = 1.8%) than in imaginative writing (1412 types, TTR = 1.3%). Hence, in their respective modalities, the more informational subcorpora

(presentational speech and informative writing) are more numerically diverse than the less informational subcorpora (conversational speech and imaginative writing).

4.3.3. Decimals

When people talk about numbers, they use more decimals in presentational speech (1.0%, 978 decimals out of 97,589 number tokens) than in conversational speech (0.5%, 297 decimals out of 63,196 number tokens). When people write about numbers, they use more decimals in informative writing (4.6%, 68,257 decimals out of 1,473,425 numbers) than in imaginative writing (0.6%, 665 decimals out of 105,133 numbers). These results show that more numbers used in the more informational subcorpora (presentational speech and informative writing) are decimals than in the less informational subcorpora (conversational speech and imaginative writing) in their respective modalities.

4.3.4. Magnitude

Figure 2 shows the proportion of number tokens in different $\log_{10}$ number ranges (i.e., magnitudes) in the different subcorpora. In all subcorpora, smaller numbers are used more often than larger numbers.

FIG. 2

When we compare the subcorpora, we see that smaller numbers are proportionally more frequent in the less informational subcorpora compared to the more informational subcorpora in their respective modalities: the $\log_{10}$ number range 0–<1 (numbers 1–9) is proportionally more frequent for conversational speech (74.7%, 47,143 out of 63,091 tokens) than presentational speech (66.4%, 64,675 out of 97,332 tokens), and for imaginative writing (78.2%, 82,175 out of 105,045 tokens) than informative writing (44.4%, 649,597 out of 1,462,196 tokens).

In comparison, larger numbers are proportionally more frequent in the more informational subcorpora than in the less informational subcorpora in their respective modalities: the $\log_{10}$ number ranges 1–9 (numbers 10–1 billion) are proportionally more frequent for presentational speech (33.6%, 32,657 out of 97,332 tokens) than conversational speech (25.3%, 15,948 out of 63,091 tokens), and for informative writing (55.6%, 812,599 out of 1,462,196 tokens) than imaginative writing (21.8%, 22,870 out of 105,045 tokens).

The proportions of number tokens become increasingly small for the higher $\log_{10}$ number ranges as larger numbers are increasingly unlikely to be discussed in general, meaning that proportional differences between the subcorpora become trivially small. Also note that the relative increase for written numbers in $\log_{10}$ number range 3–<4 is due to numbers in the range 1000–2000 occurring frequently, often referring to years.

4.3.5. Writing: Representational formats

Figure 3 displays the proportions of numbers in each $\log_{10}$ number range that were represented as a numeral, number word, or mixed number in writing. We focus on writing because the representational format of a number in writing is a decision made by the text's author, whereas in the spoken texts, this format is a matter of transcription and so does not reflect how the number was spoken in context.

Overall, the figure shows a transition from the dominance of number words for the smallest numbers, to numerals for comparatively larger numbers, to mixed numbers for the largest numbers. In particular, numbers 1–9 ($\log_{10}$ number range 0–<1, 731,772 tokens) are mostly represented by number words (66.46, 486,316 tokens) and less by numerals (33.54%, 245,456 tokens). Numbers 10–999,999 ($\log_{10}$ number range 1–<6, 811,536 tokens) are mostly denoted by numerals (92.77%, 752,834 tokens), followed by number words (7.23%, 58,647 tokens), and mixed numbers (0.01%, 55 tokens). Numbers 1 million–9,999,999 ($\log_{10}$ number range 6–<7, 11,888 tokens) are mostly represented by mixed numbers (44.03%, 5234 tokens),

followed by number words (38.45%, 4571 tokens), and then numerals (17.52%, 2083 tokens). Similarly, numbers 10 million–99,999,999 ($\log_{10}$ number range 7–<8, 6413 tokens) are mostly denoted by mixed numbers (73.85%, 4736 tokens), but, unlike the preceding number range, are denoted more by numerals (20.58%, 1320 tokens) than by number words (5.57%, 357 tokens). This range is also more likely to be denoted by mixed numbers than the previous range. We see a similar pattern for numbers 100 million–1 billion ($\log_{10}$ number range 8–9, 5632 tokens), which are mostly denoted by mixed numbers (72.46 %, 4081 tokens), followed by numerals (16.46%, 927 tokens), and then number words (11.08%, 624 tokens).

FIG. 3

5. Discussion

This study is the largest, most comprehensive analysis of number use to date. Our results have a number of important implications. First, they confirm the findings of previous studies (Coupland, 2011; Dehaene & Mehler, 1992; Dorogovtsev et al., 2005; Jansen & Pollmann, 2001) that frequency declines as numerical magnitude increases, and that round numbers are used more frequently than unround numbers of a similar magnitude. As discussed in detail in section 2, small numbers may be used more often due to the ease of mentally processing these numbers (Dehaene & Mehler, 1992), because they are more relevant to discuss (Cummins, 2015, p. 32), and because we use scales that keep numbers small (Coupland, 2011, pp. 34–35). Round numbers may be used more frequently due to their psychological salience (Van der Henst & Sperber, 2004) and cognitively accessibility or simplicity (e.g., Dehaene, 1997), and also because the exact value may be unknown (Ruud et al., 2014) or irrelevant to discuss (e.g., Van der Henst et al., 2002). People may also use round numbers to avoid seeming pedantic (e.g., Lin, 2013), and for strategic manipulation (Cummins & Franke, 2021). However, our data do not allow us to differentiate between these different

factors: we can state that smaller numbers and round numbers are used more often than larger and unround numbers, but our data do not tell us why.

Our findings clearly show that round numbers are not created equal: some are used more often and so may be seen as 'rounder' than others. First, even multiples of ten without 10-ness, 2-ness, 2.5-ness, or 5-ness (e.g., 810, 10,070) – and to a much lesser extent, multiples of five without 10-ness, 2-ness, 2.5-ness, or 5-ness (e.g., 815, 10,075) – are used more often than non-multiples. Second, numbers with more roundness properties are more frequent than numbers with fewer roundness properties, extending Jansen and Pollman's (2001) findings by showing that this result also applies if we treat being a multiple of ten or being a multiple of five as roundness properties. Third, supporting Cummins' (2015, p. 34) speculation that some roundness properties may be more important than others in determining roundness, we find that 10-ness is the most important factor, followed by 2.5-ness, 5-ness, 2-ness and being a multiple of ten (which were tied), and, lastly, being a multiple of five.

The concept of 'roundness' may thus not be either/or but rather may be a radial category with graded membership (Lakoff, 1987, Ch. 6), consistent with prototype theory (Rosch, 1973; Taylor, 2003). Numbers with more roundness properties may be closer to the centre of this category (i.e., be more prototypical of it) because they have more features associated with it (e.g., 10-ness, 2-ness), just as a robin is a more prototypical example of a bird than a penguin, because a robin has the feature 'can fly' that is associated with the *bird* category (e.g., Rosch, 1975; Rossiter & Best, 2013). A radial category of roundness is illustrated in Figure 4. A more complete depiction of roundness as a radial category would account for the different influence of different roundness properties (e.g., 10-ness > 2-ness, 2.5-ness and Multiple of 10).

FIG. 4

Our results also suggest that number use is less precise at higher magnitudes: people round larger numbers to a greater extent than smaller numbers (see also Coupland, 2011). This aspect of numerical communication may reflect the less precise nature of numerical cognition at higher magnitudes (Cheyette & Piantadosi, 2019; McCrink & Wynn, 2007; Xu & Spelke, 2000), which is captured with the Weber-Fechner law (DeWind et al., 2015; Shepard et al., 1975). Because this explanation appeals to the notion that we find quantities more difficult to discriminate between at higher magnitudes, it probably only applies to cardinal numbers (i.e., those denoting quantities) rather than, for example, the calendar years in our dataset, whose numerical value denotes their position in an ordinal sequence of years, rather than the number of years per se. For years, larger numbers (i.e., more recent years) may be easier to discriminate between than smaller numbers (i.e., less recent years), because more recent years may be easier to recall accurately. For example, while it may be easy to recall whether one went to Milan in either 2017 or 2018, it may be harder to recall whether one went to Rome in either 2007 or 2008.[5]

We also identified numbers that were used frequently because they are culturally salient. For example, confirming previous studies (Pollmann, 1998; Pollmann & Baayen, 2001), numbers between 1000 and 2000 were especially frequent as many were ordinal numbers that referred to years, particularly recent ones, which are usually more relevant to present discussion. The number 999 was also used relatively often in the BNC as it is one of the UK's emergency services numbers, a nominal use of 999. Crucially, if the number 999 were not culturally salient, we would hardly expect it to be used at all, because in all other contexts it would probably be rounded to 1000. Rather than reflecting cultural salience, other frequently used numbers seemed to reveal issues of unrepresentativeness in the BNC. The most obvious example is in the presence of ordinal page numbers from texts from the

[5] We thank an anonymous reviewer for this point and example.

Keesings Contemporary Archives, which probably reflects the texts sampled for inclusion in the BNC rather than prominent use of these numbers outside this context. Other examples include the use of numbers as names for products, like the SaverPlus 2001 savings plan. While the nominal use of numbers in branding can certainly lead to numbers being culturally salient, such as 747 (Boeing jets), 501 (Levi jeans), and 57 (Heinz), it is doubtful whether the number 2001 is salient outside the BNC, at least not due to its use in this particular brand name. Future studies could conduct diachronic analyses to determine whether different numbers have gained or lost cultural salience over time. As cultures evolve, so too may the preference for communicating about different numbers, and the ways about which these numbers are communicated. For example, Chrisomalis (2020) demonstrates that the popularity of the expression '1.2 million' has surpassed the equivalent expressions 'twelve hundred thousand' and 'one million two hundred thousand' over the years 1800–2000.

This study also investigated number frequencies across registers: across more informational (presentational speech, informative writing) and less informational (conversational speech, imaginative writing) contexts. We found that, in their respective modalities, the more informational subcorpora were more numerically diverse (people used more varied numbers) and contained more decimals and larger numbers. In writing, we found that informative writing was more numerically dense (people used more numbers in general) than imaginative writing. Overall, these measures suggest that numbers are integral to contexts in which communicating information is important. People may use numbers in informational registers to quantify phenomena for the benefit of one's audience – such as when listing ingredients for a recipe (e.g., 2 tablespoons of tomato purée, ½ a teaspoon of dried marjoram) and describing how many people the meal will serve (e.g., 4 people) (see BBC Good Food, 2022) – or to precisely record financial transactions (Coupland, 2011). People may also use numbers to establish credibility (Koetsenruijter, 2011) or convey an air of objectivity (Porter, 1995). The fact that we found register differences is consistent with

previous studies that found linguistic differences between English registers (Biber, 2009a, 2009b; Biber et al., 1999; Biber & Clark, 2002; Biber & Gray, 2010; Conrad & Biber, 2009; Egbert & Mahlberg, 2020; Love et al., 2019; Neumann, 2014).

Comparing representational formats in writing, we found that the numbers 1–9 were mostly represented as number words, whereas 10–999,999 were mostly represented as numerals, and 1 million–1 billion were mostly denoted by a mix of numerals and multiplier words, consistent with the recommendations of many writing style guides (e.g., Office for National Statistics, 2022). These conventions may stem from the fact that larger numbers, especially unround numbers, are often longer to write and harder to parse when written as number words (e.g., 'one hundred thousand, seven hundred and fifty five') than as numerals (e.g., '100,755'). However, even numerals may be difficult to parse for very large numbers (e.g., '1,000,000,000'), which may explain why a mix of numerals and number words are preferred at these magnitudes (e.g., '1 billion'). Our results are consistent with Coupland's (2011) in that we found that number words were used proportionally more often for smaller numbers. However, unlike this previous study, we found that number words were dominant over numerals for the range 1–9. This discrepancy may be attributable to the fact that the internet contains numerals from all languages, which was contrasted with number words in English, whereas our analysis compares numerals and number words used exclusively in English.

People talk and write about numbers in many different ways, which are difficult to account for exhaustively in a study like this that paints with broad brushstrokes to capture numerical communication on a large scale. In the interest of feasibility when working with large-scale data, blanket rules were established for what number expressions would be considered, which were necessarily imperfect. Our methodology automatically identified words that are typically used to denote numbers (e.g., 'one', 'hundred'), but some numerical expressions were not identified (e.g., 'nought', 'a dozen', 'a couple'), or were probably

captured as separate numbers even though they were part of a larger expression (e.g., ‘two and a half thousand’ = ‘two’ and ‘thousand’; ‘one over three’ = ‘one’ and ‘three’). As a result, we did not examine more non-standard or colloquial ways of referring to numbers (e.g., ‘a couple hundred’), which may be more prevalent in speech than in writing, potentially leading us to underestimate number use in speech. Our dataset may also be slightly noisy or incomplete generally for the reasons described above. However, we believe that the results we report would hold with a more complete identification of numbers, due to the strength of the patterns observed and the sheer size of the dataset, and because our results replicate many of the patterns revealed in previous corpus analyses (Coupland, 2011; Dehaene & Mehler, 1992; Dorogovtsev et al., 2005; Jansen & Pollmann, 2001).

To replicate our findings, it may be instructive to reproduce these analyses with another English language corpus, such as the Corpus of Contemporary American English (COCA; Davies, 2008), or a corpus in a different language, for example, from the TenTen Corpus Family (Jakubíček et al., 2013). Doing so would reveal whether our results apply across language varieties – for example, from British (BNC) to American English (COCA) – and to other languages from the same (e.g., German) or a different (e.g., Arabic) language family. It may also reveal differences in the numbers that are salient across cultures. For example, we might expect 911 (rather than 999) to be frequent in COCA, as 911 is the US’s emergency services number. Other corpora also have a different makeup of registers than the BNC, affording different kinds of comparative analyses than the more versus less informational comparison reported here. A replication could also be conducted with the 2004 version of the BNC (Brezina et al., 2021; Love et al., 2017) when the full written English texts are downloadable (only the spoken English texts are currently available).

People use some numbers more often than others. This large-scale study has shed light on the numbers people use more often, in what contexts, and why, in connection to magnitude, roundness, cultural salience, and register, and has identified the formats with

which writers choose to represent these numbers. In doing so, it has constructed a detailed profile of number use in spoken and written English.

References

Ayonrinde, O. A., Stefatos, A., Miller, S., Richer, A., Nadkarni, P., She, J., Alghofaily, A., & Mngoma, N. (2021). The salience and symbolism of numbers across cultural beliefs and practice. *International Review of Psychiatry*, *33*(1–2), 179–188. https://doi.org/10.1080/09540261.2020.1769289

Barchas-Lichtenstein, J., Voiklis, J., Attaway, B., Santhanam, L., Parson, P., Thomas, U. G., Isaacs-Thomas, I., Ishwar, S., & Fraser, J. (2022). Number soup: Case studies of quantitatively dense news. *Journalism Practice*, *0*(0), 1–28. https://doi.org/10.1080/17512786.2022.2099954

Batorsky, B., Ledvosky, A., Yarkoni, T., & Groove, B. (2021). *Word2Number*. https://w2n.readthedocs.io/en/latest/

BBC Good Food. (2022). *Chilli con carne recipe*. BBC Good Food. https://www.bbcgoodfood.com/recipes/chilli-con-carne-recipe

Beltrama, A., Solt, S., & Burnett, H. (2022). Context, precision, and social perception: A sociopragmatic study. *Language in Society*, 1–31. https://doi.org/10.1017/S0047404522000240

Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition*, *124*(2), 156–182. https://doi.org/10.1016/j.cognition.2012.05.005

Biber, D. (2009a). Are there linguistic consequences of literacy? Comparing the potentials of language use in speech and writing. In D. R. Olson & N. Torrance (Eds.), *The Cambridge Handbook of Literacy* (pp. 75–91). Cambridge University Press. https://www.cambridge.org/core/books/cambridge-handbook-of-literacy/are-there-linguistic-consequences-of-literacy-comparing-the-potentials-of-language-use-in-speech-and-writing/3210F2D761DCC022B4CDAA048C55D3AA

Biber, D. (2009b). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275–311. https://doi.org/10.1075/ijcl.14.3.08bib

Biber, D., & Clark, V. (2002). How long can you go without a verb?: Historical shifts in modification patterns with complex noun phrase structures. In T. Fanego, J. Pérez-Guerra, & M. J. López-Couso (Eds.), *English Historical Syntax and Morphology: Selected papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000.Volume 1* (pp. 43–66). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.223.06bib

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (Eds.). (1999). *Longman Grammar of Spoken and Written English*. Longman.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* (1st edition). O'Reilly Media.

BNC Consortium. (2007). *British National Corpus* (Version 3). BNC Consortium. http://www.natcorp.ox.ac.uk/

Brezina, V., Hawtin, A., & McEnery, T. (2021). The Written British National Corpus 2014 – design and comparability. *Text & Talk*, *41*(5–6), 595–615. https://doi.org/10.1515/text-2020-0052

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, *103*(6), 1131–1136. https://doi.org/10.1037/h0037361

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411.

Butterworth, B. (1999). *The Mathematical Brain*. Macmillan.

Butterworth, B., Reeve, R., Reynolds, F., & Lloyd, D. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *Proceedings of the National Academy of Sciences*, *105*(35), 13179–13184. https://doi.org/10.1073/pnas.0806045105

Chandra, R. V., & Varanasi, B. S. (2015). *Python Requests Essentials*. Packt Publishing Ltd.

Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences*, *116*(36), 17729–17734. https://doi.org/10.1073/pnas.1819956116

Chrisomalis, S. (2020). *Reckonings: Numerals, Cognition, and History*. The MIT Press.

Conrad, S., & Biber, D. (2009). *Real Grammar: A Corpus-based Approach to English*. Pearson/Longman.

Cotterill, J. (2007). 'I think he was kind of shouting or something': Uses and abuses of vagueness in the British courtroom. In J. Cutting (Ed.), *Vague Language Explored* (pp. 97–114). Palgrave Macmillan UK. https://doi.org/10.1057/9780230627420_6

Coupland, N. (2011). How frequent are numbers? *Language and Communication*, *31*(1), Article 1. https://doi.org/10.1016/j.langcom.2010.09.001

Cummins, C. (2015). *Constraints on Numerical Expressions*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199687909.001.0001

Cummins, C., & Franke, M. (2021). Rational interpretation of numerical quantity in argumentative contexts. *Frontiers in Communication*, *6*. https://www.frontiersin.org/articles/10.3389/fcomm.2021.662027

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics*, *45*(6), 557–566. https://doi.org/10.3758/BF03208063

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29.

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016

Dorogovtsev, S., Mendes, J. F., & Oliveira, J. (2005). Frequency of occurrence of numbers in the World Wide Web. *Physica A: Statistical Mechanics and Its Applications*, *360*, 548–556. https://doi.org/10.1016/j.physa.2005.06.064

Dupras, V., Ortiz, E. R., Szalaï, I., Lodato, M., & Harper, B. (2021). *Num2Words*. Savoir-faire Linux Inc.

Egbert, J., & Mahlberg, M. (2020). Fiction – one register or two?: Speech and narration in novels. *Register Studies*, *2*(1), 72–101. https://doi.org/10.1075/rs.19006.egb

Fernández-i-Marín, X. (2016). ggmcmc: Analysis of MCMC Samples and Bayesian Inference. *Journal of Statistical Software*, *70*(9). https://doi.org/10.18637/jss.v070.i09

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression.* (3rd ed.). Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Francis, W. N., & Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin.

Gibbs Jr., R. W., & Bryant, G. A. (2008). Striving for optimal relevance when answering questions. *Cognition*, *106*(1), 345–369. https://doi.org/10.1016/j.cognition.2007.02.008

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*(5695), 496–499. https://doi.org/10.1126/science.1094492

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Speech Arts* (pp. 41–58). Academic Press.

Gunasti, K. (2016). Consumer reactions to round numbers in brand names. *Marketing Letters*, *27*(2), 309–322.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), Article 7825. https://doi.org/10.1038/s41586-020-2649-2

Ifrah, G. (1987). *From One to Zero: A Universal History of Numbers*. Penguin Books.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen Corpus Family. *Proceedings of the 7th International Corpus Linguistics Conference*, 125–127.

Jansen, C. J. M., & Pollmann, M. M. W. (2001). On round numbers: Pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics*, *8*(3), 187–201. https://doi.org/10.1076/jqul.8.3.187.4095

JetBrains. (2021). *PyCharm* (2021.1.1).

Kassambara, A. (2020). *ggpubr: 'ggplot2' based publication ready plots* (0.4.0). https://CRAN.R-project.org/package=ggpubr

Kay, M. (2022). *tidybayes: Tidy data and geoms for Bayesian models*. http://mjskay.github.io/tidybayes/>

Koetsenruijter, A. W. M. (2011). Using numbers in news increases story credibility. *Newspaper Research Journal*, *32*(2), 74–82. https://doi.org/10.1177/073953291103200207

Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In E. Hinrichs & J. Nerbonne (Eds.), *Theory and Evidence in Semantics* (pp. 109–132). CSLI Publications.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.

Lasersohn, P. (1999). Pragmatic halos. *Language*, *75*(3), 522–551. https://doi.org/10.2307/417059

Lin, Y.-L. (2013). Vague language and interpersonal communication: An analysis of adolescent intercultural conversation. *International Journal of Society, Culture & Language*, *1*(2), 69–81.

Love, R., Brezina, V., McEnery, T., Hawtin, A., Hardie, A., & Dembry, C. (2019). Functional variation in the Spoken BNC2014 and the potential for register analysis. *Register Studies*, *1*(2), 296–317. https://doi.org/10.1075/rs.18013.lov

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations: *International Journal of Corpus Linguistics*, *22*(3), 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

McCarthy, M. (2006). *Explorations in Corpus Linguistics*. Cambridge University Press Marketing Department.

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science*, *18*(8), 740–745. https://doi.org/10.1111/j.1467-9280.2007.01969.x

McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press/Taylor & Francis Group.

McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, *445*, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Neumann, S. (2014). Cross-linguistic register studies: Theoretical and methodological considerations. *Languages in Contrast*, *14*(1), 35–57. https://doi.org/10.1075/lic.14.1.03neu

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, *6*(3), Article 3. https://doi.org/10.1038/nrn1626

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), Article 6. https://doi.org/10.1038/nrn.2016.40

Office for National Statistics. (2022). *Writing Numbers*. Office for National Statistics. https://style.ons.gov.uk/house-style/writing-numbers/

Pica, P., Lemer, C., Dehaene, S., & Izard, V. (2004). Exact and approximative arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Pokryshevskaya, E. B., & Antipov, E. A. (2015). A study of numerological superstitions in the apartments market. *Economics Bulletin*, *35*(1), 83–88.

Pollmann, T. (1998). On forgetting the historical past. *Memory & Cognition*, *26*(2), 320–329. https://doi.org/10.3758/BF03201143

Pollmann, T., & Baayen, R. H. (2001). Computing historical consciousness. A quantitative inquiry into the presence of the past in newspaper texts. *Computers and the Humanities*, *35*(3), 237–253.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (1st edition). Princeton University Press.

Python Software Foundation. (2021). *Python* (3.7). https://www.python.org/

R Core Team. (2020). *R.* (4.0.3) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Rath, G. J. (1966). Randomization of humans. *The American Journal of Psychology*, *79*(1), 97–103. https://doi.org/10.2307/1420712

Richardson, L. (2007). *Beautiful Soup Documentation*. https://buildmedia.readthedocs.org/media/pdf/beautiful-soup-4/latest/beautiful-soup-4.pdf

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

Rosch, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

Rossiter, C., & Best, W. (2013). "Penguins don't fly": An investigation into the effect of typicality on picture naming in people with aphasia. *Aphasiology*, *27*(7), 784–798. https://doi.org/10.1080/02687038.2012.751579

RStudio Team. (2022). *RStudio: Integrated Development Environment for R* (2022.7.1.554) [R]. RStudio, PBC. http://www.rstudio.com/

Ruud, P., Schunk, D., & Winter, J. (2014). Uncertainty causes rounding: An experimental study. *Experimental Economics*, *17*(3), 391–413.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, *7*(1), 82–138. https://doi.org/10.1016/0010-0285(75)90006-7

Sigurd, B. (1988). Round numbers. *Language in Society*, *17*(2), 243–252.

Slowikowski, K. (2021). *ggrepel: Automatically position non-overlapping text labels with 'ggplot2'* (0.9.1) [R]. https://CRAN.R-project.org/package=ggrepel

Taylor, J. R. (2003). *Linguistic Categorization* (3rd ed). Oxford University Press.

Trudgill, P. (1983). *On Dialect: Social and Geographical Perspectives*. New York University Press.

Van der Henst, J.-B., Carles, L., & Sperber, D. (2002). Truthfulness and relevance in telling the time. *Mind and Language*, *17*(5), 457–466. https://doi.org/10.1111/1468-0017.00207

Van der Henst, J.-B., & Sperber, D. (2004). Testing the cognitive and communicative principles of relevance. In I. A. Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 141–171). Palgrave Macmillan UK. https://doi.org/10.1057/9780230524125_7

Van Rossum, G. (2020). *The Python Library Reference, Release 3.8.2* (2.2.1). Python Software Foundation.

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

Wickham, H., & Seidel, D. (2020). *scales: Scale Functions for Visualization* (1.1.1) [R]. https://CRAN.R-project.org/package=scales

Wiese, H. (2003). *Numbers, language, and the human mind*. Cambridge University Press.

Wiese, H. (2004). *Numbers, Language, and the Human Mind*. Cambridge University Press.

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Routledge.

Winter, B., & Bürkner, P.-C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and Linguistics Compass*, *15*(11), e12439. https://doi.org/10.1111/lnc3.12439

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/s0010-0277(99)00066-9

Zillmann, D., & Brosius, H.-B. (2000). *Exemplification in communication: The influence of case reports on the perception of issues* (pp. x, 156). Lawrence Erlbaum Associates Publishers.

Roundness Property	Examples	
	Has property	Does not have property
Multiple of 5	15	57
	765	313
	94,300	61,692
	525,635	865,387
Multiple of 10	90	96
	760	592
	73,270	92,133
	198,250	285,989
10-ness	80	47
	300	523
	3000	18,341
	600,000	527,374
2-ness	60	64
	200	897
	1400	31,128
	100,000	587,084
2.5-ness	50	58
	125	903
	10,000	87,882
	500,000	140,770
5-ness	100	84
	4500	751
	45,000	17,329
	350,000	187,772

Table 1. Examples of numbers with properties associated with being a round number: Multiple of 5, Multiple of 10, 10-ness, 2-ness, 2.5-ness, and 5-ness. Numbers can have more than one of these properties at once (e.g., 100 has all six roundness properties; 60 is a multiple of five and ten and has 10-ness and 2-ness).

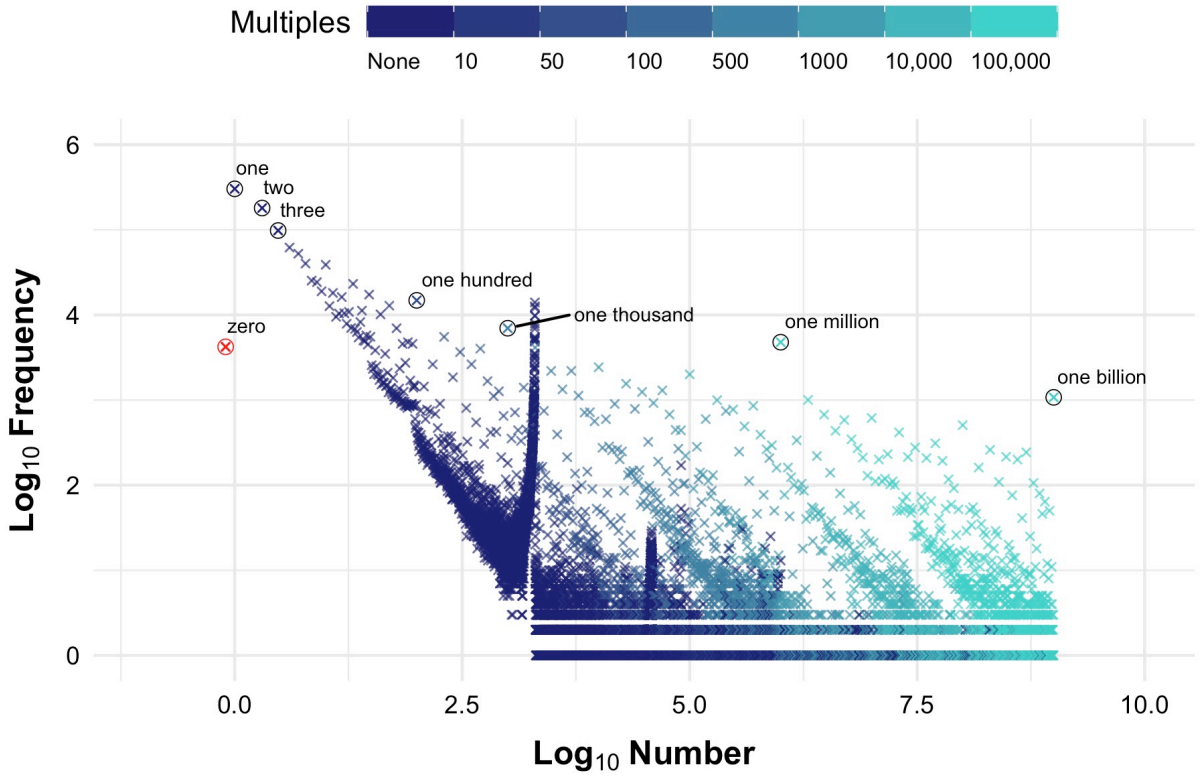


Figure 1. Frequencies for all integers that appear in the British National Corpus in any representational format. Both axes are base-10 logarithmically scaled ($\log_{10}$). Non-multiples and multiples of 10, 50, 100, 500, 1000, 10,000, and 100,000 are colour coded on a categorical scale from dark blue to light turquoise (see legend). The number 0 is in red to highlight that we have manually coded its $\log_{10}$ value as –0.1 to visualise it on a $\log_{10}$ scale, as the logarithm of 0 is not defined. The numbers 0, 1, 2, 3, 100, 1000, 1,000,000, and 1,000,000,000 are circled and labelled.

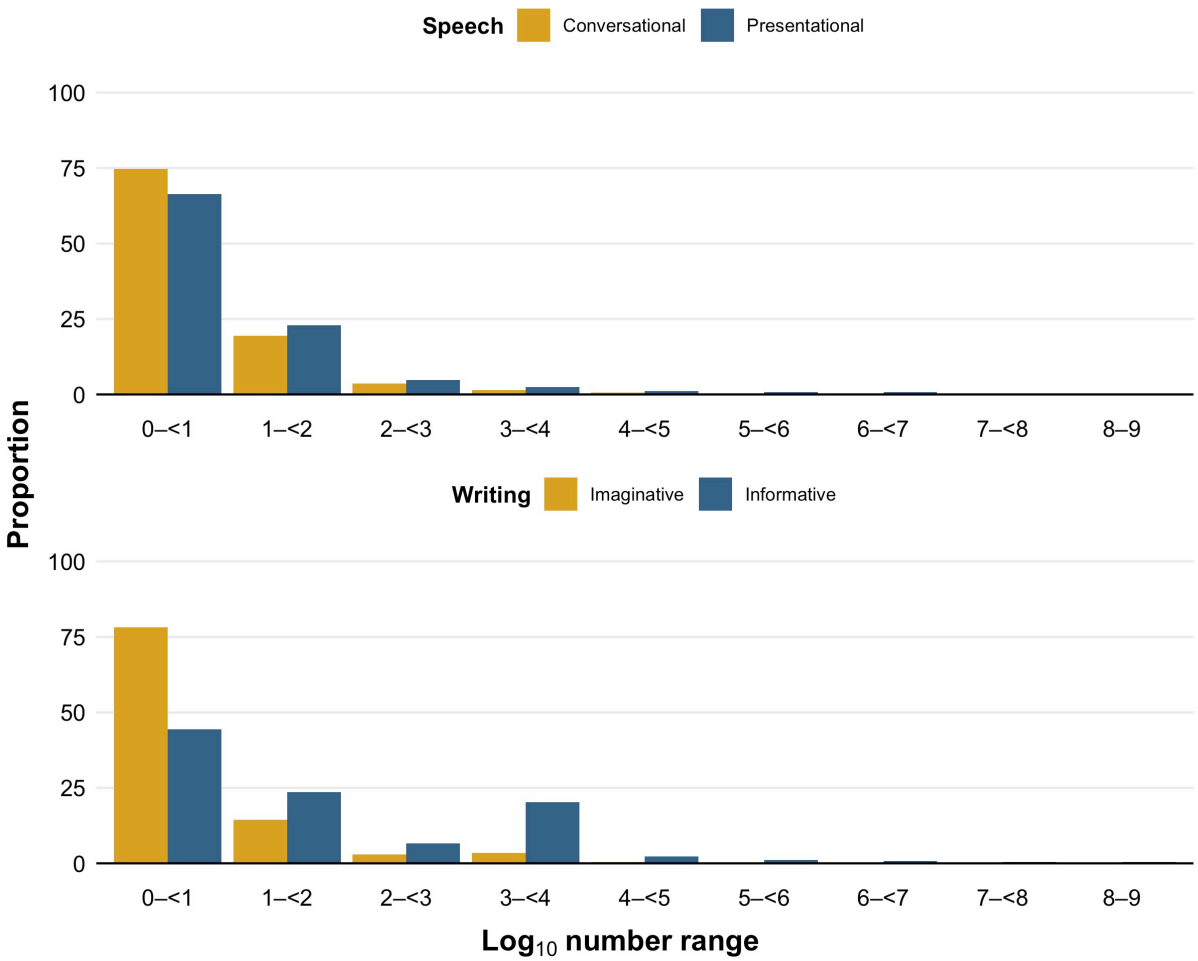


Figure 2. Proportions of number tokens in different $\log_{10}$ number ranges in the spoken and written subcorpora. Proportions are out of all number tokens in each subcorpus.

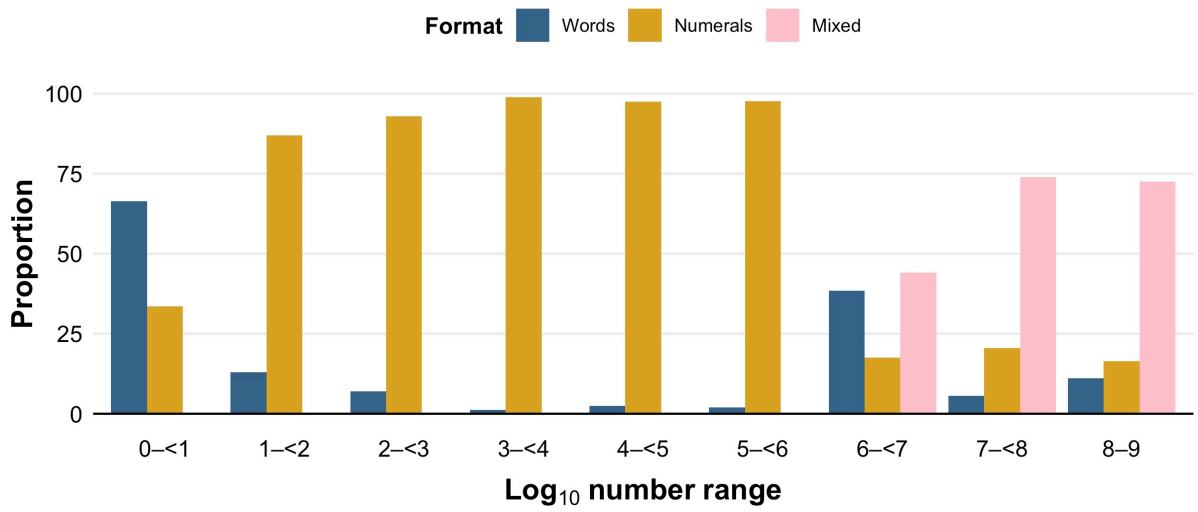


Figure 3. Proportions of number tokens in each $\log_{10}$ number range that were written in different representational formats. Proportions are out of all number tokens in each $\log_{10}$ number range.

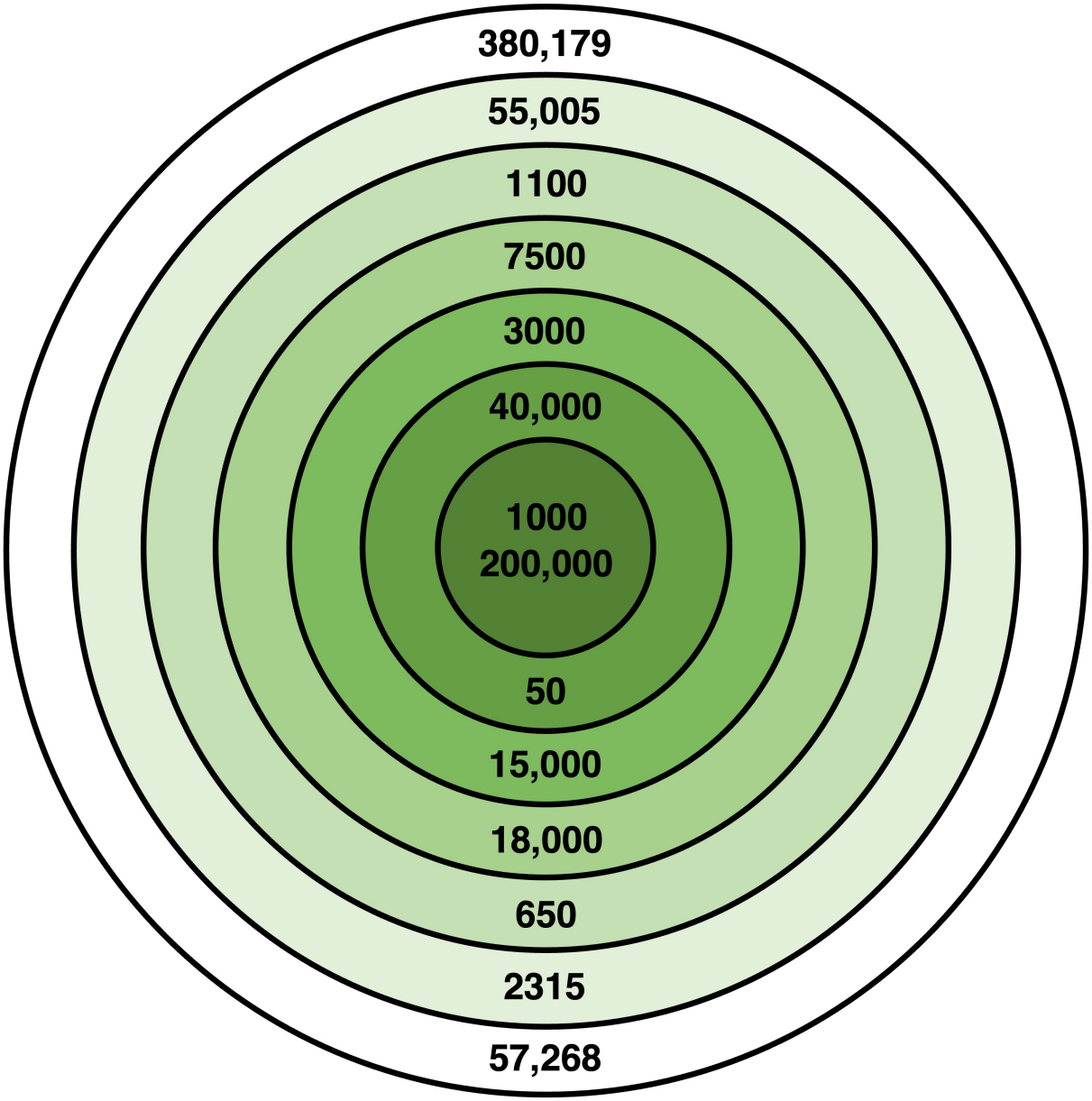


Figure 4. Roundness as a radial category, where green indicates roundness and white indicates the absence of roundness. The most prototypical or 'roundest' numbers are in the centre circle (i.e., numbers with all six roundness properties: 10-ness, 2-ness, 2.5-ness, 5-ness, being a multiple of ten, and being a multiple of five). Progressing outward through the inner rings are less prototypical or 'less round' numbers with five, four, three, two, or one roundness properties. The outer ring shows numbers that have no roundness properties.

6. PAPER 4

Woodin, G., Winter, B. & Perlman, M. (in prep). The iconic impetus of language: From sensorimotor simulation to the lexicon.

- **Conceptualization:** Greg Woodin, Marcus Perlman.
- **Data curation:** NA.
- **Formal analysis:** NA.
- **Investigation:** NA.
- **Methodology:** NA.
- **Project administration:** Greg Woodin, Marcus Perlman.
- **Supervision:** Bodo Winter, Marcus Perlman.
- **Validation:** Bodo Winter, Marcus Perlman.
- **Visualization:** Greg Woodin.
- **Writing – original draft:** Greg Woodin.
- **Writing – review & editing:** Greg Woodin, Bodo Winter, Marcus Perlman.

The iconic impetus of language: From sensorimotor simulation to the lexicon

Greg Woodin, Bodo Winter, and Marcus Perlman

English Language and Linguistics, University of Birmingham

Abstract

Iconicity refers to the resemblance between a communicative form and its meaning. Historically, linguists have argued that arbitrariness, rather than iconicity, is a key design feature of language. In this paper, we flip this long-received paradigm on its head, placing iconicity at the forefront of our understanding of language evolution. Focusing on spoken languages, we propose a model of iconicity that operates across three timescales: psychological, interactional, and cultural-historical (glossogenetic). In this model, sensorimotor simulations drive the use of ad-hoc iconic forms and modulations in social interactions. Communicatively useful iconic forms are repeated and shared via cultural transmission (Kirby, 2001; Kirby & Hurford, 2002; Scott-Phillips & Kirby, 2010), which gradually leads to their conventionalization into the lexicon. Over time, iconic forms tend to (but do not always) lose their iconicity through processes of language change, such as phonetic reduction, in response to pressures such as efficiency and discriminability. In response to this deiconization, sensorimotor simulations drive the continual use of ad-hoc iconic forms and modulations in social interactions, which, through conventionalization, replenish levels of iconicity in the lexicon. Because the deiconizing processes involved in conventionalization are motivated and hence non-random, we argue that 'arbitrariness' is

misused as an antonym for iconicity. We also argue that conventionalization is not actively biased against iconicity, but is rather iconicity blind – it does not 'care' about iconicity, and so is likely (but not certain) to erode it. Based on these ideas, we argue that iconicity is a design feature of language, whereas there is no similar evolutionary pressure toward deiconization. The primacy of iconicity in language evolution is what we call the iconic impetus of language.

Key words: iconicity; simulation; embodied cognition; gesture; evolution

1. Introduction

Iconicity is the resemblance between a communicative form and its meaning (Peirce, 1931). In spoken languages, onomatopoeias are iconic by imitating the sounds to which they refer (Bredin, 1996; Rhodes, 1994; Sarvasy, 2016), such as the interjection *ka-boom!*, which represents an explosion (e.g., Taylor, 2007), or *cuckoo*, which mimics the call of the eponymous bird (Marttila, 2011). There are also many examples of iconicity in signed languages, such as the American Sign Language sign *drink*, in which the signer mimes drinking from a cup by making a C-shape with their dominant hand, and tipping it toward their mouth.

In spite of these obvious examples of iconicity, the long received view in linguistics holds that iconicity is the exception to the rule of *arbitrariness* (e.g., Hockett, 1960; Levelt et al., 1999; Newmeyer, 1992; Pinker & Bloom, 1990), which is often defined in opposition to iconicity as the lack of resemblance between a form and its meaning (e.g., Gasser, 2004; Lockwood & Dingemanse, 2015; Sidhu et al., 2021). If a communicative form does not resemble its meaning, the choice to use this form is assumed to be arbitrary – random and unmotivated, established solely by language-specific convention. Ferdinand de Saussure (1959) illustrated this point by demonstrating that the same concept is often denoted by different words in different spoken languages (see also Locke, 1689/1979). For example, the English word *cup* and the French word *tasse* have the same meaning while not sharing a single phoneme.

While Saussure acknowledged the existence of iconicity in spoken languages, he downplayed its significance. Throughout the history of linguistics, the position that arbitrariness is primary has persisted (e.g., Hockett, 1960; Levelt et al., 1999; Newmeyer,

1992; Pinker & Bloom, 1990), with Hockett (1960) even elevating arbitrariness to a defining property that separates human language from animal communication systems. In cognitive science, arbitrariness has been emphasised as an evolutionary pressure on the vocabularies of languages (Christiansen & Chater, 2016; Christiansen & Monaghan, 2016; Gasser, 2004; Monaghan et al., 2011), and the prevalence of iconicity in signed languages is partly responsible for the longstanding dismissal of these languages as mere pantomimes or elaborate manual gesture systems (see Goldin-Meadow & Brentari, 2017). More recently, however, many researchers have argued that, alongside arbitrariness, iconicity is also a defining feature of linguistic systems (e.g., Dingemanse et al., 2015; Lockwood & Dingemanse, 2015; Perniss et al., 2010; Perniss & Vigliocco, 2014; Vinson et al., 2021).

In this paper, we turn the Saussurean principle of arbitrariness on its head, instead placing iconicity at the forefront of our understanding of languages and their evolution. We propose a model of iconicity in languages that operates across three timescales: psychological, interactional, and cultural-historical (glossogenetic). At the psychological level, we propose that sensorimotor simulations drive the use of ad-hoc iconic depictions of sensorimotor content in interactions. These iconic depictions include the creation of novel iconic signals, such as manual gestures and non-linguistic vocalizations, and the modulation of existing signals in ways that enhance their iconicity (see Perlman & Woodin, 2021), such as with iconic prosody in spoken language (e.g., Perlman, Clark, et al., 2015). When these iconic forms are repeated by language users and spread via cultural transmission (Kirby, 2001; Kirby & Hurford, 2002; Scott-Phillips & Kirby, 2010), they are gradually conventionalized and gain membership into the lexicon. Across cultural-historical time, conventionalized iconic forms tend to gradually lose their iconicity – what Flaksman (2015,

2017, 2020) calls *deiconization* – as they evolve to be more efficient, discriminable, and so on. In reaction to these various deiconizing processes, sensorimotor simulations drive the continual use of ad-hoc iconic forms and modulations in interactions, which, via conventionalization, gradually replenish levels of iconicity in the lexicon.

In this model of language evolution, communication is iconic from the start. Then, via conventionalization, other language processes lead to deiconization. These processes are not random or unmotivated; hence we argue that 'arbitrariness' is misused as an antonym for iconicity. In this way, our model builds on previous writing that has argued that arbitrariness is not a design feature of language (Baron, 1981; Deuchar, 1984; Johnston & Schembri, 2007). Our model also argues that conventionalization is not actively biased against iconicity, but is merely *iconicity blind* – it does not 'care' about iconicity, and so is likely (but not certain) to erode it. Based on these ideas, we argue that iconicity is a design feature of language, but that there is no similar evolutionary pressure toward deiconization. We call the primacy of iconicity in language evolution the *iconic impetus of language.*

In our exposition of this model, we focus our attention on spoken languages. We do this because attempting to do justice to both spoken and signed languages would be impossible without sacrificing depth. Also, signed languages have not been researched as extensively in relation to many of the topics we discuss, such as sensorimotor simulation. While we do sometimes refer to research on signed languages, it is always to make some comparison to spoken languages. We believe the language model we propose logically applies also to signed languages, but more research and a whole paper-length treatment would be needed to convincingly make this argument.

The overall structure of the paper is as follows. In section 2, we describe the iconic strategies that speakers of spoken languages use to depict sensorimotor content in their utterances. In section 3, we survey the literature on the role of sensorimotor simulation in language comprehension and production to argue that sensorimotor simulation is the mechanism that gives rise to iconic forms of expression. In section 4, we describe how sensorimotor simulation can lead to the production of ad-hoc iconic signals, how and why these signals become conventionalized into iconic words, and how the affordances of the vocal channel lead to different concentrations of iconicity for different meanings in spoken languages compared to signed languages. In section 5, we discuss the process by which iconic words tend to lose their iconicity as they undergo conventionalization, while system-wide iconicity is maintained via the continual coinage of iconic forms, which is driven by sensorimotor simulation. In section 6, we summarise the model and argue that iconicity constitutes an evolutionary pressure, whereas deiconization does not.

2. Iconicity in Multimodal Communication

In this section, we explicate the different iconic forms that are used by speakers of spoken languages. We explore these iconic forms across modalities, because speakers do not just utilise the voice (auditory) but also the hands with the performance of manual gestures (visuospatial). The four main categories of iconicity we discuss are iconic words, manual gestures, non-linguistic vocalizations, and prosody. These are not an exhaustive list of all types of iconicity, as there are other iconic strategies available for language users to exploit, such as facial expressions (e.g., Nölle, Chen, et al., 2021; Nölle, Garrod, et al., 2021), but these are beyond the scope of the present paper. We finish the section by demonstrating

how these iconic strategies can and often are combined in speakers' utterances, being used opportunistically by speakers for the purpose of creating depictive performances of sensorimotor content.

2.1. Iconic Words

In the introduction, we discussed onomatopoeias, which are iconic words that imitate the sounds to which they refer (Bredin, 1996; Rhodes, 1994; Sarvasy, 2016). For instance, *ka-boom!* depicts an explosion (e.g., Taylor, 2007), and *cuckoo* depicts the sound of the eponymous bird (Marttila, 2011). Other examples include *bang, peep, chirp,* and *hiss,* which all denote different sounds.

In addition to these saliently iconic onomatopoeias, some languages have entire grammatical categories of iconic words called *ideophones,* also known as mimetics or expressives in different descriptive traditions (Akita & Pardeshi, 2019; Dingemanse, 2018; Voeltz & Kilian-Hatz, 2001). These words go beyond the depiction of just sound to depict a range of sensory modalities (Dingemanse, 2012), including motion, size, brightness, and touch. Examples of ideophones include the Japanese words *gorogoro* and *korokoro*, which respectively mean 'a heavy/light object rolling repeatedly' (Kita, 1997; Perniss et al., 2010). In these words, the contrast in voicing between /g/ and /k/ depicts the weight of the object, and the reduplicated syllables depict the repetition of the rolling movement. Dingemanse (2012) also cites examples of ideophones from other languages, including the Somali word *juluq,* whose sound resembles the noise associated with its meaning 'to gulp down (something solid) without chewing'.

Iconicity is, however, not confined to explicitly imitative words such as onomatopoeias

and ideophones. *Sound symbolism* refers to the resemblance between certain sounds in words and their meanings (Knoeferle et al., 2017), which can be found in the more prosaic part of the lexicon (Blasi et al., 2016; Haynie et al., 2014; Johansson et al., 2019, 2020; Johansson & Zlatev, 2013; Joo, 2020; Wichmann et al., 2010; Winter et al., 2022). For example, statistical analyses have shown that sounds with high frequency components such as /i/, /ɪ/ and /t/ occur more frequently in small-size adjectives, such as *teeny, itty, meagre, miniscule, little,* and *pee-wee*, whereas low back vowels occur more frequently in large-size adjectives like *large, colossal, hulking, whopping,* and *gargantuan* (Winter & Perlman, 2021). This pattern is theorised to be motivated by the fact that small objects, instruments, and animals tend to produce higher frequency sounds than large ones . Sound-symbolic correspondences have also been reported in the semantic field of shape, with many words for round objects having rounded vowels (e.g., /u/ in *hoop* /huːp/), voiced consonants (e.g., /g/ in *globe* /gləʊb/), and bilabial consonants (e.g., /b/ in *balloon* /bəluːn/), in contrast to words for spiky objects, such as *spike*, *fork*, *cactus*, and *shrapnel* (Blasi et al., 2016; Joo, 2020; Sidhu et al., 2021).

2.2. Iconic Gestures

When people talk, they often also gesture with their hands – manual gestures, which we refer to simply as *gestures*. It is now recognized that many and maybe even most utterances involve gesture alongside speech (Enfield, 2009; Goldin-Meadow, 2005a; Kendon, 2004, 2014; McNeill, 1992). Many of these gestures resemble their meaning and so can be described as iconic. For example, speakers may pinch their index finger and thumb together to denote smallness, with the small space between the finger and thumb depicting small

size, whereas speakers may move their open hands apart from each other to depict largeness (Woodin et al., 2020). Furthermore, a speaker might depict *smoking* by moving two fingers back and forth in front of their mouth as if smoking a cigarette (see Kendon, 1988).

2.3. Iconic Vocalizations

Another strategy through which speakers can express iconicity is non-linguistic *vocalizations* (hereafter: vocalizations), including verbal or vocal imitation and mimicry, which involve depicting a sound using imitative vocal noises (Edmiston et al., 2018; Lemaitre et al., 2016; Sarvasy, 2016). An example of a vocalization would be the imitation of a car, either with an unhealthy engine that makes "a worry noise, kind of a "(low growing noise)", or with a healthy engine that "goes "[gargling sound]," has a real deep lion-purring sound to it" (H. H. Clark & Gerrig, 1990, p. 332). As another example, one participant in a study conducted by Perlman and Lupyan (2018) depicted the concept *water* by imitating the sound of dripping water with a series of tongue clicks. A particularly productive domain in which vocalizations are used is in the imitation of birdsong, which includes combinations of grunts, whistling, humming, and clicking (Sarvasy, 2016).

Vocalizations are similar to onomatopoeias and ideophones but are less conventionalized or entirely unconventionalized. Indeed, vocalizations may flout the phonotactic constraints of the lexicon, with the 'wildest' (Rhodes, 1994) ones resisting decompositional analysis into a sequence of phonemes, and being difficult to transcribe orthographically. Vocalizations also overlap with *quotations*, which involve directly reporting the speech of another person (H. H. Clark, 2016; H. H. Clark & Gerrig, 1990). For instance,

H. H. Clark and Gerrig (1990) give the example of “she says ‘well I’d like to buy an ant’”, where the second part of the sentence is a quotation. Vocalization can be interpreted as a kind of quotation where the quotation is not of conventionalized words. For example, Fox Tree and Tomlinson (2007, p. 100) give the example of “He was like opening his drawers but they were like stuck so he was like, ‘‘Raah aah’”. In this example, the quotation “Raah aah” is also a vocalization.

2.4. Iconic Prosody

Whereas iconic words, gestures, and vocalizations can all be produced on their own, *iconic prosody* modulates the production of existing words by altering their speed, rhythm, pitch, or timbre. For example, English speakers may articulate the word *long* as *loooong*, using vowel lengthening to depict the length of time referred to (e.g., Guerrini, 2020). People also speak more quickly when describing fast motion (e.g., *She dashes past the buildings*) compared with slow motion (e.g., *She labors slowly*) (Perlman, 2010), and participants describing the movement of a dot on a screen utter sentences with a longer duration when the dot moves slowly than when it moves quickly (Shintel et al., 2006). Other examples of prosodic modulation involve the depiction of verticality via pitch: participants spontaneously raise their vocal pitch when describing an upward-moving dot and lower their pitch when describing a downward-moving dot (Shintel et al., 2006). Iconic prosody is similar to iconic vocalization in that it uses qualities of the vocal signal to depict meanings, with the difference being that, while iconic vocalizations are used in place of conventional words, iconic prosody is “parasitic on the phonological form of an utterance” (Perlman & Gibbs, 2013, p. 23).

2.5. Combining Iconic Strategies

Iconic words, gestures, vocalizations, and prosody are all strategies through which speakers can depict meanings. As we will see, these strategies are often combined to create depictive performances. The use of these strategies may be redundant (depicting the same information) or complementary (depicting different information). In this way, speakers may opportunistically use all the iconic tools at their disposal for the purpose of depictive communication (see Perlman & Gibbs, 2013).

Figure 1 shows an example of how iconic strategies can be utilised simultaneously, with each strategy redundantly depicting the same information.[1] In this example, TV personality and journalist Ezra Klein uses the informal size adjective *teeny* /tiːni/, which is iconic as it uses the high front vowel /i/, which has a high formant frequency, to depict smallness (see Ohala, 1994). While uttering the word *teeny*, Klein raises his vocal pitch to prosodically exaggerate this word's high frequency vowel (see Perlman, Clark, et al., 2015). Simultaneously, Klein performs a pinch gesture, with the small space between his forefinger and thumb depicting smallness (Hassemer & Winter, 2016, 2018; Kendon, 2004; Woodin et al., 2020). In sum, Klein uses an iconic word, an iconic gesture, and iconic prosody in the same utterance, and these strategies are unified in that they appear to depict the same semantic content.

[1] Klein also uses the iconic word *itty-bitty* (see Winter & Perlman, 2021), referring to "teeny tiny, itty-bitty really hard to see stuff".

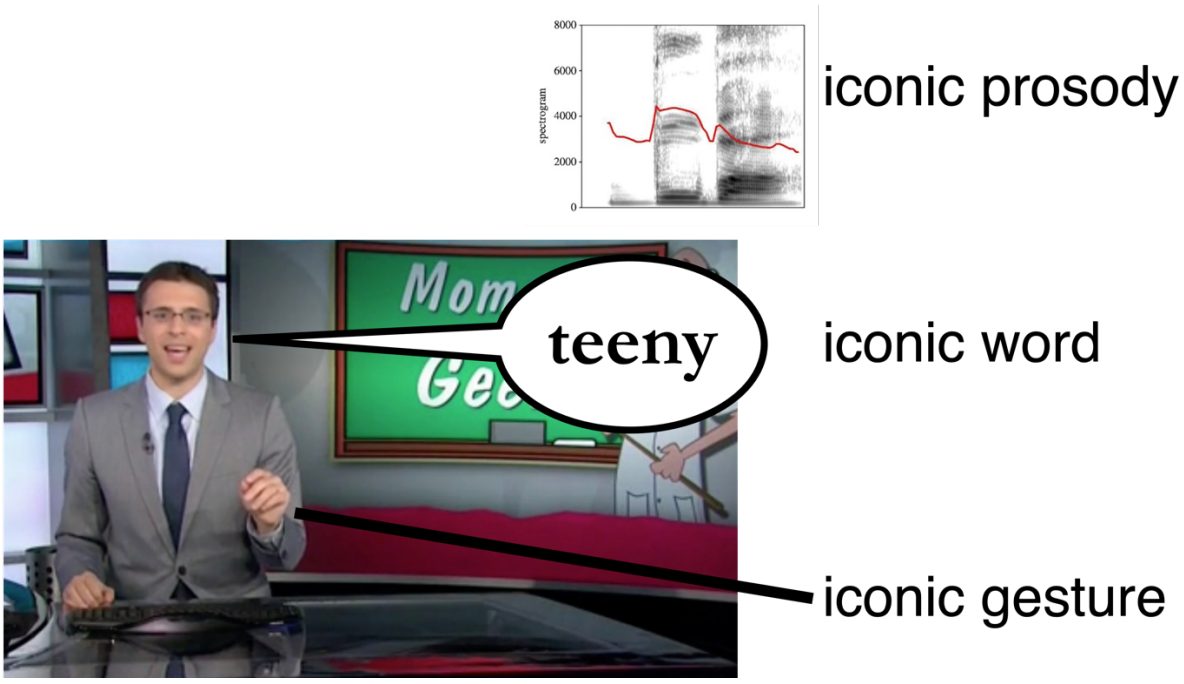


Figure 1. TV news personality and journalist Ezra Klein using the iconic word *teeny* while accentuating its high frequency /i/ vowel by raising his vocal pitch (iconic prosody), and concurrently pinching his index finger and thumb together (iconic gesture). The image of Klein has been annotated with a black line and speech bubble for illustrative purposes. Source: MSNBC, The Rachel Maddow Show, accessed via the TV News Archive: https://tinyurl.com/r3tdhrs5.

Figure 2 shows an example from an interview with a man named George Lindell, who is describing a car crash in which he was recently involved (see Perlman & Cain, 2014 for a discussion of this interview).[2] Unlike Figure 1, where the iconic strategies are utilised redundantly, Lindell uses different iconic strategies to depict complementary information. In

[2] At other points in the interview, Lindell's speech rate increases to convey the speed at which the events around him were unfolding (see Perlman, 2010). Note also Lindell's facial expression, which may imitate his facial expression during the crash.

particular, Lindell uses the semi-conventionalized iconic vocalization *BAM!* /bæm/ to depict the noise and impact of the crash, while abruptly raising the volume of his voice (iconic prosody) to depict the loudness and jarring, sudden onset of the collision. At the same time, Lindell performs an iconic gesture, holding his two clenched fists out in front of him as if gripping the car's steering wheel. By combining vocalization, prosody, and gesture in complementary ways, Lindell conjures up a vivid image of what it would have been like to experience the crash from his perspective.

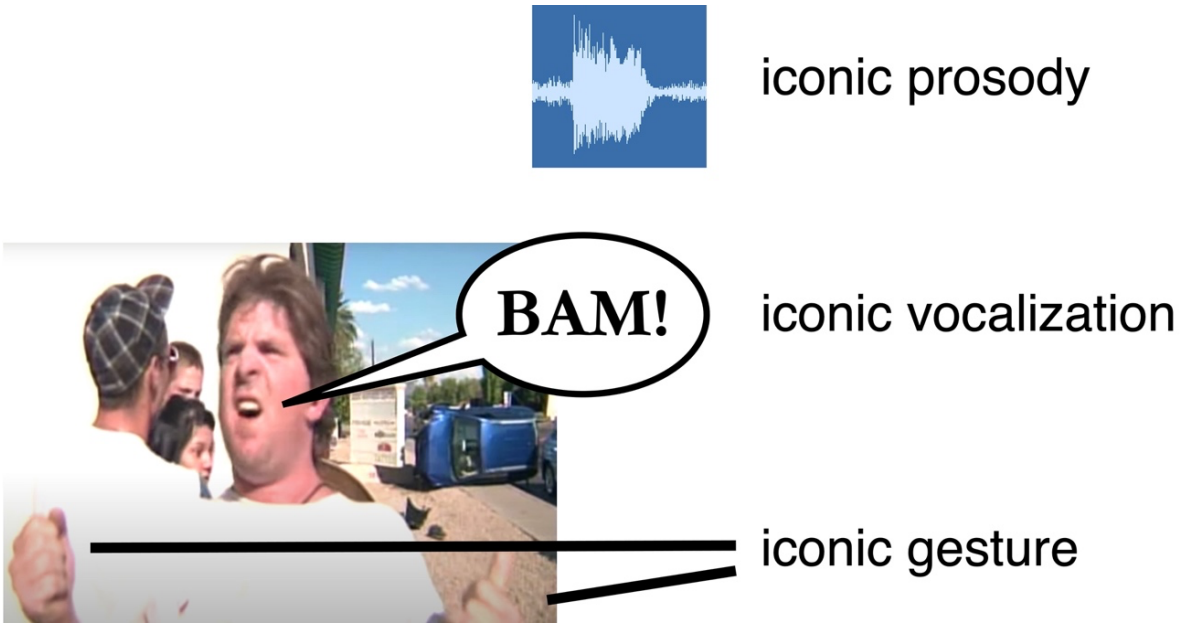


Figure 2. George Lindell describes the impact of a car crash in which he was involved using the iconic vocalization *BAM!* /bæm/, raising the volume of his voice (iconic prosody), and pretending to grip the invisible steering wheel of a car (iconic gesture). The image of Lindell has been annotated with black lines and a speech bubble for illustrative purposes. Source: Fox 10 Phoenix, accessed via YouTube: https://tinyurl.com/yc6u5pp5.

Importantly, all these forms of iconicity have the same basic function: to create a depictive performance that cues the audience to picture the scene that is being described. When communicating iconically, people use whatever resources they have to hand to do the job of depiction.

3. Sensorimotor Simulation and Iconicity

What is the psychological mechanism that gives rise to all these iconic strategies? In this section, we argue that the answer to this question is *sensorimotor simulation,* which is the mental reexperiencing of previously experienced perceptions and actions, recruiting the same areas of the brain that are activated during perception and action, but to a lesser degree (e.g., Barsalou, 1999, 2007; Decety & Jeannerod, 1995). Sensorimotor simulation is generally seen as being related to our capacity to create mental imagery (e.g., Kosslyn, 1980), with some scholars viewing imagery as a more explicit and conscious form of simulation (e.g., Bergen, 2012). In contrast to imagery, simulations are thought to happen automatically, such that, for instance, understanding a word like *kick* automatically activates leg-related areas of the brain (Dreyer et al., 2015; Hauk et al., 2004; Pulvermüller, 1999; Pulvermüller et al., 1999). Sensorimotor simulations are productive: people can simulate an infinite number of experiences by combining previously experienced perceptions and actions, a capacity which is fundamental to creativity (Barsalou, 1999; Barsalou & Prinz, 1997)

We begin this section by discussing work on the role of sensorimotor simulation in language comprehension, which concerns the idea that understanding words for concrete concepts involves the activation of sensorimotor brain areas associated with the direct

experience of the denoted concepts, as if mentally re-enacting perceptions and actions in a weaker form. We then discuss research showing that the comprehension of iconic words activates sensorimotor simulations more strongly than that of non-iconic words. The tight connection between iconicity and sensorimotor simulation in language comprehension leads us to explore the idea that sensorimotor simulation is the psychological mechanism that underlies the *production* of iconic forms, including gestures, vocalizations, and prosody.

3.1. Simulation in Language Comprehension

Behavioural and neuroscientific evidence suggests that simulation is involved in language comprehension. Early evidence suggesting that motor simulation plays a role in the processing of sentences about actions came from the Action-sentence Compatibility Effect (ACE) (Glenberg & Gallese, 2012): participants were found to perform movements with a backward motion faster after reading a sentence implying movement toward themselves (e.g., *Art handed you the pen*) than away from themselves (e.g., *You handed Art the pen*). While this particular finding has failed to replicate (Morey et al., 2021; Papesh, 2015), other studies corroborate the role of motor simulation in language comprehension. For example, fMRI scans have shown that when participants read English words such as *lick*, *pick*, and *kick*, their motor cortex exhibits patterns of activation (inferred from blood flow to different brain areas) similar to those observed in localizer scans when participants move the corresponding parts of the body (Hauk et al., 2004). Other studies using MEG and EEG recordings have also shown that action-specific areas of the brain are engaged during the comprehension of action-related language (Moseley et al., 2013; Popp et al., 2016).

While early evidence for motor simulation was open to the criticism that activation of the motor system may arise late in processing as an epiphenomenon (e.g., Mahon & Caramazza, 2008), increasing numbers of studies find that sensorimotor simulations arise early and are functionally relevant to language comprehension. For instance, Moseley et al. (2013) used imaging approaches with high temporal resolution (MEG and EEG) to demonstrate that the motor cortex becomes engaged as early as 150ms after the onset of a word. Studies have also indicated that interfering with motor cortex activity using transcranial magnetic stimulation (TMS) slows down the comprehension of hand-action words (Vukovic et al., 2017) and the learning of novel action verbs (Vukovic & Shtyrov, 2019).

More evidence for the functional relevance of sensorimotor simulation comes from lesion studies. For example, Trumpp et al. (2013) reported that a patient with a focal lesion to his auditory cortex performed more poorly in a lexical decision task when categorizing sound-related words (e.g., *bell*) compared to sound-unrelated words (e.g., *armchair*). Bak and Hodges (2003) also found that language comprehension was impaired more severely for action verbs than nouns in patients whose motor cortex had degenerated. Lastly, a study by Dreyer et al. (2015) involving two patients with different lesions to their motor cortex showed that the comprehension of action-related words referring to manual tools (e.g., *hammer*) was most affected when brain damage was limited to motor areas of the brain associated with hand movements. This research indicates that impeding sensorimotor simulation can interfere with language comprehension, pointing to the functional role of simulation in this process.

3.2. Simulation in the Comprehension of Iconicity

The dominant position in embodied cognition has undergone a shift from the view that sensorimotor simulation is always engaged to the same degree during language comprehension to a more moderate position in which neural circuits for perception and action are engaged more or less in a task-, context-, and concept-dependent manner (Barsalou, 2016; Pecher, 2018; Zwaan, 2016). For example, there is less structured sensorimotor information to simulate for words with less concrete, more abstract meanings, such as *freedom* and *democracy*, which has influenced hybrid positions where abstract concepts are seen as being supported by representational mechanisms other than simulation, such as those connected to emotional valence (Kousta et al., 2011), language (Borghi et al., 2017, 2019; Vigliocco et al., 2009), and sociality (Borghi et al., 2017, 2019). One implication of this theoretical shift is that sensorimotor simulation may be stronger and therefore play a more prominent role in the comprehension of certain communicative forms compared to others.

One candidate for communicative forms whose comprehension cues stronger simulations is iconic words. For instance, research shows that processing Japanese ideophones activates sensorimotor areas of the brain more strongly compared to non-ideophones (Kanero et al., 2014) and non-words (Osaka, 2009, 2011; Osaka et al., 2004; Osaka & Osaka, 2009). Revill et al. also reported stronger activation in intraparietal areas of the brain implicated in cross-modal and synaesthetic processing for iconic versus non-iconic words with a multi-language stimulus set presented to native English speakers. Moreover, an fMRI study by Hashimoto et al. (2006) found that hearing Japanese ideophones depicting animal sounds increased activation in brain areas associated with

processing environmental as opposed to linguistic sounds more strongly than non-mimetic animal names.

As well as iconic words, comprehending iconic prosody may involve sensorimotor simulation, which may influence people's subsequent behaviour. In a study conducted by Shintel and Nusbaum (2007), participants listened to a sentence (e.g., *The horse is brown*) and then stated whether a presented picture featured the referent of the sentence. Participants were quicker to identify pictures of stationary referents after listening to the slower sentences, whereas they were quicker to identify pictures of moving referents after listening to the faster sentences. This result suggests that iconic prosody affected participants' simulations of the scene described by the sentence, which subsequently influenced their performance on the identification task.

3.3. Simulation in the Production of Iconic Gestures

Considering the close relationship between simulation and iconicity in language comprehension, it is possible that simulation is also involved in the production of iconic forms. In their Gesture as Simulated Action framework, Hostetter and Alibali (2008, 2018) make the case for the role of sensorimotor simulation in the production of iconic gestures, arguing that, when speakers simulate motoric and visuospatial information while talking, there is increased activation of their premotor cortex (Prinz, 1997; Rizzolatti et al., 1996). Sometimes this activation spreads to the primary motor cortex, causing the performance of iconic gestures. Thus, the GSA framework seems to describe an automatic 'spill-over' of motor activation into iconicity in the manual channel. The resultant gestures depict the

content of the underlying simulation, so, for example, a person simulating the action of *smoking* might produce a gesture that imitates this action (see Kendon, 1988).

The GSA framework predicts that speakers will be more likely to gesture when sensorimotor simulation is strongly engaged during communication. Indirect evidence supports this prediction. For example, speakers who watched a cartoon vignette and listened to a description of the vignette were more likely to gesture when retelling the vignette than speakers who only listened to the description (Hostetter & Skirving, 2011). Speakers who watched the cartoon may have had a stronger sensorimotor simulation to draw from when speaking, which might explain why they gestured more frequently. In another study, speakers whose spatial skills were superior to their verbal skills gestured more often than other speakers, perhaps because they relied more on visuospatial simulations when communicating (Hostetter & Alibali, 2007). Research has additionally shown that people gesture more often when they discuss concrete concepts – which are more directly connected to sensorimotor information than abstract ones – especially when the concrete concepts are encoded in verbs denoting actions (Hadar & Krauss, 1999). Together, these various studies suggest that gesture production may increase when communication more strongly involves simulation, consistent with the idea of a direct link between simulation and gesture production.

3.4. Simulation in the Production of Iconic Prosody

The GSA framework (Hostetter & Alibali, 2018, 2008) is about gestures, but, as we saw in Figures 1 and 2, iconic gestures often occur with iconic prosody, such as when George Lindell raises the volume of his voice while gesturing as if gripping a car's steering wheel.

As we have argued, iconic gesture and iconic prosody are used for the same purpose: to create a depictive performance that triggers the audience to imagine the described experience. Indeed, iconic prosody is sometimes described as *vocal gesture* due to its gesture-like role in communication (N. Clark et al., 2013; Okrent, 2002; Perlman, 2010). If iconic gesture and prosody have the same function, there may be a unified psychological mechanism that underlies both these iconic strategies, rather than there being a special mechanism for gestures. Thus, in addition to gestures, we posit that sensorimotor simulations may underlie iconic prosody, in line with the argument made by Perlman and Gibbs (2013).

Initial evidence that can be interpreted as support for this position comes from a language production task by Perlman (2010) in which participants watched videos that depicted either fast or slow paced events. When participants subsequently described the videos, their vocal speed correlated with the speed of the events depicted in the videos. Therefore, it is possible that participants who had watched a faster paced video later simulated the fast events depicted in the video, which caused them to increase their vocal speed to reflect the speed of their simulation. Another study has shown that participants subsequently describing the movement of a dot on a screen utter sentences of a shorter (longer) duration when the dot moved quickly (slowly), and lower (raise) their vocal pitch when the dot moved downward (upward) (Shintel et al., 2006), prosodically depicting the content of their sensorimotor simulation. These studies support the idea that, like gesture, sensorimotor simulation may underlie the production of iconic prosodic modulations, but more research is needed to strengthen this claim.

3.5. Simulation in the Production of Iconic Vocalizations

The same arguments made above for iconic prosody can also be made for iconic vocalizations. Namely, iconic vocalizations are produced for the purpose of depiction, a purpose that these vocalizations share with iconic gesture and iconic prosody. We saw an example of this shared function in Figure 2, where George Lindell used the iconic vocalization *BAM!* while raising the volume of his voice (iconic prosody) and gesturing as if gripping the steering wheel of a car (iconic gesture). In this example, it seems evident that Lindell is imagining the car crash he experienced and is depicting this scene to help his audience imagine it also. In other words, Lindell is performing the contents of his sensorimotor simulation using vocalization, gesture, and prosody. Thus, if sensorimotor simulation is the psychological mechanism that gives rise to iconic gesture and iconic prosody, it may also underlie iconic vocalization (see Perlman & Gibbs, 2013). If true, sensorimotor simulation would be the motivating force behind the use of, for example, vocalizations imitating different engines (H. H. Clark & Gerrig, 1990) and ‘‘Raah aah’’ to mimic a frustrated human (Fox Tree & Tomlinson, 2007), tongue clicks to depict dripping water (Perlman & Lupyan, 2018), and grunts, whistles, hums, and clicks to imitate birdsong (Sarvasy, 2016). As with iconic prosody, additional research is needed to support the existence of a link between sensorimotor simulation and the production of iconic vocalizations.

4. From Sensorimotor Simulation to the Lexicon

Based on research into the role of sensorimotor simulation in spoken communication, we consider exactly *how* simulation may result in iconic communication in production. We

discuss the process at the neural-cognitive level, where simulation results in iconic communication via 'spill-over' from premotor to motor areas of the brain, and at the functional, pragmatic level, where iconicity arises when there is a situational need to communicate sensorimotor detail. We then explore the idea that the resultant iconic gestures, vocalizations, and prosodic modulations in social interactions may enter the lexicon via conventionalization as they are repeated and shared in these interactions. We argue that some communicative forms spread by cultural transmission (Kirby, 2001; Kirby & Hurford, 2002; Scott-Phillips & Kirby, 2010) because they have some kind of utility, and we describe the various utilities of *iconic* forms that may lead them to be repeated. We then report evidence that the lexicons of some sign systems formed from the conventionalization of iconic gestures, and argue that spoken lexicons could form partly from the conventionalization of iconic vocalizations and prosodic modulations. To finish, we discuss the affordances of the spoken modality for expressing qualitatively different sensorimotor simulations from the signed modality, and describe how these affordances affect concentrations of iconicity for different meanings in their respective lexicons.

4.1. How Simulation Gives Rise to Iconicity

At the neural-cognitive level, sensorimotor simulation may result in iconicity via activation spreading from premotor to motor areas of the brain, analogous to the process described by Hostetter and Alibali's (2008, 2018) GSA framework. The GSA framework makes its hypothesis for iconic gestures, but we extend this prediction to also include iconic vocalizations and prosody, similar to Perlman and Gibbs (2013). Specifying this process precisely is beyond the scope of this paper, but suffice it to say that gesture, vocalization,

and prosody are all actions (manual and vocal) and so could be produced via activation of the motor cortex.

At the functional, pragmatic level, iconicity may emerge when there is a contextual need to communicate sensorimotor detail, like when speakers do not share a common language, or when playing a game like charades, where speech is prohibited. In support of this idea, Lu and Goldin-Meadow (2018) found that signers used more iconic gestures to communicate differences between objects (e.g., by gesturing about their shape) when these objects were difficult to differentiate using lexical signs. In another study, Perniss et al. (2017) found that parents who imagined playing with their children injected more iconicity into their signing when the referents of the signs were not present, seemingly in an attempt to reduce referential ambiguity.

These two levels of analysis are complementary. For example, Hostetter and Alibali (2008, p. 503) discuss the idea of a *gesture threshold,* which is defined as "the level of activation beyond which a speaker cannot inhibit the expression of simulated actions as gestures". This gesture threshold can be influenced by aspects of the communicative context. For example, studies have shown that teachers can increase their gesture rate when asked to (Hostetter et al., 2006), and when they think what they are describing is difficult to understand (Alibali & Nathan, 2007). Hostetter and Alibali (2008) interpret these findings as suggesting that the teachers lowered their gesture threshold, allowing themselves to gesture more, to meet the needs of the respective interactions. Hence, according to the GSA framework, whether a speaker gestures depends on the amount of motor activation (the neural-cognitive level) and the level of the gesture threshold, which is determined partly by aspects of the speech situation (the functional, pragmatic level).

As in this paper we are broadening our focus to include other forms of iconicity in spoken communication, we reframe the gesture threshold as the *iconicity threshold,* which is the level of motor activation beyond which speakers will produce an iconic gesture, vocalization, or prosodic modulation. Hostetter and Alibali (2008) discuss other factors that they posit affect the level of the gesture threshold, such as the cognitive system's level of effort, and the strength of connections between premotor and motor areas of the speaker's brain. These factors may also affect the iconicity threshold, but we do not consider these in detail in this paper.

4.2. The Role of Conventionalization

Once a speaker has produced an iconic gesture, vocalization, or prosodic modulation, the novel iconic form becomes a candidate for *conventionalization.* Conventionalization is the process that occurs when communicative forms are used in interactions and, via cultural transmission (Kirby, 2001; Kirby & Hurford, 2002; Scott-Phillips & Kirby, 2010), are shared and repeated among language users. Over time, repetition of a communicative form leads to convergence between language users in what the form means, how it is articulated, and so on. This convergence is what is called conventionalization (e.g., Teich et al., 2021). Conventionalization has no endgame on account of it being a dynamic process, but one indicator of whether a form has undergone a substantial amount of conventionalization is that it is now considered by language users to be a 'real' word or sign, and is codified in dictionaries – it has become part of the lexicon.[3]

[3] However, there may be factors other than conventionalization that influence whether language users perceive a word or sign to be 'real', such as language attitudes. For

This process of conventionalization can cause a communicative form to change in various ways. For example, if the form can be reduced phonologically while remaining intelligible, a reduced form may be preferred for its efficiency, taking less effort and time to produce (e.g., Givón, 1991; Haiman, 1980). Examples of phonological reduction include *gonna* from the future marker *going to* (Leech, 2009)*, dunno* from *don't know* (Scheibman, 2000), and the elision of vowels in unstressed syllables in words like *national* ([næʃnl] instead of [næʃənəl]) and *preferable* ([prefrəbl] instead of [prefərəbəl]) (Gimson, 1989). Another phonological change that may occur is what we will call *phonological standardization*, where phonemes that do not conform with the phonotactics of the linguistic system are eliminated (an example of reduction) or replaced with other phonemes. This change can make the resultant communicative forms easier to remember and produce (Flaksman, 2017), by virtue of featuring phonemes with which speakers are more familiar. This process is similar to *nativization,* where words used in non-native contexts begin to deviate from their original form, influenced by features of the native language (Hsu, 2019; Kachru, 1981; see Cormier et al., 2008 for evidence from signed languages). We see this process at work in the borrowing of loan words from foreign languages with different phonotactics, such as in the deletion of /h/ and replacement of /ŋ/ with /ɲ/ for English loanwords in French, in expressions such as *hold up* and *punk* (LaCharité & Prévost, 1999; Paradis & LaCharité, 2001). Words that undergo phonological standardization also take on

example, if a word gains popularity principally among teenagers, it may be perceived to be colloquial or informal, and it may take longer for the wider language community to see it as a 'real' word, despite its evident conventionalization

features of the native language, but these words originate in iconic depiction rather than borrowing from another language.

In the longer term, the changes brought about by conventionalization may lead to *grammaticalization,* where lexical items are converted into grammatical items (e.g., Heine, 2008; Heine & Kuteva, 2007; Lessau, 1994). For example, the English past tense affix *-ed* originated in a reduction of the word *dedu,* meaning *I did* (Speyer, 2007), with this formal reduction making *dedu* more suitable for novel uses – in this case, as a grammatical item (Teich et al., 2021). In this new grammatical context, the past tense meaning of *dedu* is preserved but all other aspects of its meaning, such as its first person component, are lost in a process known as *semantic bleaching* (e.g., Traugott & Heine, 1991). Moreover, some words may undergo phonological changes like those described above only when used in certain grammatical contexts. For example, going to becomes reduced in phonological form to gonna (Leech, 2009) only when it is used as a functor: speakers might say 'I'm gonna do that', but would typically not say 'I'm gonna the shops' instead of 'I'm going to the shops'.

4.3. Why Iconic Forms Are Repeated

For iconic gestures, vocalizations, and prosodic modulations to be conventionalized and gain entry to the lexicon, they must be shared and repeated in social interactions. Not all communicative forms are created equal: for a form to spread through a community via cultural transmission (Kirby, 2001; Kirby & Hurford, 2002; Scott-Phillips & Kirby, 2010), it must have communicative utility that speakers recognise and repeat for this reason, consciously or unconsciously. Here, we consider the various utilities of iconicity that may lead iconic forms to be repeated in interactions.

First, iconicity may facilitate communication in that it may aid the relaying of sensorimotor simulations between interlocutors. In doing so, the interlocutor creates an imagined experience for their audience (see Pascual, 2014), offering them an insight into "what it would be like to hear, see, or feel what the original speaker [or signer] did" (Wade & Clark, 1993, p. 818). Iconic forms may be suitable for this function because of their direct resemblance to sensorimotor content. Perhaps owing to this resemblance, iconic forms have been shown to activate sensorimotor representations in the brain more strongly than non-iconic forms (e.g., Osaka, 2009, 2011; Osaka et al., 2004; Osaka & Osaka, 2009), which may cause one's audience to more vividly imagine the experience being described.

Iconicity may also be useful because it can resolve referential ambiguity when the referent is not present (see Perniss et al., 2017) and thus aid communication when an interlocutor encounters a form they do not know, when people do not share a common language, and so on. Supporting this idea, one large-scale study showed that users of 28 languages from 12 language families could guess the meaning of vocalizations produced by English speakers at above-chance levels (Ćwiek et al., 2021). In another study, adult listeners more reliably guessed the meaning of novel words when the prosody of these words was iconic rather than non-iconic – for example, when novel words for 'big' or 'tall' meanings had a longer duration and louder volume (Nygaard et al., 2009). It is this disambiguating feature of iconicity that has led some scholars to convincingly argue that iconicity scaffolded the creation of languages, because, in the absence of linguistic conventions, iconicity allows meaning to be communicated right from the start (e.g., Armstrong et al., 1995; Jackendoff, 2002; Stokoe, 1991).

Another communicative benefit of iconicity is that it may aid language learning. For instance, Lockwood et al. (2016) showed that Dutch participants were more successful at learning Japanese ideophones when they were shown the correct definitions of the ideophones – where there was an iconic correspondence between the form of the ideophones and their meaning – than when participants were shown a different definition. Other studies have shown that iconicity facilitates Japanese verb learning in Japanese children (Imai et al., 2008) and English children (Kantartzis et al., 2011), and that iconic words are learned earlier than non-iconic words in both English (Perry et al., 2017) and Spanish (Perry et al., 2015, 2017). Using a connectionist computational model, Gasser (2004) also showed that early word learning is facilitated by iconicity. The resemblance of iconic forms to their meanings may aid word learning because it ensures that these meanings do not have to be learnt entirely by rote.

Iconicity is also useful as a form of engagement or language play. For instance, Dingemanse and Thompson (2020) report a correlation between iconicity and funniness in lexical ratings of over 70,000 English words, citing examples such as *waddle, flop,* and *zigzag* as iconic words with a playful, informal character. Consistent with this characterization of iconic words as being less serious and more informal than non-iconic words, Kim et al. (2021) found that Korean ideophones are typically used with intimates (e.g., friends) rather than non-intimates (e.g., status superiors). In this way, the function of iconicity may be to establish and maintain informal relationships, and so may overlap with other interpersonal strategies, such as the use of humour and colloquialisms.

4.4. From Gesture to Sign

Research indicates that sign systems can form from the conventionalization and grammaticalization of iconic gestures. For example, one study demonstrated that signs in Nicaraguan Sign Language began as holistic, iconic gestures before being parsed into grammatical sub-units and recombined into larger, conventionalized syntactical structures (Senghas et al., 2004). A similar process has been documented for emerging sign languages in Israel (Sandler et al., 2005) and Brazil (Fusellier-Souza, 2006). Moreover, in predominantly speaking communities, deaf children who have not been exposed to a conventional sign language often invent systems that comprise a high number of iconic gestures, called *homesign*, that accrue many of the systematic, rule-governed elements of fully developed languages (e.g., Carrigan & Coppola, 2017; Goldin-Meadow, 2005b). Also, in situations where speech is not possible or permitted but people still need to communicate, such as in a noisy sawmill (Meissner et al., 1975), speakers often perform iconic gestures (e.g., pointing at the wrist to ask how long a job will take) that can develop into sign systems with a degree of conventionalization. Similar findings have been reported for Christian monastic orders (Barakat, 1975) and Australian Aboriginal cultures with a taboo against speaking during mourning (Kendon, 1984).

4.5. From Vocalization and Prosody to Words

Just as sign systems can form via the conventionalization of iconic gestures, spoken lexicons may form partly via the conventionalization of iconic vocalizations and prosodic modulations into fully fledged words. In fact, using computer models, Oudeyer (2005) has shown how conventionalized elements of languages, such as phonotactics, can emerge

from a system of vocalizations. For example, consider the interjection *ugh* /əx/, which communicates disgust using the /x/ phoneme, a sound resembling gagging or throat-clearing, actions associated with disgust (Flaksman, 2017). The /x/ phoneme is not present in most English accents, so this violation of English phonotactics may reveal the origin of *ugh* in iconic vocalization (Attridge, 1988; Bladon, 1977; Rhodes, 1994). The fact that *ugh* is codified by the dictionary Merriam-Webster (2020), and that variants of its pronunciation are listed as [ə], [əg], or [ək], indicate a move toward more standard English phonology, hinting that this vocalization has undergone conventionalization. The trajectory of *ugh* exemplifies how an iconic vocalization can gradually transition from a vocalization into a word, assimilating into the spoken lexicon.

New iconic words may also be coined by prosodically modifying words in the existing lexicon. For instance, the adjective *teeny* may have originated in a prosodically modified articulation of *tiny* /taɪni/, with an accentuation of this word's high frequency components, like the /ɪ/ vowel in its /aɪ/ diphthong (Perlman & Woodin, 2021). This realization of *tiny* as *teeny* may have then undergone conventionalization, similar to *teeny-tiny*, *teensy*, and *teenty* (see Online Etymology Dictionary, 2020). The derivation of *teeny* from *tiny* demonstrates how iconic words can be created from existing, latently iconic material (i.e., 'frozen' iconicity; see Murgiano et al., 2021) and become conventionalized. These new prosodic modulations may outcompete and oust the existing version of the word from the lexicon, or, as is the case with *teeny,* exist alongside it.

4.6. Lexical Distributions of Iconicity in Speech and Sign

The different affordances of gesture, vocalization, and prosody to depict sensorimotor information mean we should see differences in the meanings that are communicated through these iconic strategies, and hence that are conventionalized in the lexicons of spoken languages compared with signed languages. Most obviously, vocalization and prosody are ideally suited to depicting sounds because these iconic strategies are themselves in the auditory modality (Dingemanse et al., 2015; Perlman & Cain, 2014). Likewise, gesture is ideal for depicting manual actions, because gesture is itself a communicative form of action (e.g., Kendon, 2004; Perlman et al., 2018).

In line with these affordances, onomatopoeia is highly prevalent across spoken languages (Dingemanse et al., 2015; Perlman & Cain, 2014; Perniss et al., 2010), and rating studies also show that iconicity is highly concentrated in English and Spanish words for auditory meanings (Perlman et al., 2018; Winter et al., 2017). In comparison, iconicity in American and British Sign Languages is highly concentrated in meanings connected to actions, especially manual actions (Perlman et al., 2018; Perniss et al., 2017). Moreover, spoken languages tend to refer to parts of the vocal tract with sounds articulated using the relevant articulator, with words referring to the lips tending to contain bilabial phonemes (e.g., *mouth* /maʊθ/ contains the bilabial phoneme /m/), and words referring to the tongue tending to contain the lateral phoneme /l/ (e.g., *lick* /lɪk/) (Blasi et al., 2016; Urban, 2011; Wichmann et al., 2010). In comparison, signed languages often use the body and its constituent parts to denote themselves (Meir et al., 2013; Perlman et al., 2018).

As we have discussed, spoken languages such as Japanese and Somali, ideophones (e.g., Akita & Pardeshi, 2019; Dingemanse, 2018; Voeltz & Kilian-Hatz, 2001)

depict not just sound but extend to other domains (Dingemanse, 2012), such as motion, size, brightness, and touch (e.g., Perniss et al., 2010). In these instances, sound may be used iconically based on a cross-modal association with the target meaning (see Parise & Spence, 2012; Spence & Parise, 2012). For instance, the consonantal voicing used in the Japanese ideophones *gorogoro* and *korokoro* (Perniss et al., 2010) iconically depicts the kind of sounds made by heavy versus light objects, with these objects being the actual denotation of the ideophones. Likewise, sound symbolism (e.g., using high pitch to represent small entities; Winter & Perlman, 2021) depicts sound based on its cross-modal association with its intended meaning (e.g., size) (Fitch, 1994; Ohala, 1994; Winter et al., 2021). As these cases of iconicity map meaning from one domain to another, they are sometimes treated as being equivalent, or similar, to metaphor (Marks, 2014).

5. The Iconic Treadmill

As iconic words are conventionalized, evidence shows that they tend to lose their iconicity (Flaksman, 2015, 2017). Despite this deiconization, in her *iconic treadmill hypothesis,* Flaksman (2017) argues that iconicity is replenished in the spoken lexicon via the conventionalization of new iconic forms. In this section, we explicate the pressures that erode the iconicity of individual words, while system-wide iconicity is maintained via the production and conventionalization of new iconic forms, which we argue originate as expressions of sensorimotor simulations. Due to this iconicity maintenance, while words tend to become less iconic, linguistic systems do not (see Perlman & Woodin, 2021). We also discuss other ways in which iconicity is maintained, such as when certain words resist the language changes that would otherwise lead to deiconization.

5.1. Deiconization

There are various selective pressures in the cultural evolution of languages can lead to the deiconization of individual words. One is the pressure toward communicative efficiency. This pressure can lead to phonetic reduction, as long as the reduced word remains intelligible (e.g., Givón, 1991; Haiman, 1980), a process that is observable in simulated language evolution in the laboratory. For example, Perlman, Dale, et al. (2015) asked people to play a guessing game over many repeated interactions, an iterated version of the game of charades. Crucially, participants were not allowed to use English or any other form of conventionalized expression, encouraging them to create new iconic vocalizations. Over repeated iterations, the vocalizations became shorter and less variable. Research shows that more frequent words tend to be shorter (Piantadosi, 2014; Piantadosi et al., 2011; Zipf, 1935), because larger gains in efficiency can be made by reducing the form of words we use more often. If reduction eliminates phonemes that contribute to a word's iconicity, it becomes less iconic. For example, we have seen how the iconic interjection *ugh* /əx/ can be reduced to [ə] (Merriam-Webster, 2020). The removal of the throat-clearing /x/ phoneme makes this word less depictive of its meaning of disgust.

A second pressure is toward ease of articulation. This pressure can be met by reduction, which makes the iconic word shorter (see above), and also by *phonological standardization*, whereby non-standard phonemes are replaced with ones more typical of the language's phonology, which can make iconic words easier to articulate by replacing unfamiliar phonemes with phonemes that speakers have had more practice pronouncing (Flaksman, 2017). For example, the [g] and [k] phonemes in the alternative articulations of *ugh* [əg] and [ək] are probably easier for most native English speakers to pronounce than

/x/. As with the /x/ phoneme in *ugh*, unconventionalized phonemes in an iconic word often do the most depictive work, and hence, through the standardization of these phonemes, these words become less iconic. The price paid in iconicity for making words conform to the phonotactic constraints of a language can be observed by comparing less conventional words and more conventional words. For example, the relatively less conventional *tlok-tlok* (Flaksman, 2017) arguably more accurately mimics the sound of a horse's hooves than the more conventional *clip-clop*, but the word-initial /tl/ consonant cluster is phonotactically impermissible in standard English (see Voeltz & Kilian-Hatz, 2001 for more examples of nonstandard phonemes in iconic words).

Another pressure that can lead to deiconization is the need for words to be discriminable from one another, which may be problematic for iconic words with similar meanings (Dingemanse et al., 2015; Gasser, 2004; Monaghan et al., 2011). After all, if all bird names were iconic representations of their calls (e.g., *cuckoo, chiffchaff, kittiwake*), it may be difficult to distinguish between bird names with similar calls, unless, perhaps, any salient differences between the calls could be emphasized in the iconic word while maintaining some overall resemblance to the call (a strategy that may be increasingly difficult as more words crowd the phonological space). Making words more easily discriminable is likely to make them less iconic. For example, if the best impression of a cuckoo is *cuckoo* /kʊkuː/ but there is another similar word *cuppoo* /kʊpuː/, the word *cuckoo* may change to *cackoo* /kækuː/, which sounds less like the call it was originally imitating. The problem of discriminability is exacerbated by the fact that words for similar meanings tend to be used in similar contexts, increasing the probability of confusion (Thompson et al., 2020). For this reason, there is greater pressure toward discriminability, and hence

deiconization, in lexicons whose semantic space is more densely populated (Gasser, 2004; Sidhu & Pexman, 2018).

There is also a pressure for words to be used flexibly for different grammatical meanings in different contexts. To this end, morphemes can be added to iconic words like *bang* to inflect them for grammatical meanings such as tense (e.g., *He banged his hand on the table*) and plurality (e.g., *I heard several loud bangs*), and allow words to be used in different grammatical slots (e.g., *banged* is a verb, *bangs* is a noun). Also, this process allows iconic words to be morphosyntactically integrated into sentences, as without grammaticalization, iconic words are often limited to clause boundaries, introduced by a quotative like *went* or *like* (Buchstaller, 2002; Butters, 1980; Fox Tree & Tomlinson, 2007), such as in a sentence like *The fireworks went 'bang!'*. The incorporation of non-iconic grammatical elements into iconic words necessarily detracts from their iconicity.

Iconic words may also be deiconized by acquiring new meanings that they do not resemble. Early in the life cycle of an iconic word, the word may resemble its sensorimotor meaning too closely to be redeployed for any other meaning. However, this constraint is dispensed with once a sufficient amount of iconicity is lost via the deiconizing processes described in this section. When this happens, iconic words become *emancipated* from their depictive function, making them more suitable for redeployment for new meanings (Armstrong & Wilcox, 2007). For instance, Flaksman (2017) cites the semantic trajectory of the English word *clip*: 'a clicking sound' > 'the act of cutting off a piece of something' > 'the piece that is cut off' > 'an extract from a media text' > 'a short video'. Along this trajectory, the iconic connection between the sound of the word *clip* and its original auditory referent is broken as the word acquires a new meaning.

Emancipation makes words more susceptible to further system-wide changes that can lead to more deiconization, as there is less pressure for words to retain their original form-meaning resemblance. For example, Flaksman (2017) details how the glottal fricative /h/ in Old English *hlehhan* was originally depictive of the word's meaning of *laughter*, but a consonantal shift across the language (/h/ > /Ø/) obliterated its meaning resemblance. This consonantal shift was indiscriminate, and as such, while there was no inherent benefit in making *hlehhan* less iconic, the word was affected by this system-wide change.

5.2. Iconicity Maintenance

Despite the tendency for many iconic words to lose their iconicity over time, iconicity is replenished and hence maintained in the wider lexicon. One route by which system-wide iconicity is replenished is the coinage of new iconic words via vocalizations. For instance, Flaksman (2017) suggests that newly coined iconic forms filled the iconic void left by *hlehhan* (see above) and became conventionalized, leading to the creation of words such as *chuckle*, *giggle*, and *ha-ha*. Moreover, Perlman and Woodin (2021) moot the possibility that prosodic modulations of existing words can enhance their iconicity, and, through conventionalization, create new, distinct words or drive the existing word to become more iconic over time (e.g., *teeny* may derive from an articulation of *tiny* with an accentuation of the high frequency elements of its /aɪ/ diphthong).

Some iconic words may regain depleted iconicity due to redeployment for a new meaning. For example, the English word *pump*, originally denoting a percussive, 'hitting' sound (ca. 1000), lost its iconicity when it came to mean 'a device for compressing a fluid' (ca. 1400). Around 1883, *pump* acquired the meaning 'to produce the sound that

accompanies water pumping', recovering its depleted iconicity in a different sense to its original meaning (Afanasiev, 1984). Also, in the example of *clip* cited above, this word regained some of its depleted iconicity when it came to denote 'a short video', as its short /ɪ/ vowel and ending plosive /p/ convey a sense of 'shortness', cohering with its new meaning. In some instances, the form and meaning of the word may come to match by chance, a lucky accident of semantic change. In others, the form of an existing word may be co-opted for a new meaning due to its resemblance to this meaning, even though the word was not created with this meaning in mind. Sensorimotor simulation may be involved in the selection of the word for this meaning, as this word is still an representation of sensorimotor information, even if it is not created *de novo*.

Iconic vocalizations and prosodic modifications can also undergo processes of conventionalization without there being any erosion of iconicity. In other words, conventionalization does not inexorably lead to deiconization, though it may tend to (see Perlman & Woodin, 2021). For instance, in one study in which participants played a ten-round game of charades with non-linguistic vocalizations, the vocalizations retained their iconicity even while exhibiting features of conventionalization, such as phonetic reduction (Perlman, Dale, et al., 2015). Other studies have reported similar results (Johansson et al., 2021; Little et al., 2017; Vinson et al., 2021), indicating that iconicity is not always incompatible with other system pressures, like ease of articulation and memorability, and may be preserved where possible.

Some iconic words that are particularly useful for depiction may show resilience to iconicity-diminishing pressures. For example, the English word *peep* /piːp/, whose /iː/ vowel depicts a high pitched sound, resisted the Great Vowel Shift and did not change to a

pronunciation resembling *pipe* /paɪp/ (Durkin, 2011; Hock, 1991; Taub, 1997). The onomatopoeic bird name *cuckoo* also has not changed since Middle English at the latest (Skeat, 2005). Other examples of phonosemantically inert (Flaksman, 2013) words include *booze*, *boulder* (Brodovich, 2008) *hoot*, and *toot* (Flaksman, 2013). Winter et al. (2022) also show that /r/ in words for *rough* has existed in the Indo-European lineage at least since Proto-Indo-European, roughly 6000 years ago.

6. The Iconic Impetus of Language

In this final section, we take the arguments sketched thus far to argue for the primacy of iconicity in language evolution – the *iconic impetus of language.* First, we put together the pieces of the model that we have discussed in a modular fashion until now and present a diagram of how this model works. We then make the argument that the processes of deiconization in this model are motivated and hence non-random, which has the corollary that what is often called 'arbitrariness' in language is not actually arbitrary. Next, we show that these deiconizing processes are not actively biased against iconicity, but are merely *iconicity-blind,* and as a result tend to make words less iconic over time. We conclude the section and paper by discussing the implications of this model for an account of language origins and the evolution of languages over time.

6.1. The Model

Figure 3 presents a schematic diagram of the model outlined so far in this paper. Sensorimotor simulation impels speakers to use an iconic form of expression in their interactions when activation surpasses the speaker's iconicity threshold (see Hostetter &

Alibali, 2008, 2018). Iconic forms of expression include iconic gesture, vocalization, and prosody. Iconic gesture and vocalization are created *de novo,* whereas iconic prosody enhances the iconicity of existing words. Certain of these iconic forms may then become conventionalized as part of the lexicon if they are communicatively useful. Vocalizations become conventionalized into new iconic words, whereas iconic prosody pushes existing words to be more iconic. Conventionalization tends to (but does not always) cause deiconization in response to other system pressures toward, for example, efficiency and discriminability. In reaction to deiconization, sensorimotor simulation again causes the speaker to use an iconic form of expression, and the iconic treadmill continues (see Flaksman, 2017). As mentioned in the introduction, we predict that a parallel process of iconic gesture production, conventionalization into iconic signs, iconic enhancement of existing signs, and so on should occur in signed languages, but we limit the scope of the discussion in this paper to spoken languages.

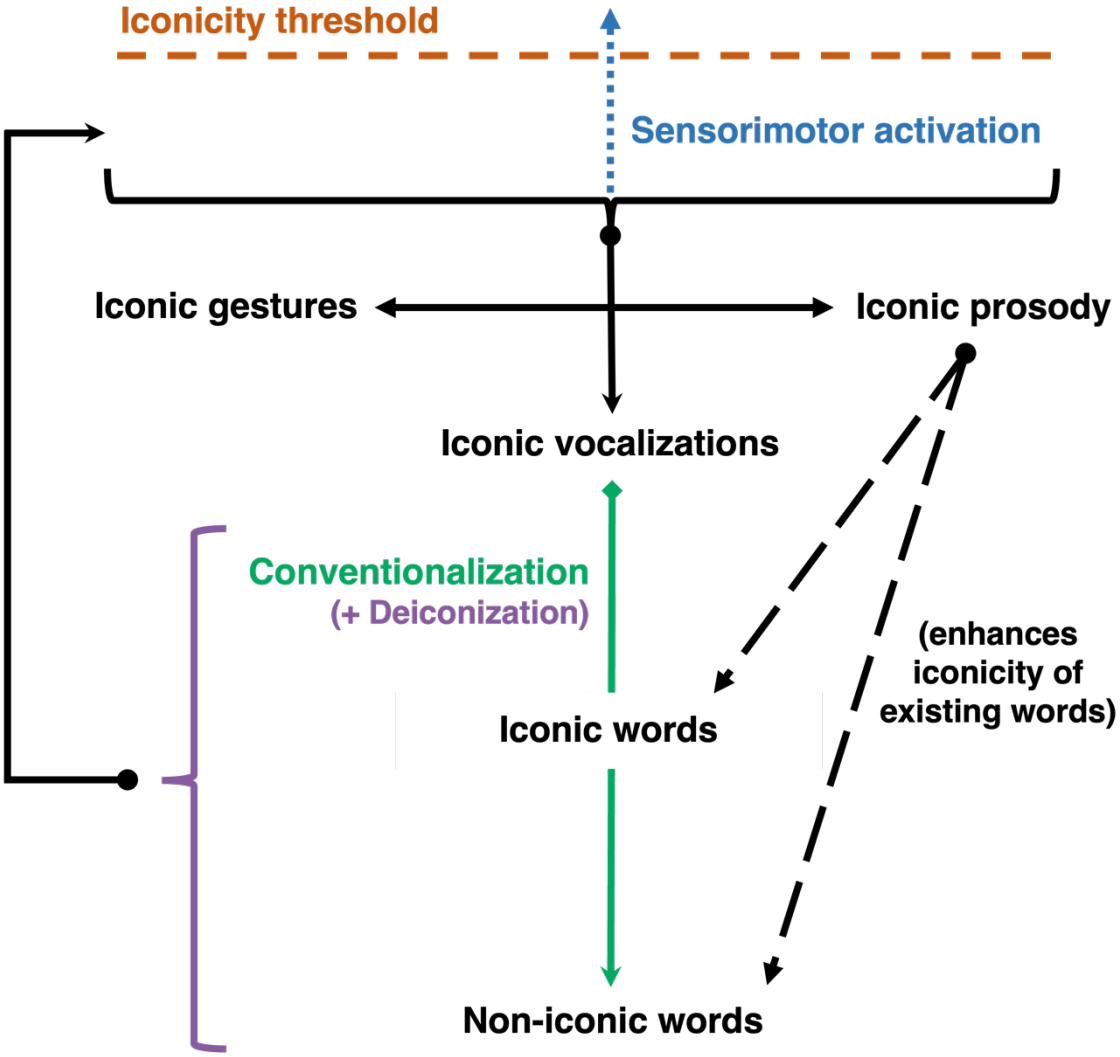


Figure 3. A schematic diagram depicting the model of language evolution described in this paper, which outlines how sensorimotor simulation drives the maintenance of iconicity in spoken lexicons. Note that deiconization is optional in the process of conventionalization, hence why it is written in parenthesis. It is deiconization, not conventionalization per se, that drives the iconic treadmill to generate more iconic forms of expression via sensorimotor simulation.

6.2. ‘Arbitrariness’ Is Not Arbitrary

When Saussure (1959, p. 69) said that the linguistic sign was arbitrary, he meant “that it is unmotivated, i.e. arbitrary in that it actually has no natural connection with the signified”. Other scholars have continued to define arbitrariness not by what it is but by what it is not (Hockett, 1960; Lockwood & Dingemanse, 2015; Sidhu et al., 2021), framing arbitrariness as the opposite of iconicity. According to this negative definition, communicative forms resemble their meanings (iconicity) or they do not (arbitrariness). If a form is not iconic, it is often assumed that the choice to use this form is random and unmotivated – hence the use of the word ‘arbitrary’.

The idea that language can be unmotivated and random is the intuitively appealing explanation for how, for example, the English word *tree* could have the same meaning as the German word *Baum.* However, synchronic comparisons between languages are misleading in this case; it is only when we focus on the same form across languages that the system looks arbitrary. When we take a diachronic perspective within a language, we see that the processes of deiconization that give rise to ‘arbitrariness’ (i.e., non-iconicity), such as phonetic reduction and standardization, are non-random and motivated – for example, by pressures toward efficiency and discriminability – and hence are not arbitrary at all.

Negative definitions of arbitrariness are incompatible with the story of language evolution told in our model, in which communication is iconic from the outset, and only becomes less iconic through deiconization, which often occurs via conventionalization. These deiconizing processes are not random, so they cannot be accurately described as arbitrary, and in fact, describing a dearth of iconicity as ‘arbitrariness’ is unhelpful as it

obscures the communicative pressures that wield significant influence in the evolution of languages. For these reasons, arbitrariness does not figure in our model of language evolution at all, with linguistic change being described in terms of iconicity, conventionalization, and deiconization.

6.3. Conventionalization Is Iconicity-blind

Despite conventionalization tending to bring about deiconization, the target of conventionalization is not deiconization itself. Conventionalization can accept both iconic and non-iconic mappings between form and meaning, requiring only that communities of language users agree on these mappings. As such, deiconization is not an inexorable component of conventionalization. Instead, conventionalization is *iconicity-blind* – it does not 'care' about iconicity, and so is likely (but not certain) to erode it in its pursuit of other goals, such as efficiency and discriminability.

Take, for example, the regular sound changes that can sweep through languages (Bybee, 2002), affecting all lexical items in which the conditioning environment for a sound change is present. Given that regular sound changes generally affect all words in the same way, these sound changes pay no regard to iconicity. Sometimes, regular sound changes leave the iconicity intact; other times, they are "phonosemantically significant" (Flaksman, 2017, p. 26), touching upon the "core, meaning-bearing phonemes of an iconic word," and thus eroding iconicity. Hence, deiconization is not the goal of this process, but it may arise nonetheless as a by-product of sound changes that do not discriminate on the basis of meaning resemblance.

To give another example, discriminability is an important design consideration of lexicons, because a lexicon with highly confusable meanings is maladaptive. To avoid confusability, some lexical items evolve to become less iconic. This evolution is toward discriminability, not deiconization per se, as the desired outcome is a system that is less confusable, which does not *necessarily* lead to deiconization for all linguistic items. Deiconization is a possible outcome of this process, because a lexical system that accepts non-iconic mappings as well as iconic ones is one way of satisfying the discriminability constraint. For this reason, some iconic forms may lose their iconicity as they undergo conventionalization.

Historical linguistics has catalogued the large number of types of language change (Campbell, 2004; Crowley & Bowern, 2010; Hock, 1991), all the way from sound changes (e.g., assimilation, dissimilation, epenthesis, lenition) to word formation processes (e.g., clipping, acronyms, compounding). None of these processes 'care' about the link between form and meaning, so they affect form without attempting to preserve form-meaning resemblance. As there are many more ways in which a unique mapping between form and meaning can be randomly destroyed than ways in which it can be randomly created, iconicity-blind pressures tend to have a deiconizing effect.

Deiconization is not selected for directly because is no *inherent* benefit to deiconization. This is unlike iconicity, which is communicatively useful for its ability to facilitate the relaying of sensorimotor simulations between speakers, resolve referential ambiguity (e.g., Perniss et al., 2017), aid language learning (e.g., Lockwood et al., 2016), and establish and maintain informal relationships (e.g., Kim et al., 2021). Understanding conventionalization as iconicity-blind rather than biased against iconicity elucidates how

iconicity can be preserved even as forms undergo conventionalization (Johansson et al., 2021; Little et al., 2017; Perlman, Dale, et al., 2015; Vinson et al., 2021), and how some conventionalized iconic words resist sound changes that affect other vocabulary items (Durkin, 2011; Flaksman, 2013; Skeat, 2005; Winter et al., 2022).

6.4. Implications for Language Origins

Iconicity is primary in the model of language evolution outlined in this paper, and it is this primacy that we call the iconic impetus of language. The iconic impetus of language leads us to propose that, just as the origins of some signed languages are traceable to the conventionalization of iconic gestures (Fusellier-Souza, 2006; Sandler et al., 2005; Senghas et al., 2004), spoken languages may also have been iconic at their inception.

One benefit of linguistic systems originating in holistic, iconic forms is that meaningful communication is available from the outset, without the need for people to agree upon conventional word-meaning pairings or grammatical and syntactical rules (e.g., Armstrong et al., 1995; Jackendoff, 2002; Stokoe, 1991). Through repeated use and cultural transmission (Kirby, 2001; Kirby & Hurford, 2002; Scott-Phillips & Kirby, 2010), these holistic forms may have been conventionalized and decomposed into syntax, morphology, and other aspects of grammar over time (see Armstrong et al., 1995; Stokoe, 1991). With this in mind, the diagram in Figure 3 may characterize the dynamic of language evolution from the very beginning, with iconicity in the driving seat.

An iconic account of language origins has precedent (e.g., Darwin, 1871; Fay et al., 2013; Givón, 1989; Perlman, Dale, et al., 2015; Ramachandran & Hubbard, 2001; Schmidtke et al., 2014). In fact, the assumption that gesture affords more potential for

iconicity than speech (e.g., Arbib et al., 2008; Armstrong & Wilcox, 2007; Corballis, 2002; Hewes et al., 1973; Tomasello, 2010) has motivated the theory that iconic gestures bootstrapped the creation of spoken languages. However, recent research has evidenced the high potential for iconicity in speech (Clark, 2016; Imai & Kita, 2014; Perlman & Cain, 2014). This reappraisal of the potential for vocal iconicity has led more recent accounts to conjecture that spoken language and gesture emerged in tandem (Kendon, 2017; Perlman, 2017). This conjecture implies that language origins are both iconic and multimodal.

Today, we can infer evidence of sensorimotor simulation on the language we encounter on a daily basis, but the invisible hand of human psychology may have been at work from the genesis of communication, moulding language to match our sensorimotor simulations.

Acknowledgments

Greg Woodin was supported by the Economic and Social Research Council Grant ES/P000711/1. Bodo Winter was supported by the UKRI Future Leaders Fellowship MR/T040505/1.

References

Afanasiev, A. (1984). Voprosy Semanticheskoi Evolutsii Leksiki: Na Materiale Angliiskix Zvukopo-drazhatelnih Suschestvitelnikh (Aspects of Vocabulary Semantic Evolution: A Study of English Imitative Words) [PhD]. University of Leningrad.

Akita, K., & Pardeshi, P. (Eds.). (2019). Ideophones, Mimetics and Expressives. In Ideophones, Mimetics and Expressives. John Benjamins Publishing Company. https://benjamins.com/catalog/ill.16

Alibali, M. W., & Nathan, M. (2007). Teacher's gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), Video Research in the Learning Sciences (pp. 349–365). Erlbaum.

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. Current Anthropology, 49(6), 1053–1063; discussion 1063-1076. https://doi.org/10.1086/593015

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). Gesture and the Nature of Language. Cambridge University Press.

Armstrong, D. F., & Wilcox, S. E. (2007). The Gestural Origin of Language. In The Gestural Origin of Language. Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195163483.001.0001/acprof-9780195163483

Attridge, D. (1988). Peculiar Language: Literature as Difference from the Renaissance to James Joyce. Methuen.

Bak, T., & Hodges, J. (2003). Kissing and dancing—A test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation. Preliminary results in patients with frontotemporal dementia. Journal of Neurolinguistics. https://doi.org/10.1016/S0911-6044(02)00011-8

Barakat, R. A. (1975). Cistercian sign language. In J. Umiker-Sebeok & T. A. Sebeok (Eds.), Monastic Sign Languages (pp. 67–322). Mouton de Gruyter.

Baron, N. S. (1981). Speech, writing, and sign: A functional view of linguistic representation. Indiana University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22(4), 577–609; discussion 610-660. https://doi.org/10.1017/s0140525x99002149

Barsalou, L. W. (2007). Grounded cognition. Annual Review of Psychology, 59(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. Psychonomic Bulletin & Review, 23(4), 1122–1142. https://doi.org/10.3758/s13423-016-1028-3

Barsalou, L. W., & Prinz, J. J. (1997). Mundane creativity in perceptual symbol systems. In Creative thought: An investigation of conceptual structures and processes (pp. 267–307). American Psychological Association. https://doi.org/10.1037/10227-011

Bergen, B. K. (2012). Louder Than Words. Basic Books.

Bladon, R. A. W. (1977). Approaching onomatopoeia. Archivum Linguisticum, 8, 156–166.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. Proceedings of the National Academy of Sciences, 113(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2019). Words as social tools: Language, sociality and inner grounding in abstract concepts. Physics of Life Reviews, 29, 120–153. https://doi.org/10.1016/j.plrev.2018.12.001

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. Psychological Bulletin, 143(3), 263. https://doi.org/10.1037/bul0000089

Bredin, H. (1996). Onomatopoeia as a figure and a linguistic principle. New Literary History, 27(3), 555–569.

Brodovich, O. I. (2008). Zvukoizobrazitelnost i zvukovie zakony (Iconicity and sound changes). In Yazik-Soznanie-Kultura-Sozium (pp. 485–489). Nauka.

Buchstaller, I. (2002). He goes and I'm like: The New Quotatives re-visited. Internet Proceedings of the University of Edinburgh Postgraduate Conference. University of Edinburgh Postgraduate Conference, Edinburgh.

Butters, R. R. (1980). Narrative go 'say'. American Speech, 55(4), 304–307. JSTOR. https://doi.org/10.2307/454573

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. Language Variation and Change, 14, 261–290.

Campbell, L. (2004). Historical linguistics. MIT Press.

Carrigan, E. M., & Coppola, M. (2017). Successful communication does not drive language development: Evidence from adult homesign. Cognition, 158, 10–27. https://doi.org/10.1016/j.cognition.2016.09.012

Christiansen, M. H., & Chater, N. (2016). Creating Language: Integrating Evolution, Acquisition, and Processing. MIT Press.

Christiansen, M. H., & Monaghan, P. (2016). Division of labor in vocabulary structure: Insights from corpus analyses. Topics in Cognitive Science, 8(3), 610–624.

Clark, H. H. (2016). Depicting as a method of communication. Psychological Review, 123(3), 324–347. https://doi.org/10.1037/rev0000026

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. Language, 66(4), 764–805. JSTOR. https://doi.org/10.2307/414729

Clark, N., Perlman, M., & Johansson Falck, M. (2013). Iconic pitch expresses vertical space (M. Borkent, B. Dancygier, & J. Hinnell, Eds.; pp. 393–410). CSLI Publications. http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-81569

Corballis, M. C. (2002). From Hand to Mouth: The Origins of Language. Princeton University Press.

Cormier, K., Schembri, A., & Tyrone, M. (2008). One hand or two? Nativisation of fingerspelling in ASL and BANZSL. Sign Language & Linguistics, 11, 3–44. https://doi.org/10.1075/sl&U38;l.11.1.03cor

Crowley, T., & Bowern, C. (2010). An Introduction to Historical Linguistics (4th ed). Oxford University Press.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., … Perlman, M. (2021). Novel vocalizations are understood across cultures. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-89445-4

Darwin, C. (1871). The Descent of Man. John Murray.

de Saussure, F. (1959). Course in general linguistics [Wade Baskin translation]. The philosophical library.
Decety, J., & Jeannerod, M. (1995). Mentally simulated movements in virtual reality: Does Fitts's law hold in motor imagery? Behavioural Brain Research, 72(1–2), 127–134.
Deuchar, M. (1984). British Sign Language. Routledge.
Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. Language and Linguistics Compass, 6(10), 654–672. https://doi.org/10.1002/lnc3.361
Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. Glossa: A Journal of General Linguistics, 3(1), Article 1. https://doi.org/10.5334/gjgl.444
Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. Trends in Cognitive Sciences, 19(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013
Dingemanse, M., & Thompson, B. (2020). Playful iconicity: Structural markedness underlies the relation between funniness and iconicity. Language and Cognition: An Interdisciplinary Journal of Language and Cognitive Science, 12(1), 203–224. https://doi.org/10.1017/langcog.2019.49
Dreyer, F. R., Frey, D., Arana, S., Saldern, S. von, Picht, T., Vajkoczy, P., & Pulvermüller, F. (2015). Is the motor system necessary for processing action and abstract emotion words? Evidence from focal brain lesions. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.01661
Durkin, P. (2011). The Oxford Guide to Etymology. Oxford University Press.
Edmiston, P., Perlman, M., & Lupyan, G. (2018). Repeated imitation makes human vocalizations more word-like. Proceedings of the Royal Society B: Biological Sciences, 285(1874), 20172709. https://doi.org/10.1098/rspb.2017.2709
Enfield, N. J. (2009). The anatomy of meaning: Speech, gesture, and composite utterances (Vol. 8). Cambridge University Press.
Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. Cognitive Science, 37(7), 1356–1367. https://doi.org/10.1111/cogs.12048
Fitch, W. T. (1994). Vocal tract length perception and the evolution of language [BA Thesis]. Brown University.
Flaksman, M. (2013). Preservation of long vowels in onomatopoeic words denoting pure tones: Phonosemantic inertia. In N. N. Kazanskii (Ed.), The Proceedings of the Joseph M. Tronsky XVII Memorial Annual International Conference Indo-European Comparative Linguistics and Classical Philology in the Institute for Linguistic Studies of Russian Academy of Sciences, St. Petersburg, (pp. 917–923). Nauka.
Flaksman, M. (2015). Diachronicheskoye Razvitie Zvuloizobrazitelnoi Leksiki Angliiskogo Yazika (Diachronic Development of English Iconic Vocabulary) [PhD]. University of St Petersburg.
Flaksman, M. (2017). Iconic treadmill hypothesis: The reasons behind continuous onomatopoeic coinage. In M. Bauer, A. Zirker, O. Fischer, & C. Ljungberg (Eds.), Dimensions of Iconicity (pp. 15–38). John Benjamins Publishing Co.
Flaksman, M. (2020). Pathways of de-iconization. Operationalizing Iconicity, 17, 75–103.
Fox Tree, J. E., & Tomlinson, J. M. (2007). The rise of like in spontaneous quotations. Discourse Processes, 45(1), 85–102. https://doi.org/10.1080/01638530701739280

Fusellier-Souza, I. (2006). Emergence and development of signed languages: From a semiogenetic point of view. Sign Language Studies, 7(1), 30–56.
Gasser, M. (2004). The origins of arbitrariness in language. Annual Conference of the Cognitive Science Society, 26.
Gimson, A. C. (1989). An Introduction to the Pronunciation of English (S. Ramsaran, Ed.; 4th ed). E. Arnold ; Distributed in the USA by Routledge, Chapman and Hall.
Givón, T. (1989). Mind, Code, and Context: Essays in Pragmatics. L. Erlbaum Associates.
Givón, T. (1991). Isomorphism in the grammatical code: Cognitive and biological considerations. Studies in Language. International Journal Sponsored by the Foundation "Foundations of Language", 15(1), 85–114. https://doi.org/10.1075/sl.15.1.04giv
Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 48(7), 905–922. https://doi.org/10.1016/j.cortex.2011.04.010
Goldin-Meadow, S. (2005a). Hearing gesture: How our hands help us think. Harvard University Press.
Goldin-Meadow, S. (2005b). The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us about How All Children Learn Language (Reprinted in paperback). Psychology Press.
Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign and language: The coming of age of sign language and gesture studies. The Behavioral and Brain Sciences, 40, e46. https://doi.org/10.1017/S0140525X15001247
Guerrini, J. (2020). Vowel quality and iconic lengthening. Proceedings of Sinn Und Bedeutung, 24(1), Article 1. https://doi.org/10.18148/sub/2020.v24i1.864
Hadar, U., & Krauss, R. K. (1999). Iconic gestures: The grammatical categories of lexical affiliates. Journal of Neurolinguistics, 12(1), 1–12. https://doi.org/10.1016/S0911-6044(99)00001-9
Haiman, J. (1980). The iconicity of grammar: Isomorphism and motivation. Language, 56(3), 515–540. JSTOR. https://doi.org/10.2307/414448
Hashimoto, T., Usui, N., Taira, M., Nose, I., Haji, T., & Kojima, S. (2006). The neural mechanism associated with the processing of onomatopoeic sounds. NeuroImage, 31(4), 1762–1770. https://doi.org/10.1016/j.neuroimage.2006.02.019
Hassemer, J., & Winter, B. (2016). Producing and perceiving gestures conveying height or shape. Gesture, 15(3), 404–424.
Hassemer, J., & Winter, B. (2018). Decoding gestural iconicity. Cognitive Science, 42(8), 3034–3049. https://doi.org/10.1111/cogs.12680
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
Haynie, H., Bowern, C., & LaPalombara, H. (2014). Sound symbolism in the languages of Australia. PLoS ONE, 9(4). https://doi.org/10.1371/journal.pone.0092852
Heine, B. (2008). Grammaticalization. In B. D. Joseph & R. D. Janda (Eds.), The Handbook of Historical Linguistics (pp. 573–601). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756393.ch18
Heine, B., & Kuteva, T. (2007). The Genesis of Grammar. Oxford University Press.

Hewes, G. W., Andrew, R. J., Carini, L., Choe, H., Gardner, R. A., Kortlandt, A., Krantz, G. S., McBride, G., Nottebohm, F., Pfeiffer, J., Rumbaugh, D. G., Steklis, H. D., Raliegh, M. J., Stopa, R., Suzuki, A., Washburn, S. L., & Wescott, R. W. (1973). Primate communication and the gestural origin of language. Current Anthropology, 14(1/2), 5–24. https://doi.org/10.1086/201401

Hock, H. H. (1991). Principles of Historical Linguistics (2nd rev. and updated ed). Mouton de Gruyter.

Hockett, C. F. (1960). The origin of speech. Scientific American, 203(3), 88–97. https://doi.org/10.1038/scientificamerican0960-88

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. Gesture, 7(1), 73–95. https://doi.org/10.1075/gest.7.1.05hos

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. Psychonomic Bulletin & Review, 15(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Hostetter, A. B., & Alibali, M. W. (2018). Gesture as simulated action: Revisiting the framework. Psychonomic Bulletin & Review. https://doi.org/10.3758/s13423-018-1548-0

Hostetter, A. B., Bieda, K., Alibali, M. W., Nathan, M., & Knuth, E. J. (2006). Don't just tell them, show them! Teachers can intentionally alter their instructional gestures. In R. Sun (Ed.), Proceedings of the 28th Annual Conference of the Cognitive Science Society (pp. 1523–1528). Erlbaum.

Hostetter, A. B., & Skirving, C. J. (2011). The effect of visual vs. Verbal stimuli on gesture production. Journal of Nonverbal Behavior, 35(3), 205–223. https://doi.org/10.1007/s10919-011-0109-2

Hsu, J.-L. (2019). The nativization of English in Taiwanese magazine advertisements. World Englishes, 38(3), 463–485. https://doi.org/10.1111/weng.12423

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. Cognition, 109(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Jackendoff, R. (2002). Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press.

Johansson, N., Anikin, A., & Aseyev, N. (2020). Color sound symbolism in natural languages. Language and Cognition, 12(1), 56–83. https://doi.org/10.1017/langcog.2019.35

Johansson, N., Anikin, A., Carling, G., & Holmer, A. (2019). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. Linguistic Typology, 24(2), 253–310.

Johansson, N., Carr, J. W., & Kirby, S. (2021). Cultural evolution leads to vocal iconicity in an experimental iterated learning task. Journal of Language Evolution, 6(1), 1–25. https://doi.org/10.1093/jole/lzab001

Johansson, N., & Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: A typological study of 101 languages. Public Journal of Semiotics, 5(1), 3–20.

Johnston, T. A., & Schembri, A. (2007). Australian Sign Language (Auslan): An Introduction to Sign Language Linguistics. Cambridge University Press.

Joo, I. (2020). Phonosemantic biases found in Leipzig-Jakarta lists of 66 languages. Linguistic Typology, 24(1), 1–12. https://doi.org/10.1515/lingty-2019-0030

Kachru, B. B. (1981). The pragmatics of non-native varieties of English. In L. E. Smith (Ed.), English for Cross-Cultural Communication (pp. 15–39). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-16572-8_2

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. PLOS ONE, 9(5), e97905. https://doi.org/10.1371/journal.pone.0097905

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. Cognitive Science, 35(3), 575–586. https://doi.org/10.1111/j.1551-6709.2010.01169.x

Kendon, A. (1984). Knowledge of sign language in an Australian Aboriginal community. Journal of Anthropological Research, 40(4), 556–576. https://doi.org/10.1086/jar.40.4.3629797

Kendon, A. (1988). How gestures can become like words. In F. Poyates (Ed.), Cross-Cultural Perspectives in Nonverbal Communication (pp. 131–141). C. J. Hogrefe.

Kendon, A. (2004). Gesture: Visible Action as Utterance. Cambridge University Press.

Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language'. Phil. Trans. R. Soc. B, 369(1651), 20130293.

Kim, H., Winter, B., & Brown, L. (2021). Beyond politeness markers: Multiple morphological and lexical differences index deferential meanings in Korean. Journal of Pragmatics, 182, 203–220. https://doi.org/10.1016/j.pragma.2021.06.006

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. IEEE Transactions on Evolutionary Computation, 5(2), 102–110. https://doi.org/10.1109/4235.918430

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), Simulating the Evolution of Language (pp. 121–148). Springer.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. Linguistics, 35, 379–415. https://doi.org/10.1515/ling.1997.35.2.379

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. Scientific Reports, 7(1), Article 1. https://doi.org/10.1038/s41598-017-05965-y

Kosslyn, S., M. (1980). Image and Mind. MIT Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674443662

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. Journal of Experimental Psychology: General, 140(1), 14.

LaCharité, D., & Prévost, P. (1999). The role of L1 and of teaching in the acquisition of English sounds by francophones. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), Proceedings of the 23rd Annual Boston University Conference on Language Development (Vol. 2, pp. 373–385). Cascadilla Press.

Leech, G. N. (2009). Change in Contemporary English: A Grammatical Study. Cambridge University Press.

Lemaitre, G., Houix, O., Voisin, F., Misdariis, N., & Susini, P. (2016). Vocal imitations of non-vocal sounds. PLOS ONE, 11(12), e0168167. https://doi.org/10.1371/journal.pone.0168167

Lessau, D. A. (1994). A Dictionary of Grammaticalization.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. The Behavioral and Brain Sciences, 22(1), 1–38; discussion 38-75. https://doi.org/10.1017/s0140525x99001776

Little, H., Perlman, M., & Eryilmaz, K. (2017). Repeated interactions can lead to more iconic signals. CogSci.

Locke, J. (1689). An Essay concerning Human Understanding. Oxford University Press.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.01246

Lockwood, G., Hagoort, P., & Dingemanse, M. (2016). How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. Collabra, 2(1). https://doi.org/10.1525/collabra.42

Lu, J. C., & Goldin-Meadow, S. (2018). Creating images with the stroke of a hand: Depiction of size and shape in sign language. Frontiers in Psychology, 9, 1276. https://doi.org/10.3389/fpsyg.2018.01276

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. Journal of Physiology-Paris, 102(1), 59–70. https://doi.org/10.1016/j.jphysparis.2008.03.004

Marttila, A. (2011). A cross-linguistic study of lexical iconicity and its manifestation in bird names. LINCOM Europa.

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago Press.

Meir, I., Padden, C., Aronoff, M., & Sandler, W. (2013). Competing iconicities in the structure of languages. Cognitive Linguistics, 24(2), 309–343. https://doi.org/10.1515/cog-2013-0010

Meissner, M., Philpott, S. B., & Philpott, D. (1975). The sign language of sawmill workers in British Columbia. Sign Language Studies, 9(1), 291–308. https://doi.org/10.1353/sls.1975.0010

Merriam-Webster. (2020). Ugh [Dictionary]. Merriam Websiter. https://www.merriam-webster.com/dictionary/ugh

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. Journal of Experimental Psychology: General, 140(3), 325–347. https://doi.org/10.1037/a0022924

Morey, R. D., Kaschak, M. P., Díez-Álamo, A. M., Glenberg, A. M., Zwaan, R. A., Lakens, D., Ibáñez, A., García, A., Gianelli, C., Jones, J. L., Madden, J., Alifano, F., Bergen, B., Bloxsom, N. G., Bub, D. N., Cai, Z. G., Chartier, C. R., Chatterjee, A., Conwell, E., … Ziv-Crispel, N. (2021). A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). Psychonomic Bulletin & Review. https://doi.org/10.3758/s13423-021-01927-8

Moseley, R. L., Pulvermüller, F., & Shtyrov, Y. (2013). Sensorimotor semantics on the spot: Brain activity dissociates between conceptual categories within 150 ms. Scientific Reports, 3(1), Article 1. https://doi.org/10.1038/srep01928

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating language in the real-world: The role of multimodal iconicity and indexicality. Journal of Cognition, 4(1), Article 1. https://doi.org/10.5334/joc.113

Newmeyer, F. J. (1992). Iconicity and generative grammar. Language, 756–796.

Nölle, J., Chen, C., Hensel, L. B., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2021). Facial expressions of emotion include iconic signals of rejection and acceptance. Journal of Vision, 21(9), 2932. https://doi.org/10.1167/jov.21.9.2932

Nölle, J., Garrod, O. G., Chen, C., Liu, M., Ince, R. A., Schyns, P. G., & Jack, R. E. (2021). Understanding the semiotic affordances of human facial expressions using data-driven methods. Protolang 7, Heinrich Heine University Düsseldorf.

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. Cognitive Science, 33(1), 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x

Ohala, J. J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), Sound Symbolism (pp. 325–347). Cambridge University Press.

Okrent, A. (2002). A modality-free notion of gesture and how it can help us with the morpheme vs. Gesture question in sign language linguistics. In R. P. Meier, K. Kormier, & D. Quinto-Pozos (Eds.), Modality and Structure in Signed and Spoken Language (pp. 175–198). Cambridge University Press.

Online Etymology Dictionary. (2020). Teeny | Origin and meaning of teeny by Online Etymology Dictionary [Dictionary]. Online Etymology Dictionary. https://www.etymonline.com/word/teeny

Osaka, N. (2009). Walk-related mimic word activates the extrastriate visual cortex in the human brain: An fMRI study. Behavioural Brain Research, 198(1), 186–189. https://doi.org/10.1016/j.bbr.2008.10.042

Osaka, N. (2011). Ideomotor response and the neural representation of implied crying in the human brain: An fMRI study using onomatopoeia1. Japanese Psychological Research, 53(4), 372–378. https://doi.org/10.1111/j.1468-5884.2011.00489.x

Osaka, N., & Osaka, M. (2009). Gaze-related mimic word activates the frontal eye field and related network in the human brain: An fMRI study. Neuroscience Letters, 461(2), 65–68. https://doi.org/10.1016/j.neulet.2009.06.023

Osaka, N., Osaka, M., Morishita, M., Kondo, H., & Fukuyama, H. (2004). A word expressing affective pain activates the anterior cingulate cortex in the human brain: An fMRI study. Behavioural Brain Research, 153(1), 123–127. https://doi.org/10.1016/j.bbr.2003.11.013

Oudeyer, P. (2005). The self-organization of combinatoriality and phonotactics in vocalization systems. Connection Science, 17(3–4), 325–341. https://doi.org/10.1080/09540090500217145

Papesh, M. H. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. Journal of Experimental Psychology: General, 144(6), e116. https://doi.org/10.1037/xge0000125

Paradis, C., & LaCharité, D. (2001). Guttural deletion in loanwords. Phonology, 18(2), 255–300. https://doi.org/10.1017/S0952675701004079

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. Experimental Brain Research, 220(3), 319–333. https://doi.org/10.1007/s00221-012-3140-6

Pascual, E. (2014). Fictive Interaction: The Conversation Frame in Thought, Language, and Discourse. John Benjamins Publishing Company.

Pecher, D. (2018). Curb your embodiment. Topics in Cognitive Science, 10(3), 501–517.

Peirce, C. S. (1931). Collected Papers of Charles Sanders Peirce (C. Hartshorne & P. Weiss, Eds.). Harvard University Press.

Perlman, M. (2010). Talking fast: The use of speech rate as iconic gesture. In F. Perrill, V. Tobin, & M. Turner (Eds.), Meaning, Form, and Body. CSLI Publications.

Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. Gesture, 14(3), 320–350. https://doi.org/10.1075/gest.14.3.03per

Perlman, M., Clark, N., & Falck, M. J. (2015). Iconic prosody in story reading. Cognitive Science, 39(6), 1348–1368. https://doi.org/10.1111/cogs.12190

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. Royal Society Open Science, 2(8). https://doi.org/10.1098/rsos.150152

Perlman, M., & Gibbs, R. W., Jr. (2013). Sensorimotor simulation in speaking, gesturing, and understanding. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), Body – Language – Communication: An International Handbook on Multimodality in Human Interaction (pp. 512–533). De Gruyter Mouton.

Perlman, M., Little, H., Thompson, B., & Thompson, R. L. (2018). Iconicity in signed and spoken vocabulary: A comparison between American Sign Language, British Sign Language, English, and Spanish. Frontiers in Psychology, 9. https://doi.org/10.3389/fpsyg.2018.01433

Perlman, M., & Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. Scientific Reports, 8(1), Article 1. https://doi.org/10.1038/s41598-018-20961-6

Perlman, M., & Woodin, G. (2021). A complete real-world theory of language should explain how iconicity remains a stable property of linguistic systems. Journal of Cognition, 4(1), Article 1. https://doi.org/10.5334/joc.166

Perniss, P., Lu, J. C., Morgan, G., & Vigliocco, G. (2017). Mapping language to the world: The role of iconicity in the sign language input. Developmental Science, 21(2), e12551. https://doi.org/10.1111/desc.12551

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. Frontiers in Psychology, 1. https://doi.org/10.3389/fpsyg.2010.00227

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130300. https://doi.org/10.1098/rstb.2013.0300

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. PLoS ONE, 10(9). https://doi.org/10.1371/journal.pone.0137147

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2017). Iconicity in the speech of children and adults. Developmental Science, 21(3), e12572. https://doi.org/10.1111/desc.12572

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. Psychonomic Bulletin & Review, 21(5), 1112–1130. https://doi.org/10.3758/s13423-014-0585-6

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. Proceedings of the National Academy of Sciences, 108(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. Behavioral and Brain Sciences, 13(4), 707–727.

Popp, M., Trumpp, N. M., & Kiefer, M. (2016). Feature-specific event-related potential effects to action- and sound-related verbs during visual word recognition. Frontiers in Human Neuroscience, 10. https://doi.org/10.3389/fnhum.2016.00637

Prinz, W. (1997). Perception and action planning. European Journal of Cognitive Psychology, 9(2), 129–154. https://doi.org/10.1080/713752551

Pulvermüller, F. (1999). Words in the brain's language. Behavioral and Brain Sciences, 22(2), 253–279. https://doi.org/10.1017/S0140525X9900182X

Pulvermüller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. Cerebral Cortex, 9(5), 497–506. https://doi.org/10.1093/cercor/9.5.497

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. Journal of Consciousness Studies, 8(12), 3–34.

Revill, K. P., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. Brain and Language, 128(1), 18–24. https://doi.org/10.1016/j.bandl.2013.11.002

Rhodes, R. (1994). Aural images. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), Sound Symbolism (pp. 276–292). Cambridge University Press.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. Cognitive Brain Research, 3(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. Proceedings of the National Academy of Sciences of the United States of America, 102(7), 2661–2665. https://doi.org/10.1073/pnas.0405448102

Sarvasy, H. (2016). Warblish: Verbal mimicry of birdsong. Journal of Ethnobiology, 36(4), 765–782. https://doi.org/10.2993/0278-0771-36.4.765

Scheibman, J. (2000). I dunno: A usage-based account of the phonological reduction of don't in American English conversation. Journal of Pragmatics, 32(1), 105–124. https://doi.org/10.1016/S0378-2166(99)00032-6

Schmidtke, D., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.00080

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. Trends in Cognitive Sciences, 14(9), 411–417. https://doi.org/10.1016/j.tics.2010.06.006

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. Science, 305(5691), 1779–1782. https://doi.org/10.1126/science.1100199

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. Cognition, 105(3), 681–690. https://doi.org/10.1016/j.cognition.2006.11.005

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. Journal of Memory and Language, 55(2), 167–177. https://doi.org/10.1016/j.jml.2006.03.002

Sidhu, D. M., & Pexman, P. M. (2018). Lonely sensational icons: Semantic neighbourhood density, sensory experience and iconicity. Language, Cognition and Neuroscience, 33(1), 25–31. https://doi.org/10.1080/23273798.2017.1358379

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. Psychonomic Bulletin & Review, 1–9. https://doi.org/10.3758/s13423-021-01883-3

Skeat, W. W. (2005). An Etymological Dictionary of the English Language. Dover Publications.

Spence, C., & Parise, C. V. (2012). The cognitive neuroscience of crossmodal correspondences. I-Perception, 3(7), 410–412. https://doi.org/10.1068/i0540ic

Speyer, A. (2007). Germanische Sprachen: Ein Vergleichender Überblick. Vandenhoeck & Ruprecht.

Stokoe, W. C. (1991). Semantic phonology. Sign Language Studies, 71, 107–114. JSTOR.

Taub, S. F. (1997). Language in the Body: Iconicity and Metaphor in American Sign Language. University of California, Berkeley.

Taylor, K. J. (2007). Ka-BOOM!: A Dictionary of Comic Book Words, Symbols & Onomatopoeia. Mora Publications.

Teich, E., Fankhauser, P., Degaetano-Ortlieb, S., & Bizzoni, Y. (2021). Less is more/more diverse: On the communicative utility of linguistic conventionalization. Frontiers in Communication, 5. https://www.frontiersin.org/article/10.3389/fcomm.2020.620275

Thompson, B., Perlman, M., Lupyan, G., Sehyr, Z. S., & Emmorey, K. (2020). A data-driven approach to the semantics of iconicity in American Sign Language and English. Language and Cognition, 12(1), 182–202. https://doi.org/10.1017/langcog.2019.52

Tomasello, M. (2010). Origins of Human Communication (1. MIT paperback ed). MIT Press.

Traugott, E. C., & Heine, B. (Eds.). (1991). Approaches to Grammaticalization (Vol. 2). John Benjamins.

Trumpp, N. M., Kliese, D., Hoenig, K., Haarmeier, T., & Kiefer, M. (2013). Losing the sound of concepts: Damage to auditory association cortex impairs the processing of sound-related concepts. Cortex, 49(2), 474–486. https://doi.org/10.1016/j.cortex.2012.02.002

Urban, M. (2011). Conventional sound symbolism in terms for organs of speech: A cross-linguistic study. Folia Linguistica, 45(1), 199–213. https://doi.org/10.1515/flin.2011.007

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. Language and Cognition, 1(2), 219–247.

Vinson, D., Jones, M., Sidhu, D. M., Lau-Zhu, A., Santiago, J., & Vigliocco, G. (2021). Iconicity emerges and is maintained in spoken language. Journal of Experimental Psychology: General, 150(11), 2293–2308. https://doi.org/10.1037/xge0001024

Voeltz, F. K. E., & Kilian-Hatz, C. (2001). Ideophones. John Benjamins Publishing Company.

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. Neuropsychologia, 96, 222–229. https://doi.org/10.1016/j.neuropsychologia.2017.01.025

Vukovic, N., & Shtyrov, Y. (2019). Learning with the wave of the hand: Kinematic and TMS evidence of primary motor cortex role in category-specific encoding of word meaning. NeuroImage, 202, 116179. https://doi.org/10.1016/j.neuroimage.2019.116179

Wade, E., & Clark, H. H. (1993). Reproduction and demonstration in quotations. Journal of Memory and Language, 32(6), 805–819. https://doi.org/10.1006/jmla.1993.1040

Wichmann, S., Holman, E. W., & Brown, C. H. (2010). Sound symbolism in basic vocabulary. Entropy, 12(4), Article 4. https://doi.org/10.3390/e12040844

Winter, B., Oh, G. E., Hübscher, I., Idemaru, K., Brown, L., Prieto, P., & Grawunder, S. (2021). Rethinking the frequency code: A meta-analytic review of the role of acoustic body size in communicative phenomena. Philosophical Transactions of the Royal Society B: Biological Sciences, 376(1840), 20200400. https://doi.org/10.1098/rstb.2020.0400

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. Glossa: A Journal of General Linguistics, 6(1), Article 1. https://doi.org/10.5334/gjgl.1646

Winter, B., Perlman, M., Perry, L., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. Interaction Studies, 18(3), 430–451. https://doi.org/10.1075/is.18.3.07win

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. Scientific Reports, 12(1), Article 1. https://doi.org/10.1038/s41598-021-04311-7

Woodin, G., Winter, B., Perlman, M., Littlemore, J., & Matlock, T. (2020). 'Tiny numbers' are actually tiny: Evidence from gestures in the TV News Archive. PLOS ONE, 15(11), e0242142. https://doi.org/10.1371/journal.pone.0242142

Zipf, G. K. (1935). The Psycho-biology of Language (pp. ix, 336). Houghton, Mifflin.

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. Psychonomic Bulletin & Review, 23(4), 1028–1034. https://doi.org/10.3758/s13423-015-0864-x

7. DISCUSSION

This thesis presented four papers that investigated the topics of conceptual metaphor, gestures, data visualization, numerical communication, iconicity, and language evolution. Paper 1 was a multimodal corpus study that investigated the gestures that speakers performed when they used linguistic metaphors that frame numerical quantity in terms of physical size. Paper 2 was an online experiment that investigated the influence of conceptual metaphors of emotional valence, as well as graphical conventions, on the interpretation of time series graphs. Paper 3 was a text-based corpus study that explored the frequencies with which different numbers are used in spoken and written English, and the factors that explain these frequencies. Paper 4 was a theoretical paper positing that sensorimotor simulation is the source of iconic forms of communication, which maintains iconicity in the lexicon, in opposition to other deiconizing pressures, over the course of language evolution.

Papers 1 and 2 corroborate the core claim of Conceptual Metaphor Theory (e.g., Gibbs, 1994; Kövecses, 2002; Lakoff & Johnson, 1980b) that linguistic metaphors reflect conceptual metaphors that structure thought. Paper 1 demonstrated that the gestures produced by speakers when they describe numerical magnitude in terms of physical size reflect the metaphor used in their speech. These metaphoric gestures suggest that the metaphors are conceptually active for these speakers (see Müller, 2008; Woodin, 2019). Paper 2 showed that people's interpretation of time series graphs is facilitated when the graph is consistent with conceptual metaphors of emotional valence. Just like linguistic metaphors are so conventionalized that they are arguably not evidence for conceptual metaphors by themselves (e.g., Black, 1993; Hanks, 2006), established graphical

conventions exist for the presentation of numerical quantity and time metaphors in line graphs (e.g., Abdul-Rahman et al., 2017; Havre et al., 2002; Wattenberg, 2005). As a consequence, Paper 1 investigated metaphoric gestures, which are less conventionalized and are produced more spontaneously than linguistic metaphors (McNeill, 1992, 2000, 2005). Similarly, Paper 2 explored conceptual metaphors of emotional valence in data visualizations, because valence metaphors are not conventionally represented in such visualizations. In both these domains, we found evidence for conceptual metaphors.

Papers 1–3 explored numerical communication in different multimodal contexts: Paper 1 studied metaphorical communication about numbers in spoken language and gestures used in news broadcasts; Paper 2 explored the metaphoric representation of numbers along the axes of line graphs; and Paper 3 analysed the frequency with which numbers are used in speech and writing. Papers 1 and 2 investigated three strategies (language, gesture, data visualization) that are used to express number metaphors, consistent with the claim of Conceptual Metaphor Theory (e.g., Gibbs, 1994; Kövecses, 2002; Lakoff & Johnson, 1980b) that conceptual metaphors have primacy over linguistic metaphors and are expressed through different outlets of communication. Paper 3 was different from Papers 1 and 2 as it did not investigate number metaphors, but explored numerical communication more broadly. This paper showed that number frequencies are connected to four factors: magnitude, roundness, cultural salience, and register. It also found that written numbers of different magnitudes are denoted by different formats (numerals, number words, mixed numbers).

Papers 1 and 3 offer insight into numerical cognition by looking at numerical communication ‘in the wild’ rather than the lab. For example, the size-based gestures of

numerical magnitude reported in Paper 1 are consistent with lab experiments that demonstrate an association between size and numerical magnitude (e.g., Andres et al., 2008; Decarli et al., 2022; Henik & Tzelgov, 1982). Moreover, the bias toward discussing smaller numbers reported in Paper 3 is consistent with random number generation tasks, which find that participants tend to verbalise smaller rather than larger numbers (see Loetscher & Brugger, 2007). Other results in Paper 3 can be interpreted in light of the numerical cognition literature. For example, the results suggested that people use round numbers more approximately at higher magnitudes, which may reflect our increasingly imprecise ability to quantify larger sets (DeWind et al., 2015; Shepard et al., 1975), and/or a logarithmically compressed mental number line where larger numbers are represented less clearly (e.g., Dehaene, 1992). Some experimental results did not replicate in the non-experimental contexts we explored. For instance, in contrast with prior studies evidencing an association between the left (right) hand and smaller (larger) numbers (Daar & Pratt, 2008; Dehaene et al., 1993; Mills et al., 2014), Paper 1 did not find that speakers performed left- versus right-handed gestures consistent with this horizontal conceptualization. This discrepancy in findings may be because speakers in Paper 1 were talking about numerical quantity using size-based language, suggesting that they were conceptualizing numbers in size-based (rather than horizontal) terms. These size-based linguistic metaphors may also have primed the size-based gestures we observed.

Paper 4 put forward the argument that sensorimotor simulation underlies the production of iconic communicative forms, such as iconic gestures, vocalizations, and prosodic modulations of existing words. In Chapter 2.10, I argued that sensorimotor simulation may also underlie metaphoric gestures, vocalizations, and prosody. Hence,

speakers in Paper 1 who produced size-based metaphoric gestures may have been simulating manipulating an object of a small or large size. According to Hostetter and Alibali's (2008, 2018) Gesture as Simulated Action framework, this motor simulation may have caused activation in the speaker's premotor cortex that then spread to their motor cortex, causing them to produce a metaphoric gesture. An equivalent process may be involved for metaphoric vocalization and prosody, as is discussed in Paper 4. Paper 4 focused on iconicity, but we have seen that many cross-modal instances of iconicity mentioned there – such as the depiction of physical size using vocal pitch – can be understood as a kind of metaphor (Marks, 2014).

Paper 4 focused on iconicity in describing the role of sensorimotor simulation in language evolution. Considering the close connection between metaphor and iconicity explicated in this thesis, metaphor may also have a role in the processes of language evolution discussed in this paper. Specifically, sensorimotor simulation may instigate the creation of metaphoric forms (metaphoric gestures, prosody, and vocalizations) that may then undergo conventionalization, replenishing levels of metaphoricity in the lexicon. For example, Perlman (2022) showed that participants produced louder vocalizations to represent larger objects than smaller objects. These metaphoric vocalizations iconically depict the source domain of volume to denote the target domain of size. If these vocalizations are sufficiently useful, they may become conventionalized and gradually become part of the lexicon. The metaphoric forms that are produced via sensorimotor simulation may also undergo further metaphoric extension to new contexts. For instance, 'teeny' iconically depicts the source domain of pitch to represent the target domain of physical size. If this word were used to describe numbers, which do not have a physical

size, this would evidence a kind of double metaphoric extension, from pitch to size to numerical quantity. However, consistent with the argument in Paper 4, the impetus of language is still iconic, rather than metaphoric: iconicity has primacy, as it underlies all metaphoric gestures, prosody, and vocalizations, whereas not all iconic communicative forms are metaphoric (e.g., onomatopoeia).

The results of Papers 2 and 3 can potentially also be interpreted in terms of sensorimotor simulation. Chapter 2.10 made the argument that the 'mental model' that people use to interpret graphs (Padilla, 2018; Vessey, 1991) may be a sensorimotor simulation that is generated in anticipation of seeing a graph. Paper 2 hints that this simulation may be informed partly by conceptual metaphors of emotional valence. The graphs that subverted the graph user's simulation (i.e., ones that subverted valence metaphors) may have been harder to interpret because the graph user was required to transform their sensorimotor simulation, or generate a new one. In relation to Paper 3, Chapter 2.10 speculated that people prefer smaller rather than larger digits in random number generation tasks because they may be simulating a logarithmically compressed number line, where smaller numbers are further apart and hence are represented more distinctly (see Dehaene, 1992). Simulating a mental number line could partly explain the bias toward smaller numbers reported in Paper 3. However, this paper discusses many alternative reasons for why people use smaller numbers more frequently than larger numbers – for example, because smaller quantities are more easily countable and are encountered more often (Cummins, 2015, p. 32).

This thesis connects research topics from across the spectrum of linguistics, cognitive science, and psychology. For example, Paper 1 demonstrates that findings in

numerical cognition (e.g., Andres et al., 2008; Decarli et al., 2022; Henik & Tzelgov, 1982) replicate in a gesture analysis (e.g., Cienki & Müller, 2008; Kendon, 2004; McNeill, 1992) informed by ideas from Conceptual Metaphor Theory (e.g., Gibbs, 1994; Kövecses, 2002; Lakoff & Johnson, 1980b). Paper 2 is an investigation on data visualizations informed by theories from cognitive science (e.g., Padilla, 2018; Vessey, 1991, 1994) and cognitive linguistics, specifically Conceptual Metaphor Theory (e.g., Hartmann et al., 2012; Ishihara et al., 2008; Meier & Robinson, 2004). Paper 3 interprets results from a corpus linguistic paradigm (e.g., Dehaene & Mehler, 1992; Francis & Kučera, 1982; Sinclair & Coulthard, 1975) in light of research into numerical cognition (e.g., Cheyette & Piantadosi, 2019; DeWind et al., 2015; Shepard et al., 1975), and performs analyses inspired by register studies (e.g., Biber, 2009, 2012; Conrad & Biber, 2009). Paper 4 explores research on sensorimotor simulation (e.g., Barsalou, 1999, 2007; Decety & Jeannerod, 1995) in explanation of iconic communication (e.g., Dingemanse et al., 2015; Perniss & Vigliocco, 2014; Vinson et al., 2021), connecting these ideas to language evolution (e.g., Arbib et al., 2008; Christiansen & Chater, 2016; Fay et al., 2013) and historical linguistics (e.g., Flaksman, 2017; Leech, 2009; Teich et al., 2021). As a body of work, this thesis makes connections from the topics explored in the main papers to sensorimotor simulation. By connecting research in fields often kept separate, this thesis has shed new light on the phenomena investigated.

This thesis utilised a range of methodologies, from a multimodal corpus analysis in Paper 1 to online experiments in Paper 2 to a text-based corpus analysis in Paper 3. In these papers, we also cited work that used different methodologies to explore the same topic, or related ones, including experiments in Papers 1 (e.g., Badets et al., 2012;

Lindemann et al., 2007; Winter & Matlock, 2013) and 3 (e.g., Buckley & Gillman, 1974; Dehaene, 1989; Koetsenruijter, 2011). Paper 2 also mostly cited experiments, but mostly ones outside the realm of data visualization, and ones conducted in person rather than online (e.g., Casasanto, 2009; Chinello et al., 2013; Vallesi et al., 2008). As we argue in Banks et al. (accepted), triangulation of data from different methodologies is beneficial for mitigating their respective limitations. For example, experiments looking at size-based number associations constrain participants' behaviour (e.g., they have to respond with button presses) and use a limited stimuli set (e.g., numbers 1–9), creating task demands that may influence the results obtained. Paper 1 instead looked at gestures, which are a much freer, more spontaneous, and arguably more ecologically valid form of expression. When speakers produced these gestures, they mentioned a wider variety of numbers (e.g., 'millions', 'two tenths of one percent') than the stimuli sets used in experiments. This methodology was also able to collect a larger sample than is typically possible in experiments. However, more ecologically valid approaches may lack the rigorous control that is possible in the lab, and TV broadcasts tend to be messier than data collected in experiments, requiring time-consuming filtering to useable videos. By triangulating results from different methodologies and finding equivalent results, we gain confidence that these results reflect a genuine finding, and not just a product of the methodology used in that study.

The work reported here suggests several future research directions. For example, Bouissac (2008) argues that gestures such as those shown in Paper 1 iconically imitate speech rather than being truly metaphoric. While we argue in Paper 1 for the genuine metaphoricity of these gestures, it would nonetheless provide stronger evidence for

conceptual metaphors of numerical quantity if metaphoric gestures occurred with non-metaphorical words, like ‘more’ and ‘less’, as then gestures would not have a linguistic metaphor to imitate. Studies could explore conceptual metaphors of time by looking at whether speakers move their hands leftward when saying ‘earlier’ and rightward when saying ‘later’ (e.g., Ishihara et al., 2008; Vallesi et al., 2008, 2011). Studies could also explore conceptual metaphors of emotional valence by investigating whether speakers move their hands upward when saying ‘better’ and downward when saying ‘worse’ (e.g., Casasanto & Dijkstra, 2010; Globig et al., 2019; Meier & Robinson, 2004). Already the methodology presented in Paper 1 is capable of collecting larger samples to answer such questions than most previous gesture research (cf. Casasanto & Jasmin, 2010; Hinnell, 2018; Lempert, 2011). Nevertheless, sample sizes could be drastically upscaled with automated methods, such as pose detection software OpenPose (Cao et al., 2019). At the very least, OpenPose could assist in the removal of unanalysable videos, saving large amounts of time: in an ongoing project (Woodin et al., 2022) where 35,000 videos were analysed by OpenPose, 26,717 videos were excluded automatically. Ideally, OpenPose could be used to analyse gesture movement in news broadcasts, but this is a work-in-progress due to the messiness of this kind of data. Automated methods have, however, been used in previous work to analyse gestures (e.g., Pouw et al., 2020) and signs (e.g., Östling et al., 2018) in videos where the hands are clearly visible.

Paper 4 is a purely theoretical paper, so empirical data is needed to directly support its main premise. If sensorimotor simulation is the psychological mechanism from which iconic gestures, prosody, and vocalizations originate, the use of one of these iconic strategies (e.g., iconic gesture) may predict another (e.g., iconic prosody), and vice versa.

For example, based on the size-based gestures observed in Paper 1, one could look at whether speakers that perform these iconic gestures also raise their vocal pitch when referring to smaller numbers, and lower their vocal pitch when referring to larger numbers (see Fitch, 1994; Ohala, 1994; Winter et al., 2021). Doing so could further interrogate the idea that there are "degrees of metaphoricity" (Woodin, 2019), with the use of more activation indicators (Müller, 2008) – gesture and prosody – indicating that the linguistic metaphor is more conceptually active. Moreover, Perlman (2010) showed that people talk more quickly when describing fast motion (e.g., 'She dashes past the buildings') than when describing slow motion (e.g., 'She labors slowly'). Based on this finding, one might investigate whether there is a comparable increase in the speed of gestures that people produce when speaking quickly about fast events. Another idea would be to take iconicity rating norms (e.g., Winter et al., 2017) and examine whether words rated as more iconic tend to occur with other iconic forms of expression, like gestures. These are just three examples of the questions that could be interrogated. The automated method described above (Cao et al., 2019) could be used in conjunction with speech analysis software like Praat (Boersma & Weenink, 2022) to conduct a large-scale, multimodal analysis of iconic and metaphoric communication. Doing so would be of great benefit methodologically and would be applicable to many research questions other than the ones discussed here.

Papers 2 and 3 suggest future research directions in their respective topics. For instance, Paper 2 demonstrated that the influence of conceptual metaphors of emotional valence was not sufficient to justify inverting the axes of time series graphs, subverting graphical convention. Future studies could explore graph types where there is no conventional ordering of categories along the y-axis, like stacked bar charts, and test

whether ordering these categories in line with their emotional valence improves their interpretability. This valence-based ordering could be pitted against a quantity-based ordering, where the category with 'more' is placed higher on the y-axis. Doing so could determine whether either of these conceptual metaphors (emotional valence, quantity) matters for the interpretation of these graphs, and if so, which is more important. In Paper 3, we suggested that our results could be reproduced with the Corpus of Contemporary American English (COCA; Davies, 2008) rather than the British National Corpus (BNC Consortium, 2007). In addition to being a larger-scale replication of our study (1 billion versus 100 million words), this could help to identify cultural differences in number use across British English and American English. Moreover, a Principal Component Analysis (Rao, 1964; see Huang et al., 2016) could automatically identify the key dimensions of variation in number use by identifying numbers that tend to cluster together in texts. Performing this analysis would presumably uncover the dimensions identified in Paper 3, but may also identify other dimensions that the original paper did not predict would be relevant.

Overall, this thesis has shown that linguistic metaphors may reflect conceptual metaphors that are evidenced experimentally, which include conceptual metaphors of numerical quantity. These conceptual metaphors are expressed multimodally through metaphoric and iconic gesture, prosody, and vocalizations. Conceptual metaphors of numerical quantity and time may be represented in data visualizations, and conceptual metaphors of emotional valence influence the interpretation of these visualizations. Moreover, the frequency with which numbers are used may be influenced by conceptual metaphors of numerical quantity and other cognitive factors, such as our increasingly

imprecise ability to quantify larger sets. Iconic gestures, vocalizations, and prosodic modulations, such as the metaphoric examples addressed above, may be the outward expression of sensorimotor simulations. Sensorimotor simulation may be involved in conceptual metaphor, and so may help to explain the other results reported here. By synthesising insights from a wide range of research areas and methodologies across the spectrum of linguistics, cognitive science, and psychology, this thesis has shown that metaphor, iconicity, and multimodal numerical communication are interfaces between language and cognition.

BIBLIOGRAPHY

Abdul-Rahman, A., Walton, S., Bemis, K., Lein, J. G., Coles, K., Silver, D., & Chen, M. (2017). Re-spatialization of time series plots. *Information Visualization*, *16*(4), 275–290. https://doi.org/10.1177/1473871616666395

Akita, K. (2017). The linguistic integration of Japanese ideophones and its typological implications. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *62*(2), 314–334. https://doi.org/10.1017/cnj.2017.6

Akita, K., & Pardeshi, P. (Eds.). (2019). Ideophones, Mimetics and Expressives. In *Ideophones, Mimetics and Expressives*. John Benjamins Publishing Company. https://benjamins.com/catalog/ill.16

Aleotti, S., Di Girolamo, F., Massaccesi, S., & Priftis, K. (2020). Numbers around Descartes: A preregistered study on the three-dimensional SNARC effect. *Cognition*, *195*, 104111. https://doi.org/10.1016/j.cognition.2019.104111

Andres, M., Ostry, D. J., Nicol, F., & Paus, T. (2008). Time course of number magnitude interference during grasping. *Cortex*, *44*(4), 414–419. https://doi.org/10.1016/j.cortex.2007.08.007

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1063; discussion 1063-1076. https://doi.org/10.1086/593015

Aryani, A., Hsu, C., & Jacobs, A. M. (2019). Affective iconic words benefit from additional sound–meaning integration in the left amygdala. *Human Brain Mapping*, *40*(18), 5289–5300. https://doi.org/10.1002/hbm.24772

Aulet, L. S., Yousif, S. R., & Lourenco, S. F. (2021). Spatial–numerical associations from a novel paradigm support the mental number line account. *Quarterly Journal of Experimental Psychology*, *74*(10), 1829–1840. https://doi.org/10.1177/17470218211008733

Autry, K. S., Jordan, T. M., Girgis, H., & Falcon, R. G. (2020). The development of young children's mental timeline in relation to emergent literacy skills. *Journal of Cognition and Development*, *21*(1), 1–22. https://doi.org/10.1080/15248372.2019.1664550

Azhar, M., Chen, Y., & Campbell, J. I. D. (2021). Reading direction and spatial effects in parity and arithmetic tasks. *Psychological Research*, *85*(6), 2186–2196. https://doi.org/10.1007/s00426-020-01397-y

Badets, A., Bouquet, C. A., Ric, F., & Pesenti, M. (2012). Number generation bias after action observation. *Experimental Brain Research*, *221*(1), 43–49. https://doi.org/10.1007/s00221-012-3145-1

Banks, B., Borghi, A. M., Fargier, R., Fini, C., Jonauskaite, D., Mazzuca, C., Montalti, M., Villani, C., & Woodin, G. (accepted). Consensus paper: Current perspectives on abstract concepts and future research directions. *Journal of Cognition*, 1–26. https://doi.org/10.5334/ joc.238

Barrett, L. F., & Russell, J. A. (2016). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*. https://journals.sagepub.com/doi/10.1111/1467-8721.00003

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–609; discussion 610-660. https://doi.org/10.1017/s0140525x99002149

Barsalou, L. W. (2007). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., & Prinz, J. J. (1997). Mundane creativity in perceptual symbol systems. In *Creative thought: An investigation of conceptual structures and processes* (pp. 267–307). American Psychological Association. https://doi.org/10.1037/10227-011

Bavelas, J., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture*, *2*(1), 1–17. https://doi.org/10.1075/gest.2.1.02bav

Bergen, B. K. (2012). *Louder Than Words*. Basic Books.

Bergen, B. K., & Chan Lau, T. T. (2012). Writing direction affects how people map space onto time. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00109

Biber, D. (2009). *Dimensions of Register Variation*. Cambridge University Press.

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, *8*(1), 9–37. https://doi.org/10.1515/cllt-2012-0002

Black, M. (1993). More about metaphor. In *Metaphor and Thought* (pp. 19–41). Cambridge University Press.

Blackwell, N. L., Perlman, M., & Fox Tree, J. E. (2015). Quotation as a multimodal construction. *Journal of Pragmatics*, *81*, 1–7. https://doi.org/10.1016/j.pragma.2015.03.004

Bladon, R. A. W. (1977). Approaching onomatopoeia. *Archivum Linguisticum*, *8*, 156–166.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

BNC Consortium. (2007). *British National Corpus* (Version 3). BNC Consortium. http://www.natcorp.ox.ac.uk/

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer.* (6.2.10).

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

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

Boroditsky, L., & Gaby, A. (2010). Remembrances of times east: Absolute spatial representations of time in an Australian aboriginal community. *Psychological Science*, *21*(11), 1635–1639. https://doi.org/10.1177/0956797610386621

Bouissac, P. (2008). The study of metaphor and gesture: A critique from the perspective of semiotics. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 277–282).

Bredin, H. (1996). Onomatopoeia as a figure and a linguistic principle. *New Literary History*, *27*(3), 555–569. https://doi.org/10.1353/nlh.1996.0031

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, *103*(6), 1131–1136. https://doi.org/10.1037/h0037361

Calbris, G. (1990). *The Semiotics of French Gestures*. Indiana University Press.

Cameron, L. (1999). Operationalising 'metaphor' for applied linguistic research. In G. Low & L. Cameron (Eds.), *Researching and Applying Metaphor* (pp. 3–28). Cambridge University Press. https://doi.org/10.1017/CBO9781139524704.004

Cameron, L. (2003). *Metaphor in Educational Discourse* (1st ed.). Continuum.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2019). *OpenPose: Realtime multi-person 2D pose estimation using part affinity fields* (arXiv:1812.08008). arXiv. https://doi.org/10.48550/arXiv.1812.08008

Casasanto, D. (2008). Similarity and proximity: When does close in space mean close in mind? *Memory & Cognition*, *36*(6), 1047–1056. https://doi.org/10.3758/MC.36.6.1047

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351–367. https://doi.org/10.1037/a0015854

Casasanto, D. (2011). Different bodies, different minds: The body specificity of language and thought. *Current Directions in Psychological Science*, *20*(6), 378–383. https://doi.org/10.1177/0963721411422058

Casasanto, D. (2014). Experiential origins of mental metaphors: Language, culture, and the body. In *The power of metaphor: Examining its influence on social life* (pp. 249–268). American Psychological Association. https://doi.org/10.1037/14278-011

Casasanto, D. (2017). The hierarchical structure of mental metaphors. In B. Hampe (Ed.), *Metaphor: Embodied Cognition and Discourse* (pp. 46–61). Cambridge University Press.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593. https://doi.org/10.1016/j.cognition.2007.03.004

Casasanto, D., & Bottini, R. (2014a). Mirror reading can reverse the flow of time. *Journal of Experimental Psychology: General*, *143*(2), 473–479. https://doi.org/10.1037/a0033297

Casasanto, D., & Bottini, R. (2014b). Spatial language and abstract concepts. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(2), 139–149. https://doi.org/10.1002/wcs.1271

Casasanto, D., & Chrysikou, E. G. (2011). When left is 'right'. Motor fluency shapes abstract concepts. *Psychological Science*, *22*(4), 419–422. https://doi.org/10.1177/0956797611401755

Casasanto, D., & Dijkstra, K. (2010). Motor action and emotional memory. *Cognition*, *115*(1), 179–185. https://doi.org/10.1016/j.cognition.2009.11.002

Casasanto, D., & Henetz, T. (2012). Handedness shapes children's abstract concepts. *Cognitive Science*, *36*(2), 359–372. https://doi.org/10.1111/j.1551-6709.2011.01199.x

Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLOS ONE*, *5*(7), e11805. https://doi.org/10.1371/journal.pone.0011805

Casasanto, D., & Jasmin, K. (2012). The hands of time: Temporal gestures in English speakers. *Cogl*, *23*(4), 643–674. https://doi.org/10.1515/cog-2012-0020

Castaño, E., Gilboy, E., Feijóo, S., Serrat, E., Rostan, C., Hilferty, J., & Cunillera, T. (2018). Hand position and response assignment modulate the activation of the valence-space conceptual metaphor. *Cognitive Science*, *42*(7), 2342–2363. https://doi.org/10.1111/cogs.12669

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia*, *37*(4), 395–402. https://doi.org/10.1016/s0028-3932(98)00108-0

Chen, M., & Lin, C.-H. (2021). What is in your hand influences your purchase intention: Effect of motor fluency on motor simulation. *Current Psychology*, *40*(7), 3226–3234. https://doi.org/10.1007/s12144-019-00261-6

Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences*, *116*(36), 17729–17734. https://doi.org/10.1073/pnas.1819956116

Chinello, A., de Hevia, M. D., Geraci, C., & Girelli, L. (2013). Finding the spatial-numerical association of response codes (SNARC) in signed numbers: Notational effects in accessing number representation. *Functional Neurology*, *27*(3), 177–185.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525X1500031X

Cienki, A. (1998). Metaphoric gestures and some of their relations to verbal metaphoric expressions. In J.-P. Koenig (Ed.), *Discourse and Cognition: Bridging the Gap* (pp. 189–204). CSLI Publications.

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In *The Cambridge Handbook of Metaphor and Thought* (pp. 483–502).

Cipora, K., Soltanlou, M., Reips, U.-D., & Nuerk, H.-C. (2019). The SNARC and MARC effects measured online: Large-scale assessment methods in flexible cognitive effects. *Behavior Research Methods*, *51*(4), 1676–1692. https://doi.org/10.3758/s13428-019-01213-5

Clark, N., Perlman, M., & Johansson Falck, M. (2013). *Iconic pitch expresses vertical space* (M. Borkent, B. Dancygier, & J. Hinnell, Eds.; pp. 393–410). CSLI Publications. http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-81569

Conrad, S., & Biber, D. (2009). *Real Grammar: A Corpus-based Approach to English*. Pearson/Longman.

Conson, M., Cinque, F., Barbarulo, A. M., & Trojano, L. (2008). A common processing system for duration, order and spatial information: Evidence from a time estimation task. *Experimental Brain Research*, *187*(2), 267–274. https://doi.org/10.1007/s00221-008-1300-5

Coolidge, F. L., & Overmann, K. A. (2012). Numerosity, abstraction, and the emergence of symbolic thinking. *Current Anthropology*, *53*(2), 204–212. https://doi.org/10.1086/664818

Cooney, S. M., Holmes, C. A., & Newell, F. N. (2021). Children's spatial–numerical associations on horizontal, vertical, and sagittal axes. *Journal of Experimental Child Psychology*, *209*, 105169. https://doi.org/10.1016/j.jecp.2021.105169

Cooperrider, K., Slotta, J., & Núñez, R. (2022). The ups and downs of space and time: Topography in Yupno language, culture, and cognition. *Language and Cognition*, *14*(1), 131–159. https://doi.org/10.1017/langcog.2021.25

Cox, D. (2004). The art and science of visualization: Metaphorical maps and cultural models. *Technoetic Arts*, *2*(2), 71–80.

Cummins, C. (2015). *Constraints on Numerical Expressions*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199687909.001.0001

Ćwiek, A., & Fuchs, S. (2019). Iconic prosody is rooted in sensori-motor properties: Fundamental frequency and the vertical space. *CogSci*, 1572–1578.

Daar, M., & Pratt, J. (2008). Digits affect actions: The SNARC effect and response selection. *Cortex*, *44*(4), 400–405. https://doi.org/10.1016/j.cortex.2007.12.003

Dalmaso, M., Schnapper, Y., & Vicovaro, M. (2022). When time stands upright: STEARC effects along the vertical axis. *Psychological Research*. https://doi.org/10.1007/s00426-022-01693-9

Dancygier, B., & Sweetser, E. (2014). *Figurative Language* (1. publ). Cambridge University Press.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/

De Cruz, H. (2008). An extended mind perspective on natural number representation. *Philosophical Psychology*, *21*(4), 475–490. https://doi.org/10.1080/09515080802285289

de Hevia, M. D. (2021). How the human mind grounds numerical quantities on space. *Child Development Perspectives*, *15*(1), 44–50. https://doi.org/10.1111/cdep.12398

de la Vega, I., Graebe, J., Härtner, L., Dudschig, C., & Kaup, B. (2015). Starting off on the right foot: Strong right-footers respond faster with the right foot to positive words and with the left foot to negative words. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00292

De Ruiter, J. P. (2007). Postcards from the mind: The relationship between speech, imagistic gesture and thought. *Gesture*, *7*(1), 21–38.

de Ruiter, J. P. (2007). Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture*, *7*(1), 21–38. https://doi.org/10.1075/gest.7.1.03rui

Decarli, G., Veggiotti, L., & de Hevia, M. D. (2022). The link between number and action in human infants. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-07389-9

Decety, J., & Jeannerod, M. (1995). Mentally simulated movements in virtual reality: Does Fitts's law hold in motor imagery? *Behavioural Brain Research*, *72*(1–2), 127–134.

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics*, *45*(6), 557–566. https://doi.org/10.3758/BF03208063

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. https://doi.org/10.1016/0010-0277(92)90049-n

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In *Sensorimotor Foundations of Higher Cognition*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231447.003.0024

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29.

Derbyshire, N., Ellis, R., & Tucker, M. (2006). The potentiation of two components of the reach-to-grasp action during object categorisation in visual memory. *Acta Psychologica*, *122*(1), 74–98. https://doi.org/10.1016/j.actpsy.2005.10.004

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: A Journal of General Linguistics*, *3*(1), Article 1. https://doi.org/10.5334/gjgl.444

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2012). The sound of thickness: Prelinguistic infants' associations of space and pitch. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, 6.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*(5), 613–621. https://doi.org/10.1177/0956797612457374

D'Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the Bouba-Kiki Paradigm. *Language and Speech*, *57*(3), 367–393. https://doi.org/10.1177/0023830913507694

Dorogovtsev, S., Mendes, J. F., & Oliveira, J. (2005). Frequency of occurrence of numbers in the World Wide Web. *Physica A: Statistical Mechanics and Its Applications*, *360*, 548–556. https://doi.org/10.1016/j.physa.2005.06.064

Duffy, S. E. (2014). The role of cultural artifacts in the interpretation of metaphorical expressions about time. *Metaphor and Symbol*, *29*(2), 94–112. https://doi.org/10.1080/10926488.2014.889989

Edmiston, P., Perlman, M., & Lupyan, G. (2018). Repeated imitation makes human vocalizations more word-like. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1874), 20172709. https://doi.org/10.1098/rspb.2017.2709

Enfield, N. J. (1999). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances* (First paperback ed). Cambridge University Press.

Evans, V., & Green, M. (2011). *Cognitive Linguistics: An Introduction* (Repr). Edinburgh Univ. Press.

Fauconnier, G. (1994). *Mental spaces: Aspects of Meaning Construction in Natural Language*. Cambridge University Press.

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive Science*, *37*(7), 1356–1367. https://doi.org/10.1111/cogs.12048

Felisatti, A., Ranzini, M., Blini, E., Lisi, M., & Zorzi, M. (2022). Effects of attentional shifts along the vertical axis on number processing: An eye-tracking study with optokinetic stimulation. *Cognition*, *221*, 104991. https://doi.org/10.1016/j.cognition.2021.104991

Fischer, M. H., Dewulf, N., & Hill, R. L. (2005). Designing bar graphs: Orientation matters. *Applied Cognitive Psychology*, *19*(7), 953–962. https://doi.org/10.1002/acp.1105

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition: From single digits to arithmetic. *Quarterly Journal of Experimental Psychology*, *67*(8), 1461–1483. https://doi.org/10.1080/17470218.2014.927515

Fishwick, P. A. (2006). Metaphoric mappings: The art of visualization. In *Aesthetic Computing* (pp. 89–114). MITP. https://ieeexplore.ieee.org/document/6285441

Fitch, W. T. (1994). *Vocal tract length perception and the evolution of language* [BA Thesis]. Brown University.

Flaksman, M. (2017). Iconic treadmill hypothesis: The reasons behind continuous onomatopoeic coinage. In M. Bauer, A. Zirker, O. Fischer, & C. Ljungberg (Eds.), *Dimensions of Iconicity* (pp. 15–38). John Benjamins Publishing Co.

Forker, D. (2020). Elevation as a grammatical and semantic category of demonstratives. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01712

Francis, W. N., & Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin.

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, *34*(8), 1430–1451. https://doi.org/10.1111/j.1551-6709.2010.01105.x

Fuhrman, O., McCormick, K., Chen, E., Jiang, H., Shu, D., Mao, S., & Boroditsky, L. (2011). How linguistic and cultural forces shape conceptions of time: English and Mandarin time in 3D. *Cognitive Science*, *35*(7), 1305–1328. https://doi.org/10.1111/j.1551-6709.2011.01193.x

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*(3), B87–B95. https://doi.org/10.1016/S0010-0277(02)00234-2

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: Evidence from days of the week. *Cortex*, *40*(1), 171–172. https://doi.org/10.1016/S0010-9452(08)70938-9

Gibbs, R. W., Jr. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge University Press.

Gibbs, R. W., & Matlock, T. (2008). Metaphor, imagination, and simulation: Psycholinguistic evidence. In *The Cambridge Handbook of Metaphor and Thought* (pp. 161–176). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.011

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(7), 905–922. https://doi.org/10.1016/j.cortex.2011.04.010

Globig, L. K., Hartmann, M., & Martarelli, C. S. (2019). Vertical head movements influence memory performance for words with emotional content. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00672

Göbel, S. M., Shaki, S., & Fischer, M. H. (2011). The cultural number line: A review of cultural and linguistic influences on the development of number processing. *Journal of Cross-Cultural Psychology*, *42*(4), 543–565. https://doi.org/10.1177/0022022111406251

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*(2), 279–297.

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign and language: The coming of age of sign language and gesture studies. *The Behavioral and Brain Sciences*, *40*, e46. https://doi.org/10.1017/S0140525X15001247

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, *67*(2), 274–284. https://doi.org/10.3758/bf03206491

Guerrini, J. (2020). Vowel quality and iconic lengthening. *Proceedings of Sinn Und Bedeutung*, *24*(1), Article 1. https://doi.org/10.18148/sub/2020.v24i1.864

Gunasti, K. (2016). Consumer reactions to round numbers in brand names. *Marketing Letters*, *27*(2), 309–322.

Hanks, P. (2006). Metaphoricity is gradable. In A. Stefanowitsch & S. Giries Th. (Eds.), *Corpus-Based Approaches to Metaphor and Metonymy* (pp. 17–35). Mouton de Gruyter.

Hartmann, M., Gashaj, V., Stahnke, A., & Mast, F. W. (2014). There is more than 'more is up': Hand and foot responses reverse the vertical association of number magnitudes. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1401–1414. https://doi.org/10.1037/a0036686

Hartmann, M., Grabherr, L., & Mast, F. W. (2012). Moving along the mental number line: Interactions between whole-body motion and numerical cognition. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(6), 1416–1427. https://doi.org/10.1037/a0026706

Hartmann, M., Martarelli, C. S., Mast, F. W., & Stocker, K. (2014). Eye movements during mental time travel follow a diagonal line. *Consciousness and Cognition*, *30*, 201–209. https://doi.org/10.1016/j.concog.2014.09.007

Hartmann, M., & Mast, F. W. (2012). Moving along the mental time line influences the processing of future related words. *Consciousness and Cognition*, *21*(3), 1558–1562. https://doi.org/10.1016/j.concog.2012.06.015

Hartmann, M., Mast, F. W., & Fischer, M. H. (2016). Counting is a spatial process: Evidence from eye movements. *Psychological Research*, *80*(3), 399–409. https://doi.org/10.1007/s00426-015-0722-5

Hashimoto, T., Usui, N., Taira, M., Nose, I., Haji, T., & Kojima, S. (2006). The neural mechanism associated with the processing of onomatopoeic sounds. *NeuroImage*, *31*(4), 1762–1770. https://doi.org/10.1016/j.neuroimage.2006.02.019

Hassemer, J., & Winter, B. (2018). Decoding gestural iconicity. *Cognitive Science*, *42*(8), 3034–3049. https://doi.org/10.1111/cogs.12680

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

Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, *8*(1), 9–20. https://doi.org/10.1109/2945.981848

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory* (pp. xix, 335). Wiley.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. https://doi.org/10.3758/BF03202431

Heurley, L. P., Brouillet, T., Coutté, A., & Morgado, N. (2020). Size coding of alternative responses is sufficient to induce a potentiation effect with manipulable objects. *Cognition*, *205*, 104377. https://doi.org/10.1016/j.cognition.2020.104377

Hinnell, J. (2018). The multimodal marking of aspect: The case of five periphrastic auxiliary constructions in North American English. *Cognitive Linguistics*, *29*(4), 773–806.
Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97. https://doi.org/10.1038/scientificamerican0960-88
Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, *7*(1), 73–95. https://doi.org/10.1075/gest.7.1.05hos
Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495
Hostetter, A. B., & Alibali, M. W. (2018). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-018-1548-0
Hostetter, A. B., & Skirving, C. J. (2011). The effect of visual vs. Verbal stimuli on gesture production. *Journal of Nonverbal Behavior*, *35*(3), 205–223. https://doi.org/10.1007/s10919-011-0109-2
Hostetter, A. B., & Sullivan, E. (2011). Gesture production during spatial tasks: It's not all about difficulty. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1965–1970).
Huang, Y., Guo, D., Grieve, J., & Kasakoff, A. (2016). Understanding US regional linguistic variation with Twitter data analysis. *Computers, Environment and Urban Systems*, *59*, 244–255.
Hudson, R. (1985). The limits of subcategorization. *Linguistic Analysis*, *15*, 233–255.
Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences*, *103*(51), 19599–19604. https://doi.org/10.1073/pnas.0609485103
Iani, C., Ferraro, L., Maiorana, N. V., Gallese, V., & Rubichi, S. (2019). Do already grasped objects activate motor affordances? *Psychological Research*, *83*(7), 1363–1374. https://doi.org/10.1007/s00426-018-1004-9
Ishihara, M., Keller, P. E., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: Evidence for the STEARC effect. *Cortex*, *44*(4), 454–461.
Johnson, R. C. (1967). Magnitude symbolism of English words. *Journal of Verbal Learning and Verbal Behavior*, *6*(4), 508–511. https://doi.org/10.1016/S0022-5371(67)80008-2
Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, *8*(2), 113–139. https://doi.org/10.1016/0885-2014(93)90008-S
Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLOS ONE*, *9*(5), e97905. https://doi.org/10.1371/journal.pone.0097905
Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press.
Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, *35*, 379–415. https://doi.org/10.1515/ling.1997.35.2.379
Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*, 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Koetsenruijter, A. W. M. (2011). Using numbers in news increases story credibility. *Newspaper Research Journal*, *32*(2), 74–82. https://doi.org/10.1177/073953291103200207

Kosslyn, S. (1994). *Image and Brain: The Resolution of the Imagery Debate*. MIT Press.

Kövecses, Z. (2002). *Metaphor: A Practical Introduction*. Oxford University Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.

Lakoff, G., & Johnson, M. (1980a). Conceptual metaphor in everyday language. *The Journal of Philosophy*, *77*(8), 453–486. https://doi.org/10.2307/2025464

Lakoff, G., & Johnson, M. (1980b). *Metaphors We Live By*. University of Chicago Press.

Lakoff, G., & Johnson, M. L. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic books.

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychological Bulletin*, *136*(6), 1045–1067. https://doi.org/10.1037/a0020970

Lang, P., Bradley, M., & Cuthbert, B. (1997). *Motivated attention: Affect, activation and action* (P. J. Lang, B. N. Cuthbert, & R. F. Simons, Eds.). https://www.scienceopen.com/document?vid=87378d88-f5c6-41c9-a7f4-c4c45bdbb67b

Langacker, R. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press. https://books.google.co.uk/books/about/Cognitive_Grammar.html?id=UKVNKz0ZRqwC

Leech, G. N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge University Press.

Lemaitre, G., Houix, O., Voisin, F., Misdariis, N., & Susini, P. (2016). Vocal imitations of non-vocal sounds. *PLOS ONE*, *11*(12), e0168167. https://doi.org/10.1371/journal.pone.0168167

Lemaitre, G., Susini, P., Rocchesso, D., Lambourg, C., & Boussard, P. (2014). Non-verbal imitations as a sketching tool for sound design. In M. Aramaki, O. Derrien, R. Kronland-Martinet, & S. Ystad (Eds.), *Sound, Music, and Motion* (pp. 558–574). Springer International Publishing. https://doi.org/10.1007/978-3-319-12976-1_34

Lempert, M. (2011). Barack Obama, being sharp: Indexical order in the pragmatics of precision-grip gesture. *Gesture*, *11*(3), 241–270. https://doi.org/10.1075/gest.11.3.01lem

Leone, M. J., Salles, A., Pulver, A., Golombek, D. A., & Sigman, M. (2018). Time drawings: Spatial representation of temporal concepts. *Consciousness and Cognition*, *59*, 10–25. https://doi.org/10.1016/j.concog.2018.01.005

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1–38; discussion 38-75. https://doi.org/10.1017/s0140525x99001776

Li, H. (2020). Less is more: Motor fluency impairment and body-specific representation of valence. *Review of Cognitive Linguistics. Published under the Auspices of the Spanish Cognitive Linguistics Association*, *18*(2), 519–534. https://doi.org/10.1075/rcl.00069.li

Lindemann, O., Abolafia, J. M., Girardi, G., & Bekkering, H. (2007). Getting a grip on numbers: Numerical magnitude priming in object grasping. *Journal of Experimental*

Psychology. Human Perception and Performance, *33*(6), 1400–1409. https://doi.org/10.1037/0096-1523.33.6.1400

Locke, J. (1689). *An Essay concerning Human Understanding*. Oxford University Press.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01246

Loetscher, T., Bockisch, C. J., & Brugger, P. (2008). Looking for the answer: The mind's eye in number space. *Neuroscience*, *151*(3), 725–729. https://doi.org/10.1016/j.neuroscience.2007.07.068

Loetscher, T., Bockisch, C. J., Nicholls, M. E. R., & Brugger, P. (2010). Eye position predicts what number you have in mind. *Current Biology: CB*, *20*(6), R264-265. https://doi.org/10.1016/j.cub.2010.01.015

Loetscher, T., & Brugger, P. (2007). Exploring number space by random digit generation. *Experimental Brain Research*, *180*(4), 655–665. https://doi.org/10.1007/s00221-007-0889-0

Marks, L. E. (2014). *The Unity of the Senses: Interrelations Among the Modalities*. Elsevier Science. http://qut.eblib.com.au/patron/FullRecord.aspx?p=1882879

Marshall, C. B. (1968). Living on the left hand of God. *Theology Today*, *25*(2), 195–200. https://doi.org/10.1177/004057366802500206

McCormick, K., Kim, J. Y., List, S., & Nygaard, L. C. (2015). Sound to meaning mappings in the bouba–kiki effect. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1565–1570). Cognitive Science Society.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

McNeill, D. (Ed.). (2000). *Language and Gesture*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620850

McNeill, D. (2005). *Gesture and Thought*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226514642.001.0001

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Association between affect and vertical position. *Psychological Science*, *15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*, 119–141. https://doi.org/10.1075/gest.4.2.02mel

Merten, K., & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience*, *21*(2), 333–346. https://doi.org/10.1162/jocn.2008.21032

Miles, L. K., Tan, L., Noble, G. D., Lumsden, J., & Macrae, C. N. (2011). Can a mind have two time lines? Exploring space-time mapping in Mandarin and English speakers. *Psychonomic Bulletin & Review*, *18*(3), 598–604. https://doi.org/10.3758/s13423-011-0068-y

Mills, K. J., Rousseau, B. R., & Gonzalez, C. L. R. (2014). A cross-sectional developmental examination of the SNARC effect in a visually-guided grasping task. *Neuropsychologia*, *58*, 99–106. https://doi.org/10.1016/j.neuropsychologia.2014.04.001

Moseley, R. L., Pulvermüller, F., & Shtyrov, Y. (2013). Sensorimotor semantics on the spot: Brain activity dissociates between conceptual categories within 150 ms. *Scientific Reports*, *3*(1), Article 1. https://doi.org/10.1038/srep01928

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520. https://doi.org/10.1038/2151519a0

Müller, C. (2008). *Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View*. University of Chicago Press.

Murphy, G. L. (1996). On metaphoric representation. *Cognition*, *60*(2), 173–204. https://doi.org/10.1016/0010-0277(96)00711-1

Murphy, G. L. (1997). Reasons to doubt the present evidence for metaphoric representation. *Cognition*, *62*(1), 99–108. https://doi.org/10.1016/S0010-0277(96)00725-1

Newmeyer, F. J. (1992). Iconicity and generative grammar. *Language*, 756–796.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*(1), 149–157. https://doi.org/10.1016/s0896-6273(02)01144-3

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, *28*, 225–252. https://doi.org/10.1146/annurev.anthro.28.1.225

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*(1), 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*(1), 1–18. https://doi.org/10.1159/000261678

Ohala, J. J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound Symbolism* (pp. 325–347). Cambridge University Press.

Okrent, A. (2002). A modality-free notion of gesture and how it can help us with the morpheme vs. Gesture question in sign language linguistics. In R. P. Meier, K. Kormier, & D. Quinto-Pozos (Eds.), *Modality and Structure in Signed and Spoken Language* (pp. 175–198). Cambridge University Press.

Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature*, *414*(6860), 197–200. https://doi.org/10.1038/35102562

Osaka, N. (2009). Walk-related mimic word activates the extrastriate visual cortex in the human brain: An fMRI study. *Behavioural Brain Research*, *198*(1), 186–189. https://doi.org/10.1016/j.bbr.2008.10.042

Osaka, N. (2011). Ideomotor response and the neural representation of implied crying in the human brain: An fMRI study using onomatopoeia1. *Japanese Psychological Research*, *53*(4), 372–378. https://doi.org/10.1111/j.1468-5884.2011.00489.x

Osaka, N., & Osaka, M. (2009). Gaze-related mimic word activates the frontal eye field and related network in the human brain: An fMRI study. *Neuroscience Letters*, *461*(2), 65–68. https://doi.org/10.1016/j.neulet.2009.06.023

Osaka, N., Osaka, M., Morishita, M., Kondo, H., & Fukuyama, H. (2004). A word expressing affective pain activates the anterior cingulate cortex in the human brain: An fMRI study. *Behavioural Brain Research*, *153*(1), 123–127. https://doi.org/10.1016/j.bbr.2003.11.013

Östling, R., Börstell, C., & Courtaux, S. (2018). Visual Iconicity across sign languages: Large-scale automated video analysis of iconic articulators and locations. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00725

Overmann, K. A. (2013). Material scaffolds in numbers and time. *Cambridge Archaeological Journal*, *23*(1), 19–39. https://doi.org/10.1017/S0959774313000024

Padakannaya, P., Devi, M. L., Zaveria, B., Chengappa, S. K., & Vaid, J. (2002). Directional scanning effect and strength of reading habit in picture naming and recall. *Brain and Cognition*, *48*(2–3), 484–490.

Padilla, L. M. (2018). A case for cognitive models in visualization research. *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*, 69–77. https://doi.org/10.1109/BELIV.2018.8634267

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: A cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, *3*, 29. https://doi.org/10.1186/s41235-018-0120-9

Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O., & Bertini, E. (2015). *How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques* (SSRN Scholarly Paper ID 2566968). Social Science Research Network. https://papers.ssrn.com/abstract=2566968

Parsons, P. (2018). *Conceptual Metaphor Theory as a foundation for communicative visualization design*.

Pascual, E. (2014). *Fictive Interaction: The Conversation Frame in Thought, Language, and Discourse*. John Benjamins Publishing Company.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences*, *19*(10), 590–602. https://doi.org/10.1016/j.tics.2015.08.003

Peirce, C. S. (1931). *Collected Papers of Charles Sanders Peirce* (C. Hartshorne & P. Weiss, Eds.). Harvard University Press.

Perlman, M. (2010). Talking fast: The use of speech rate as iconic gesture. In F. Perrill, V. Tobin, & M. Turner (Eds.), *Meaning, Form, and Body*. CSLI Publications.

Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*(3), 320–350. https://doi.org/10.1075/gest.14.3.03per

Perlman, M., Clark, N., & Falck, M. J. (2015). Iconic prosody in story reading. *Cognitive Science*, *39*(6), 1348–1368. https://doi.org/10.1111/cogs.12190

Perlman, M., & Gibbs, R. W., Jr. (2013). Sensorimotor simulation in speaking, gesturing, and understanding. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (pp. 512–533). De Gruyter Mouton.

Perlman, M., & Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-20961-6

Perlman, M., Paul, J., & Lupyan, G. (2022). Vocal communication of magnitude across language, age, and auditory experience. *Journal of Experimental Psychology. General*, *151*(4), 885–896. https://doi.org/10.1037/xge0001103

Perlman, M., Paul, J. Z., & Lupyan, G. (2015). Congenitally deaf children generate iconic vocalizations to communicate magnitude. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, & C. D. Jennings (Eds.), *Proceedings of the 37th Annual Cognitive Science Society Meeting (CogSci 2015)* (pp. 315–320).

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00227

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130300. https://doi.org/10.1098/rstb.2013.0300

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–727.

Pitt, B., & Casasanto, D. (2020). The correlations in experience principle: How culture shapes concepts of time and number. *Journal of Experimental Psychology. General*, *149*(6), 1048–1070. https://doi.org/10.1037/xge0000696

Plevnik, J. (1991). Son of man seated at the right hand of God: Luke 22,69 in Lucan Christology. *Biblica*, *72*(3), 331–347.

Popp, M., Trumpp, N. M., & Kiefer, M. (2016). Feature-specific event-related potential effects to action- and sound-related verbs during visual word recognition. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00637

Pouw, W., Trujillo, J. P., & Dixon, J. A. (2020). The quantification of gesture–speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. *Behavior Research Methods*, *52*(2), 723–740. https://doi.org/10.3758/s13428-019-01271-9

Pulvermüller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex*, *9*(5), 497–506. https://doi.org/10.1093/cercor/9.5.497

Rao, C. R. (1964). The use and interpretation of Principal Component Analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, *26*(4), 329–358.

Revill, K. P., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language*, *128*(1), 18–24. https://doi.org/10.1016/j.bandl.2013.11.002

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Ruiz Fernandéz, S., Lachmair, M., & Rahona, J. J. (2014). Human mental representation of time in the vertical space. *Proceedings of the 6th International Congress of Medicine in Space and Extreme Environments (ICMS)*. 6th International Congress of Medicine in Space and Extreme Environments (ICMS, Berlin.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295x.110.1.145

Ruud, P., Schunk, D., & Winter, J. (2014). Uncertainty causes rounding: An experimental study. *Experimental Economics*, *17*(3), 391–413.

Safar, J. (2020). “When you were that little…”: From Yucatec Maya height-specifier gestures to Yucatec Maya Sign Language person-classifier signs. *Gesture*, *19*(1), 1–40. https://doi.org/10.1075/gest.19007.saf

Sarvasy, H. (2016). Warblish: Verbal mimicry of birdsong. *Journal of Ethnobiology*. https://doi.org/10.2993/0278-0771-36.4.765

Sasaki, K., Yamada, Y., & Miura, K. (2015). Post-determined emotion: Motor action retrospectively modulates emotional valence of visual images. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1805), 20140690. https://doi.org/10.1098/rspb.2014.0690

Saussure, F. de. (1966). *Course in General Linguistics*. McGraw-Hill.

Seno, T., Kawabe, T., Ito, H., & Sunaga, S. (2013). Vection modulates emotional valence of autobiographical episodic memories. *Cognition*, *126*(1), 115–120. https://doi.org/10.1016/j.cognition.2012.08.009

Shaki, S., & Fischer, M. H. (2008). Reading space into numbers – a cross-linguistic comparison of the SNARC effect. *Cognition*, *108*(2), 590–599. https://doi.org/10.1016/j.cognition.2008.04.001

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, *16*(2), 328–331. https://doi.org/10.3758/PBR.16.2.328

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, *7*(1), 82–138. https://doi.org/10.1016/0010-0285(75)90006-7

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, *55*(2), 167–177. https://doi.org/10.1016/j.jml.2006.03.002

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, *28*(4), 1390–1398. https://doi.org/10.3758/s13423-021-01883-3

Sigurd, B. (1988). Round numbers. *Language in Society*, *17*(2), 243–252.

Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Steen, F. F., & Turner, M. (2013). Multimodal Construction Grammar. In *Language and the Creative Mind* (pp. 255–274). CSLI Publications.

Stocker, K., Hartmann, M., Martarelli, C. S., & Mast, F. W. (2016). Eye movements reveal mental looking through time. *Cognitive Science*, *40*(7), 1648–1670. https://doi.org/10.1111/cogs.12301

Taylor, K. J. (2007). *Ka-BOOM!: A Dictionary of Comic Book Words, Symbols & Onomatopoeia*. Mora Publications.

Teich, E., Fankhauser, P., Degaetano-Ortlieb, S., & Bizzoni, Y. (2021). Less is more/more diverse: On the communicative utility of linguistic conventionalization. *Frontiers in Communication*, *5*. https://www.frontiersin.org/article/10.3389/fcomm.2020.620275

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 830–846.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, *8*(6), 769–800. https://doi.org/10.1080/13506280042000144

Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science*, *3*(3), 499–535. https://doi.org/10.1111/j.1756-8765.2010.01113.x

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, *23*(4), 515–557. https://doi.org/10.1016/0010-0285(91)90005-9

Ultan, R. (1978). Size-sound symbolism. In J. Greenberg (Ed.), *Universals of Human Language* (pp. 525–567). Stanford University Press.

Vallesi, A., Binns, M. A., & Shallice, T. (2008). An effect of spatial-temporal association of response codes: Understanding the cognitive representations of time. *Cognition*, *107*(2), 501–527. https://doi.org/10.1016/j.cognition.2007.10.011

Vallesi, A., McIntosh, A. R., & Stuss, D. T. (2011). How time modulates spatial responses. *Cortex*, *47*(2), 148–156. https://doi.org/10.1016/j.cortex.2009.09.005

Van Leeuwen, T., & Kress, G. (1995). Critical layout analysis. *Internationale Schulbuchforschung*, *17*(1), 25–43.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature*. *Decision Sciences*, *22*(2), 219–240. https://doi.org/10.1111/j.1540-5915.1991.tb00344.x

Vessey, I. (1994). The effect of information presentation on decision making: A cost-benefit analysis. *Information & Management*, *27*(2), 103–119. https://doi.org/10.1016/0378-7206(94)90010-8

Vinson, D., Jones, M., Sidhu, D. M., Lau-Zhu, A., Santiago, J., & Vigliocco, G. (2021). Iconicity emerges and is maintained in spoken language. *Journal of Experimental Psychology: General*, *150*(11), 2293–2308. https://doi.org/10.1037/xge0001024

Voeltz, F. K. E., & Kilian-Hatz, C. (2001). *Ideophones*. John Benjamins Publishing Company.

Wade, E., & Clark, H. H. (1993). Reproduction and demonstration in quotations. *Journal of Memory and Language*, *32*(6), 805–819. https://doi.org/10.1006/jmla.1993.1040

Walker, E. J., & Cooperrider, K. (2016). The continuity of metaphor: Evidence from temporal gestures. *Cognitive Science*, *40*(2), 481–495. https://doi.org/10.1111/cogs.12254

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

Wattenberg, M. (2005). Baby names, visualization, and social data analysis. *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 1–7. https://doi.org/10.1109/INFVIS.2005.1532122

Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language*, *99*, 122–150. https://doi.org/10.1016/j.jml.2017.09.006

Winter, B. (2014). Horror movies and the cognitive ecology of primary metaphors. *Metaphor and Symbol*, *29*(3), 151–170. https://doi.org/10.1080/10926488.2014.924280

Winter, B., & Matlock, T. (2013). More is up… and right: Random number generation along two axes. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*.

Winter, B., & Matlock, T. (2017). Primary metaphors are both cultural and embodied. In *Metaphor: Embodied Cognition and Discourse* (pp. 99–116). Cambridge University Press. https://doi.org/10.1017/9781108182324.007

Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience and Biobehavioral Reviews*, *57*, 209–219. https://doi.org/10.1016/j.neubiorev.2015.09.005

Winter, B., Oh, G. E., Hübscher, I., Idemaru, K., Brown, L., Prieto, P., & Grawunder, S. (2021). Rethinking the frequency code: A meta-analytic review of the role of acoustic body size in communicative phenomena. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1840), 20200400. https://doi.org/10.1098/rstb.2020.0400

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics*, *6*(1), Article 1. https://doi.org/10.5334/gjgl.1646

Winter, B., Perlman, M., & Matlock, T. (2013). Using space to talk and gesture about numbers: Evidence from the TV news archive. *Gesture*, *13*(3), 377–408. https://doi.org/10.1075/gest.13.3.06win

Winter, B., Perlman, M., Perry, L., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies*, *18*(3), 430–451. https://doi.org/10.1075/is.18.3.07win

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-021-04311-7

Woodin, G. (2019). *Degrees of Metaphoricity: A Quantitative Gesture Analysis* [MA]. University of Birmingham.

Woodin, G., & Winter, B. (2018). Placing abstract concepts in space: Quantity, time and emotional valence. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02169

Woodin, G., Winter, B., & Uhrig, P. (2022). *How people gesture when they talk about actions: Evidence from TV News Media* [Oral]. British Association for Applied Linguistics, Queen's University Belfast.

Yang, Z., & Purves, D. (2003). A statistical explanation of visual space. *Nature Neuroscience*, *6*(6), 632–640. https://doi.org/10.1038/nn1059

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture*, *5*(1–2), 165–190. https://doi.org/10.1163/1568537054068660

Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition*, *57*(3), 271–295. https://doi.org/10.1016/0010-0277(95)00674-3

Zima, E., & Bergs, A. (2017). Multimodality and construction grammar. *Linguistics Vanguard*, *3*(s1). https://doi.org/10.1515/lingvan-2016-1006