

This is a peer-reviewed preprint of the work below.
 Cuskley, Christine and Kees Sommer (forthcoming). The evolution of linguistic iconicity and the cross-modal cognitive suite. To appear in Olga Fisher, Kimi Akita, and Pamela Perniss (eds.), *Oxford Handbook of Iconicity in Language*. Oxford University Press: Oxford, UK.

The evolution of linguistic iconicity and the cross-modal cognitive suite

Christine Cuskley^{*1}, Kees Sommer²

¹Language Evolution, Acquisition and Development Group
 School of English Literature, Language and Linguistics
 Newcastle University, Newcastle upon Tyne, UK, NE1 7RU

²Communication & Multimedia Design
 Faculty of IT & Design
 The Hague University of Applied Sciences, The Hague, NL, 2521 EN

*Contact: christine.cuskley@ncl.ac.uk

Abstract (145 words)

This chapter will focus on the deep evolutionary history of the cognitive capacities underlying linguistic iconicity. The complex capacity for linguistic iconicity has roots in a more general cross-modal ability present throughout the animal kingdom, cross-modal transfer. Cross-modal transfer is the ability to make basic inferences about sensory properties of an object in multiple modalities based on experience from only one. This situates iconicity as a fundamentally cross-modal phenomenon; part of a broader, uniquely human cross-modal cognitive suite which includes relatively rare phenomena like synesthesia, alongside more ubiquitous phenomena like sensory metaphor and cross-modal correspondences. Evidence suggests the evolutionarily deep capacity for cross-modal transfer was honed into more sophisticated capacities underlying iconicity by an evolutionary ratchet of increased prosociality during human self-domestication. This period provided strong selective pressures for increasingly complex cross-sensory communication, and eventually, the predominantly arbitrary symbolic systems that underpin modern human language.

Keywords: Cross-modal correspondences, cross-modal transfer, sound symbolism, language evolution

Word Count (excl. ref list): 8049

1. Introduction

Iconicity occurs where there is a resemblance-based mapping between linguistic form and meaning (Sidhu and Pexman, 2018), and it is often invoked as playing a key role in language evolution (Cuskley and Kirby, 2013; Imai and Kita, 2014). This invocation rests on two specific pillars: learning and grounding. Iconicity confers well-established advantages in language learning (Nielsen and Rendall, 2012) and has been shown to play a key role particularly in child language acquisition (e.g., Kantartzis, Imai and Kita, 2011). In learning, an iconic connection between a linguistic form and its meaning makes a mapping easier to store and

retrieve (Gasser, 2004; Lockwood, Dingemanse and Hagoort, 2016; see also Nielsen and Dingemanse, 2021). In terms of grounding, a form connected to its meaning can provide the foundation for a shared convention: in a scenario where two people don't already share a communication system and wish to establish one, tying form to meaning is a key means to getting communication "off the ground". For example, if multiple members of a population naturally associate close vowels with smaller entities, an utterance like "*tiku*" to refer to a small animal in a complex scene - alongside other multimodal grounding like deictic pointing - allows for rapid formation of shared form-meaning pairings. In other words, forms (including vocalisations and gestures) can refer to something other than themselves, by symbolising their meaning in a shared intuitive manner. A shared understanding of which forms intuitively map to which meanings can accelerate the construction of a lexicon (Imai and Kita, 2014).

Both the advantages of learning and grounding fit seamlessly into our increasingly multimodal understanding of language: an integration of vocalisation and manual-facial gestures was likely key in the face-to-face niche in which language evolved, and bolsters language acquisition (Vigliocco, Perniss and Vinson, 2014; Murgiano, Motamedi and Vigliocco, 2021; Rasenberg *et al.*, 2022). While vocalisations in and of themselves have the capacity to be iconically grounded (Cuskley, 2013; Perlman, Dale and Lupyan, 2015), the greater visuo-spatial affordances of the manual-gestural modality play a key role in bolstering this. Importantly, this is not an all or nothing process: iconicity is graded, and can shade into or be bolstered by linguistic conventions (Ahlner and Zlatev, 2010). Occhino and colleagues (2017) provide an illustrative example of this from sign languages: while the sign for tree is iconically motivated in different languages (e.g., Danish Sign Language, Chinese Sign language, American Sign Language), it nonetheless takes a different form in each language. In other words, there are many ways to express a single meaning iconically. This entails the fact that iconicity can be, but isn't necessarily, *transparent* in that it has an obvious meaning for a completely naïve onlooker or hearer. Form-meaning pairings can be iconic for some particular language users while being opaque or arbitrary to others (Occhino *et al.*, 2017); forms can have a variety of guises, occurring in just one modality at a time (e.g., as a sound), or as a multimodal unit (e.g., a sound temporally synchronized with gesture and touch; Hodge and Ferrara, 2022).

While both learning and grounding are well-established areas of evidence for the role of iconicity in language evolution, these only partially cover the questions we need to answer for a fully comprehensive understanding of the evolution of language iconicity (and other areas of language; see Spike, 2017). Tinbergen (1963) identified a systematic way of thinking about

behavioral phenomena in terms of proximate and ultimate views (Bateson and Laland, 2013). Proximate explanations focus on *causal mechanisms* (how a trait works in an organism, e.g., neurological structures underlying it), and *development* (how a trait develops over an individual organism's lifespan; see Perniss and Vigliocco, 2014). Ultimate explanations focus on the *adaptive function* of a trait (what it evolved "for"), and the trait's *phylogeny* (its evolutionary history, including related traits in other species). The ultimate adaptive function of iconicity has been clearly articulated in terms of the aforementioned grounding of emergent language systems (Cuskley, 2013; Cuskley and Kirby, 2013; Imai and Kita, 2014; Vigliocco, Perniss and Vinson, 2014). Here, we focus on ultimate phylogenetic mechanisms of iconicity, particularly its evolutionary precursors and selective pressures which led to the broader cross-modal cognitive suite which enables linguistic iconicity.

To answer more ultimate phylogenetic questions, we can use what we know about the proximate factors of a trait as a starting point. Sidhu and colleagues (2018) identify five mechanisms underlying sound symbolic associations, which in turn form the basis for vocal iconicity: statistical learning, shared (low or high level) properties, neurological factors, evolved species-general associations, and linguistic patterns. This provides a useful framework for the kinds of proximate factors which *can* underlie these associations, but they note that these aren't mutually exclusive for any given association. Particularly within a Tinbergian evolutionary framework, these mechanisms offer different levels of explanation for a given association. Identifying neurological structures involved in sound symbolic associations and statistical learning provides explanatory power in mechanistic and developmental terms at the proximate level. On the other hand, evolved species-general traits provide a partial ultimate explanation, confined mainly to associations between size and pitch (Ohala, 1995). Linguistic patterns have both a proximate and ultimate side: in terms of proximate development, they form the basis of statistical learning, while in an ultimate sense, they may arise from the amplification of general sound symbolic biases in the cultural transmission of language (Imai and Kita, 2014).

Shared properties, both at high and low levels, seem to have the most explanatory power; therefore, *how* we have the capacity to compute and connect shared properties is a key question. For example, close front vowels and small shapes share the property of smallness, and stop consonants and angular shapes have connotations of being fast and tense (Sapir, 1929; Bozzi and Flores D'Arcais, 1967; Sidhu and Pexman, 2018, table 3, p. 1631). *Why* do close front vowels have the property of smallness for us, and *how* did we evolve the capacity to store

this connection? Why are stop consonants or angular shapes especially fast and how did they come to be so? What is the deeper evolutionary history of the capacity for linguistic iconicity?

This chapter will aim to fill this gap, by proposing an evolutionary path from cross-modal transfer into the human cross-modal cognitive suite. Section 2 will demonstrate that cross-modal capacities underlying language iconicity share close ties to the evolutionarily older capacity for cross-modal transfer. Section 3 will outline other phenomena, including cross-modal correspondences, cross-modal metaphor, and synesthesia, which descended from cross-modal transfer to form the uniquely human *cross-modal cognitive suite*. Section 4 will tie the emergence of this cross-modal cognitive suite to human self-domestication, which has been identified as a key process in the evolution of language structure and cultural transmission (Thomas and Kirby, 2018). The environment created by self-domestication acted as a key selective pressure to expand our existing cross-modal capacities to include language iconicity, which has fundamentally social and communicative functions. Finally, Section 5 will provide concluding thoughts on the evolution of linguistic iconicity and the cross-modal cognitive suite.

2. Cross-modal transfer: a precursor of language iconicity

2.1 Understanding cross-modal transfer

Briefly, cross-modal transfer¹ is the ability to utilize sensory information from one modality to make inferences about, or undertake actions in, another. For example, we readily make assumptions based on the appearance of an object about how it would feel (Saito *et al.*, 2003), even if we've only ever seen it. In other words, we use input from only one modality (vision) to make inferences about properties from another (touch). This ability is pervasive across our sensory modalities, with evidence of cross-modal transfer from touch to olfaction (Demattè *et al.*, 2006), vision to taste (Slocombe, Carmichael and Simner, 2016), and sound to vision (Rose, 2013). Developmentally in humans, cross-modal transfer emerges early and robustly: for example, measures of cardiac response show that infants match intensities of light and sound as early as three weeks old (Lewkowicz and Turkewitz, 1980).

¹ Note that there is substantial variation in the use of terminology related to concepts covered in this chapter. The term “cross-modal” is sometimes used interchangeably with multisensory, multimodal, cross-sensory, intersensory, intermodal, and a variety of other terms generally used idiosyncratically across the literature. This entry uses *cross-modal transfer* to refer to the phenomenon of transferring experience in one modality directly to another (e.g., inferring visual features based on tactile experience), *cross-modal correspondences* to refer to associations between modalities which seem to be broadly shared in a population regardless of their underlying mechanism (including cross-modal correspondences underpinned by cross-modal transfer), and *multimodal* to refer to representations or interactions involving multiple modalities.

Cross-modal transfer can seem like an impressive leap of inference, and indeed for a long time was considered to be uniquely human (Davenport, Rogers and Russell, 1973). However, the capacity for cross-modal transfer is essential for the function of embodied action in an organism: it is a basic component of a functional multisensory system. For example, accurate inferences about how a substance might taste (and thus, its nutritional or caloric value) based on what it looks or smells like are basic cognitive tools for survival (Simner, Cuskey and Kirby, 2010). Likewise, coordinating simple movements in three-dimensional space requires cross-modal transfer: visual input allows an organism to coordinate effective movements using the sense of the body in space (proprioception). In other words, even our distant primate ancestors would have required cross-modal transfer to function in their physical environment: the basic ability to coordinate a successful leap to (and grasp of) a previously unencountered branch requires cross-modal transfer from vision to proprioception (Ramachandran, 2004). Performing this kind of task based *only* on visual input of a novel scene would have been essential even for our ancestors tens of millions of years ago.

2.2 Cross-modal transfer in other species

Given this basic function of cross-modal transfer within multisensory systems, it is unsurprising that it is a capacity shared widely across the animal kingdom. Until the second half of the twentieth century, many researchers argued the ability was uniquely human (discussed in Davenport, Rogers and Russell, 1973), potentially arising as a result of our linguistic cognition (Geschwind, 1974). Nonetheless, research has shown that a wide array of phylogenetically distant species share the capacity to varying degrees. Both our closest ape relatives and more distantly related monkeys have demonstrated this capacity (Cowey and Weiskrantz, 1975; Ettlinger and Jarvis, 1976). Early work also showed that rats trained to discriminate the intensity of a sound (e.g., performing a task for a reward with a high intensity stimulus, but not a low intensity one) spontaneously transfer this training to light (Over and Mackintosh, 1969). Schumacher and colleagues (2016) trained fish on discrimination of light and electrical stimuli in the weakly electric fish, *Ganthonemus petersii*. They found that the fish can transfer discrimination training both between light and electro-sensory perception, and that this transfer is particularly sensitive to extremes. Solvi and colleagues (2020) trained bumblebees to discriminate between different shapes in total darkness – meaning they had only experienced the shapes haptically. The bees later encountered the shapes visually, but could not touch them; they nonetheless maintained discrimination performance in the visual

modality. This was also reversible: bees were able to discriminate haptically between shapes (in the dark) having only ever seen them previously.

Overall, testing for cross-modal transfer in other species can be tricky: it relies primarily on discrimination training, which is time intensive and requires captive populations. Given the practical difficulty with operationalizing cross-modal transfer, its apparent survival value, and the phylogenetically diverse array of species demonstrating the ability, it is quite possible that current research has underestimated the ubiquity of cross-modal transfer.

3. The cross-modal cognitive suite

While early theoretical work suggested language might underpin cross-modal “processing” (Geschwind, 1974), this was later abandoned as a rich body of evidence emerged that cross-modal transfer was significantly more evolutionarily widespread than language (Ettlinger and Wilson, 1990). However, the reverse – that cross-modal abilities are a prerequisite for language – has amassed substantial evidence. More specifically, we propose that cross-modal transfer has blossomed in primates – and particularly humans – into a more sophisticated cross-modal cognitive suite. This suite provides us with neurological and behavioral tools to not only make, but *store and extend* cross-modal representations. These representations underpin our ability to learn arbitrary symbols, and tie together atypical sensory phenomena like synesthesia with more abstract and mundane sensory phenomena like cross-modal metaphor.

3.1 Cross-modal correspondences

The human cross-modal cognitive suite extends beyond cross-modal transfer to more arbitrary cross-modal correspondences. These are shared associations that involve either the extension or even the apparent uncoupling of the clear connective mechanisms inherent to cross-modal transfer. We generalise cross-modal correspondences that arise from statistical learning: small animals make high pitched sounds, so we associate higher pitched sound with smaller shapes in general (Ohala, 1995; Mondloch and Maurer, 2004). However, many cross-modal correspondences have no clear mechanism. Cross-modal correspondences are often described as “non-arbitrary” because they tend to be shared across participants at high rates (Spence, 2011; Slobodenyuk *et al.*, 2015). But there is often no transparent mechanism at play. In terms of linguistic iconicity, there is some ineffable “resemblance based mapping” between

forms; but we don't always have a sense of what underlies the resemblance, beyond the fact that we seem to share a strong intuition that there is one.

For example, we associate high pitch with abstract shape angularity: higher pitches are more angular, while low pitch is round (Marks, 1987). This association is robust across participants, but there is no obvious physical mechanism to connect pitch and angularity. Some researchers argue that all cross-modal correspondences might be reduced to statistical learning: they are the result of picking up on co-occurrences in the environment, like that larger objects tend to be heavier (Spence, 2011). Proponents of this approach suggest that statistical regularities in the environment may also underlie less obvious associations. In some cases, research has substantiated this: Parise and colleagues (2014) have shown that robust associations between pitch and spatial location (high pitch² tends to be high in space, low pitch low) are reflected in the statistical properties of soundscapes. In other words, high pitched sounds tend to actually come from higher elevations in natural soundscapes.

But for many associations, mechanistic explanations which rely on statistical regularities are tenuous, and have more of a feel of “just-so” stories. For example, Parise and Spence (2012) suggest that for pitch and angularity, “harder objects resonate at higher frequencies and tend to scatter into sharper pieces than softer objects when broken” (Slobodenyuk *et al.*, 2015, p. 1380). Thus, we associate higher pitch with more angular objects. But even if this held, it would be a very complex chain of cross-modal correspondences: tying the shape of a two-dimensional object to its three dimensional form, simulating the behavior of this 3D form when broken, associating this with hardness, hardness to resonant frequency, and resonant frequency of an object to pitch. This particular mechanistic chain is untested, and there may be a simpler explanation for shape-pitch correspondences; but the point is that this *kind* of explanatory mechanism involves a complex cross-modal chain, bringing this well beyond the norm of more basic cross-modal transfer. Moreover, these kinds of explanations are especially common for the kinds of linguistic cross-modal correspondences that underlie iconicity. The next section provides a brief overview of linguistic cross-modal correspondences and some of their proposed mechanisms. Following that, we connect these explicitly to two other linguistic elements of the cross-modal cognitive suite: synesthesia and metaphor.

3.2 Linguistic cross-modal correspondences

² Note that the metaphorical language used to describe pitch (high vs low) in English reflects this association; this will be discussed further in section 3.3.

Work on linguistic cross-modal correspondences can be traced back in large part to early work by Köhler (1929) on shape and Sapir (1929) on size. Briefly, findings based on this work have shown that shape features are associated with a variety of phonetic, phonological, and articulatory features. In consonants, voicing, manner of articulation and sonority have been associated with roundedness (e.g., Nielsen and Rendall, 2011; D’Onofrio, 2014; Cuskley, Simner and Kirby, 2017), while stops are more angular and continuants tend to be more rounded (Westbury, 2005; Nielsen and Rendall, 2012). In terms of vowels, both closer and more front vowels are associated with more angular shape (Ozturk, Krehm and Vouloumanos, 2013), and back-rounded vowels are associated with rounder shapes³ (Maurer, Pathman and Mondloch, 2006). Linguistic cross-modal correspondences involving shape are sensitive to orthographic effects: though the effect is persistent across cultures with diverse writing systems (Ćwiek, et al., 2022), the acquired cross-modal shape-sound mappings inherent in alphabetic writing systems can interfere with these effects in certain kinds of tasks (Cuskley, Simner and Kirby, 2017). Additionally, it is possible that orthography *reflects* linguistic cross-modal correspondences: letters *look* like they sound, to some degree (Jee, Tamariz and Shillcock, 2018).

In terms of size, close front vowels are associated with small entities, and open back vowels are associated with large entities (Thompson and Estes, 2011; Ohtake and Haryu, 2013; for more detailed reviews of size and angularity symbolism, see Sidhu and Pexman, 2018; Westbury *et al.*, 2018). While the majority of work on linguistic cross-modal correspondences has tended to focus on shape and size, recent work has also focused on motion (Cuskley, 2013), smell (Speed *et al.*, 2021), taste (Simner, Cuskley and Kirby, 2010), color (Cuskley *et al.*, 2019; Johansson, Anikin and Aseyev, 2020), and emotions (Adelman, Estes and Cossu, 2018) (see also Dingemanse, Perlman and Perniss, 2020 for a detailed overview of experimental approaches to linguistic iconicity).

Linguistic cross-modal correspondences often focus specifically on *sound* symbolic associations (Hinton, Nichols and Ohala, 2006; Sidhu and Pexman, 2018), probing the bounds of how sound maps onto other modalities, responding to the historically underrated potential for vocal iconicity (Perlman, Dale and Lupyan, 2015; Johansson, Carr and Kirby, 2021). However, particularly in the context of language evolution, the term *linguistic cross-modality*

³ It’s worth noting that roundedness and frontness are confounded in many studies which look at these features. Like cognitive sciences generally (Blasi et al., 2022; Henrich, Heine and Norenzayan, 2010), the vast majority of these studies focus on English speakers, for whom there are no rounded front vowel phonemes, and the majority of back vowels are rounded.

is more apt, extending beyond sound symbolism in two important ways. First, the term *linguistic cross-modal correspondences* encompasses sound symbolic associations, but also gestural symbolism in multimodal linguistic interaction, which may be used to infuse non-transparent vocalizations with iconicity (Cuskley and Kirby, 2013; Rasenberg *et al.*, 2022). Second, a focus on linguistic cross-modal correspondences underscores their considerable potential to account not only for grounding in spoken languages, but also signed languages which occur primarily in the manual/visual channel. Overall, the term *linguistic cross-modal correspondence* better fits the “multimodal, polysemiotic, and plurifunctional” nature of iconicity (Hodge and Ferrara, 2022).

Multimodal interactions are of crucial importance given the face-to-face niche of linguistic interaction. Symbolic alignment emerges earlier and more reliably in multimodal contexts (Rasenberg *et al.*, 2022), and gestural signals arguably have greater affordance for transparent cross-modal mappings, particularly to the visual modality (Hockett, 1978). There is little, if any, work on “gestural cross-modal correspondences” *per se*, mainly because gestures tend to be much more obviously iconic (often in part because of their transparency), and so this work falls under the umbrella of manual or gestural iconicity. There is still a great deal to understand about the mechanisms at play in creating and understanding iconic gestures (see e.g., Hassemer and Winter, 2018; Ortega and Özyürek, 2020). However, the mechanisms connecting, for example, small gestures and small quantities (e.g., Woodin *et al.*, 2020), are more obvious than, for example, those which connect shape roundness to voicing. These tend to be more straightforward extensions of cross-modal transfer (smaller quantities tend to occupy less space), as opposed to the complex (and often hypothetical) syllogistic cross-modal chains often invoked to explain cross-modal correspondences like those between roundedness and voicing.

The proximate mechanisms underlying a particular cross-modal correspondence (including those that are linguistic in nature) will range from being relatively transparent and simple (small things make high pitched sounds) to more opaque and complex. Each correspondence – linguistic or not – requires further study to illuminate its specific properties and underlying proximate mechanisms. Nonetheless, these associations can be considered extensions of the general and evolutionarily deep capacity for cross-modal transfer. Linguistic cross-modal correspondences, and cross-modal correspondences more broadly, are often characterized as innate or “hard-wired” (see Spence and Deroy, 2012 for an overview of arguments about innateness cross-modal correspondences generally, and Sidhu and Pexman,

2018 for sound symbolic associations). In the first instance, there are general philosophical and conceptual issues with claims dealing in concepts like innateness (Mameli and Bateson, 2011) and the degree to which behaviors are “hard-wired” (Lilienfeld *et al.*, 2015). More importantly, if we wish to illuminate the phylogeny of linguistic cross-modal correspondences (and thus, iconicity), it is more fruitful to consider the *capacity* for cross-modal correspondences in general as the entity we need to explain the evolution of, rather than specific instantiations.

In other words, regardless of the source of a particular cross-modal correspondence, it is our capacity for making and using these associations that is unique, and this has been used to particular effect in language. Moreover, humans haven’t just expanded the capacity for cross-modal transfer into more opaque cross-modal correspondences but moved further in two ways. First, we utilize this capacity in the creation and interpretation of sensory metaphors, which are a major source of expansion within a vocabulary. Second, in some individuals, cross-modal cognition has gone beyond ineffable association and into the realm of sensory experience in the form of synesthesia.

3.2 Metaphor

Leaving a detailed discussion of the relationship between metaphor and iconicity aside (see Hiraga, 2005), this section will focus on the relationship between cross-modal correspondences and metaphor. Some metaphors have such a strong cross-modal basis that we barely recognize them as metaphorical. In English, we refer to sounds with higher frequency as high pitch, and so associations between pitch and position in space seem almost literal. Metaphors work because they are based on the kinds of shared extended affordances inherent to, and sometimes reflecting, shared cross-modal correspondences. Pitch-height metaphors are conventionalized in English, but even efforts to explain them in literal terms lead to more metaphor. The supposedly dry language of physics that describes pitch in terms of *high frequency* is technically somewhat metaphorical; fundamental frequency changes as a function of the distance between crests in a sound wave, usually quantified in terms of how many waves pass within a second (Hz) – this is more literally how *fast* the sound wave is cycling (amplitude, experienced as loudness, is conventionally represented as the height of a sound wave on a 2D plane).

Like there are many roads to iconicity, there are also many roads to metaphor. Height is not a universal metaphor for talking about variations in sound frequency; languages use a variety of metaphors in this domain including weight, age, strength, and thickness (Eitan and

Timmers, 2010; Dolscheid *et al.*, 2013). Most readers can likely guess which way these metaphors go with little effort and no experience in the relevant languages: high pitch is light while low pitch is heavy (Kpelle, Liberia; Stone, 1981), high pitch is young and low pitch is old (Suyà people of the Amazon Basin; Zbikowski, 1998), and high pitch is weak while low pitch is strong (Bashi, central Africa; Merriam and Merriam, 1964; Eitan and Timmers, 2010). Like English, Dutch uses height metaphors for pitch; Farsi on the other hand uses thin and thick to refer to high and low pitch respectively (Dolscheid *et al.*, 2013), while Japanese uses both height and thickness metaphors. Dolscheid and colleagues (2013) found that this affects performance in a purely psychophysical task: when asked to sing back pure tones while looking at a screen, Dutch speakers reproduce higher pitch tones when a line is presented at the top of the screen at the same time as the tone, and lower pitches when a line is presented at the bottom of the screen. On the other hand, Farsi speakers display thickness interference: they produce a lower pitch when a thicker line is displayed, and a higher pitch when a thinner line is displayed. Dutch speakers initially don't display thickness interference, but after brief training on using thickness metaphors to talk about pitch, thickness interference appears in the tone reproduction task.

This shows a bi-directional cross-modal relationship: not only does how we perceive pitch affect how we talk about pitch, but how we talk about pitch can affect performance in a perceptual task. Despite not being the pitch metaphor of preference for all languages, visuospatial elevation does seem to map to pitch consistently across developmental stages (Walker *et al.*, 2018) and cultures (Parkinson *et al.*, 2012). Children who speak pitch-thickness languages, like Farsi and Turkish, are quicker to learn perceptual mappings between pitch and thickness than children who speak a pitch-height language like German (Shayan *et al.*, 2014). In other words, these correspondences both manifest in languages, and can be intensified or fine-tuned by language learning (Holler *et al.* 2022). Our capacity to make novel metaphors, and internalize them as conventions, is part of the same cross-modal cognitive suite underlying linguistic iconicity. The neurological mechanisms underpinning this connection will be discussed further in section 3.4.

3.3 Synesthesia

Synesthesia is a phenomenon wherein a stimulus in one modality (known as the *inducer*) not only triggers the usual sensory experience (e.g., hearing sound), but also a response from at least one additional modality, known as the *concurrent* (e.g., seeing color as a result of hearing sound). Importantly, for synesthetes, the relationship between inducer and

concurrent goes beyond the more widespread cross-modal correspondences described earlier in this section: the concurrent is automatic, and perceptually real (although this can take many forms; see van Leeuwen, 2013, for a discussion). For example, the most widely attested kind of synesthesia is colored graphemes, wherein participants reliably see printed letters in particular colors (Simner *et al.*, 2005). Various behavioral studies have shown this effect is genuine (Ward *et al.*, 2010), but there is also neurological evidence: grapheme-color synesthetes show activation in the V4 color area of the visual cortex within 110ms of the onset of a black and white grapheme stimulus (Brang *et al.*, 2010). While synesthesia was once thought to be rare and exhibit much higher prevalence among women, it is now understood to be fairly prevalent (occurring in approximately 5% of the population), with an even gender distribution (Simner and Carmichael, 2015). It also shows genetic patterns, with over 40% of synesthetes reporting at least one synesthetic relative in their immediate family (Barnett *et al.*, 2008) and some research suggesting genetic similarity between families with a higher prevalence of synesthetes (Tomson *et al.*, 2011).

The relationship between synesthesia and cross-modal correspondences has been a matter of debate. On the one hand, some have argued that human neonates are all strongly synesthetic, but that normal neural pruning creates division between sensory modalities in development such that most adults have cross-modal correspondences, but not synesthesia (Maurer, Gibson and Spector, 2013). On the other hand, many have argued that synesthesia and cross-modal correspondences should be considered distinct phenomena (Deroy and Spence, 2013). However, much of this debate surrounds the level of granularity at which we consider the phenomena. While the specific nature of synesthetic inducers and concurrents might be highly idiosyncratic, there do tend to be trends across synesthetes (e.g., for grapheme and color, *A* is more likely to be red, *B* is more likely to be blue), and these are echoed by shared cross-modal correspondences in the general population (Simner *et al.*, 2005; Cuskley *et al.*, 2019).

This “echo” is, however, less specific: synesthetes show strong stability *and* specificity of their associations over time. Shared cross-modal correspondences might be temporally stable (e.g., people associate the letter *A* with red reliably when tested at different times), but they are more relative: *A* is redder than other letters, but it is not a specific shade of red. For synesthetes, concurrents are both temporally stable and specific: they exhibit hue similarity to a far more specific degree than non-synesthetes (Rothen *et al.*, 2013), and this consistency can span decades (Simner and Logie, 2008). Overall, while synesthesia is rightly considered a

distinct phenomenon (defined primarily by the perceptual reality of automatically and involuntarily experiencing a concurrent), there is little doubt that it is fundamentally related to more general cross-modal correspondences in some important ways (Deroy and Spence, 2013; Bankieris and Simner, 2015).

The relative versus specific nature of cross-modal correspondences and synesthesia can also be thought of in terms of structure. In a vowel-color association task across over a thousand Dutch speakers, Cuskley and colleagues (2019) found that a relative minority demonstrated the consistency in choosing specific colors across trials generally used as a diagnostic criteria for synesthesia. However, a majority of participants – including most synesthetes – showed structure in their mappings: that is, vowels that sounded similar were likely to have similar colors, even if those colors weren't consistent across trials. This indicates that the cross-modal cognitive suite involves a general preference for cross-modal structure, but a flexibility with respect to the precise nature of that structure. Synesthetes have more definite and specific preferences, but these could play a key role in seeding the cross-modal correspondences in a population (Cuskley and Kirby, 2013).

There are at least 60 distinct varieties of synesthesia that have been documented, including tasting words, hearing colors, and seeing emotions (Simner, 2012). But within this diversity, a majority of attested synesthesias involve linguistic inducers; either in terms of graphemes, phonemes, or the lexicon (Simner, 2007). Most obviously, this points to a particularly important relationship between language and cross-modality. But it also underscores the key role played by learning even in such an automatic, perceptually real phenomenon – a phenomenon *so* real and automatic, that many synesthetes report being unaware that their experience was unusual (Auvray, Deroy and Matthen, 2015). Even proponents of a largely innate language faculty concede that language specific properties (like phoneme contrasts and lexical items) must be learned (Pleyer and Hartmann, 2019); indeed, the lexicon is one area of language which is plastic into adulthood and beyond. Given its patterns of heredity (Brang and Ramachandran, 2011), there does appear to be a strong genetic basis for synesthesia, but it is polygenetic and non-specific: it is a predisposition for a trait that is shaped by learning and the environment, not a specific cross-modal connection.

Grapheme-related synesthesias are a particularly interesting case for the role of learning in both synesthesia and cross-modal phenomena more generally. Not only are graphemes learned, but they are effortfully learned *after* the sound system of a language has been acquired. Graphemic mappings already entail an initially effortful cross-modal mapping between sound

and shape that eventually becomes so automatic that the sight of a written representation can activate a phonological one (Ziegler, Ferrand and Montant, 2004). Moreover, reading is not universally learned – not every contemporary language has a written form (signed languages tend not to, and many spoken languages lack one), and historically, literacy was far from ubiquitous across the population. In other words, graphemes are not necessarily a “normal” part of development. Nonetheless, they often seem to be how synesthesia is expressed; even for non-synesthetes, graphemes are an acquired automatic, perceptually real shape-sound correspondence. In short, language related synesthesias underscore the flexible nature of cross-modal phenomena, as a *capacity* for creating associations between modalities that adapt to and interact with individual experience.

3.4 Neurological underpinnings of the cross-modal cognitive suite

Neuroscientists have been aware of specialized areas of the brain related to experience in single modalities (e.g., somatosensory cortex for touch, the visual cortex in the occipital lobe, and auditory cortex within the temporal lobes; Purves *et al.*, 2004), for centuries. However, the neurological activity involved in multisensory experience cannot be pinpointed in the same way we can locate e.g., “vision” alone; rather, it seems to be a network of structures working together which integrate information from various subsystems (Sadaghiani and Wirsich, 2020). Despite its relatively late discovery, this distributed network is far from peripheral; in fact, many have argued that the integration of sensory and experiential information is a key component of what we experience as consciousness (Searle, 2004; Crick and Koch, 2005; Koch *et al.*, 2016).

Nonetheless, this cross-modal neurological network does seem to have an array of different “hubs” throughout the brain, including sensory specific cortices, as well as the superior temporal sulcus, the intraparietal sulcus, the insular claustrum, the frontal cortex, and the superior colliculus, among others (Calvert, 2001). Ettlinger and Wilson (1990) argue that these areas do not seem to store multi-sensory representations, rather, they “relay” cross-sensory connections. Particularly relevant to the evolution of cross-modal transfer is that these brain areas do not seem to be unique to humans, not only in terms of their presence, but in terms of their role in cross-modal tasks across a wide variety of species, including rats, cats, and monkeys (Crick and Koch, 2005).

While we shouldn't necessarily expect to locate a specific “cross-modality area” in the brain for reasons we will discuss shortly, some particular structures show interesting behavioral

and evolutionary patterns relevant to the cross-modal cognitive suite. The intraparietal sulcus (IPS) is a large groove in the cerebral cortex located in the parietal lobe (sulcus being Latin for “fold or wrinkle”), which plays a role in cross-modal correspondences (including cross-modal transfer), metaphor, and synesthesia. The IPS plays a key role as a relay of information for visual, spatial, and auditory systems in primates. This extends to a role in cross-modal correspondences more generally, particularly as these relate to cross-modal effects in speech perception (Kilian-Hütten, Formisano and Vroomen, 2017). The right IPS has been implicated in perceptual metaphor processing (Sadaghiani, Maier and Noppeney, 2009), and more specifically in novel metaphor production, as a kind of informational relay in a complex network between perceptual and executive function areas in the brain (Beaty, Silvia and Benedek, 2017). The IPS also seems to play a role in synesthesia. In a grapheme-colour task, grapheme-colour synesthetes have greater activation in the left IPS than non-synesthetes (Rouw and Scholte, 2010), and the IPS also shows notable activity in numerical-spatial synesthesias (Cohen Kadosh, Cohen Kadosh and Henik, 2007; Arend *et al.*, 2017).

Moreover, there is evidence for evolutionary changes to the IPS in humans. The degree of folding in a mammalian brain - also known as *gyrification* - is key to the expansion of available cortical surface area without necessarily increasing overall brain volume (Tallinen *et al.*, 2014). Gyrification within a confined cranial capacity increases cortical surface area not just by creating new folds, but also by deepening existing ones. The depth of the IPS groove is markedly enlarged in humans relative to macaques (Grefkes and Fink, 2005), and deeper in humans (22-26mm, Ludwig *et al.*, 2013) than chimpanzees (11-12mm, Bogart *et al.*, 2014). Simultaneously, these sulcal depth differences show lateral asymmetries, with greater differences in depth between left and right sides of the brain in chimpanzees relative to macaques (Bogart *et al.*, 2012; Sawada, 2020 for additional detail on lateralization). This asymmetry in IPS depth is further intensified between humans and chimpanzees and has been specifically connected to language and tool use (Binofski *et al.*, 2016). This asymmetry suggests that some hemispheric specialization related to cross-modality was present in our last common ancestor with chimpanzees. Overall, this points to a distributed network crucial for the cross-modal cognitive suite, with the IPS as a key relay: one which has undergone significant changes increasing cortical surface area particularly in humans.

While the IPS is perhaps a common theme, the distributed nature of this cross-modal network means that this is by no means the only brain area implicated in cross-modal experience (see e.g., Calvert, 2001 on cross-modal processing; Schmidt and Seger, 2009 on

metaphor; Rouw, Scholte and Colizoli, 2011 on synesthesia). Debates about the relationship between widespread cross-modal phenomena (including cross-modal correspondences and sensory metaphor) and synesthesia often focus on the diversity in these cross-modal experiences (Deroy and Spence, 2013). The variety of neural structures underpinning cross-modal transfer makes this system ripe for *exaptation*: “features that enhance fitness but were not built by natural selection for their current role” (Gould and Vrba, 1982, p. 4). Where brain areas are functionally (and thus, adaptively) specialized (e.g., as in the visual cortex for vision), those systems develop robustly and are developmentally canalized. On the other hand, systems which are more distributed tend to be more plastic: they can respond to learning and experience throughout the lifespan, and exhibit more individual and cultural variation (Barrett, 2012). This feature of cross-modal systems has made them a prime candidate for expansion and exaptation into a broader cross-modal cognitive suite.

This also underscores the argument that, at least from an evolutionary perspective, we should veer away from treating cross-modal experiences as confined traits with varying deep or superficial similarities. This is not to say we shouldn’t study phenomena like metaphor, synesthesia, or visuo-motor coordination in their own right, much like inhibition (one dimension of executive function) and visual encoding (one dimension of memory) still remain vital areas of study independent of the broader cognitive systems to which they belong. Rather, in evolutionary context, cross-modal experiences are more usefully framed as a cognitive suite or system, alongside other multidimensional cognitive systems like executive function (Gilbert and Burgess, 2008) and memory (Squire, 2004). Our cross-modal cognitive suite is not confined to specific behavior(s), but is a *capacity* to make associations across modalities with varying degrees of automaticity, perceptual activation, and abstraction. The next section outlines the potential selective pressures which might have led basic cross-modal transfer to have expanded into this broader system, particularly as a tool for iconicity, and an adaptation for communication.

4. From cross-modal to communicative

Sections 1 and 2 have introduced the relationship between linguistic iconicity and cross-modal transfer, while section 3 has detailed the much broader human cross-modal cognitive suite. Evidence regarding the cross-modal cognitive suite points to a generalized cross-modal network in the brain, potentially revolving around considerable growth in the depth of the intraparietal sulcus (IPS). We propose that this network was originally used primarily for basic

coordination tasks in cross-modal transfer, and was exapted for more generalized cross-modal correspondences. This system expanded behaviorally and neurologically, growing into both more abstract and general-purpose use for linguistic metaphor, and more idiosyncratic but perceptually real phenomena like synesthesia.

There is still a missing link here: exaptation requires selective pressure. While many animals (including humans) use cross-modal transfer to dynamically navigate the world around them as individuals, humans seem unique in their propensity to use cross-modality for social communication. What selective pressure led to its more social and communicative use in humans? Increasingly, researchers have come to appreciate the significant potential role of human self-domestication on our evolution, particularly as this pertains to many aspects of language, including social cognition and cultural transmission (Benítez-Burraco, Ferrettti, and Progovac 2021; Thomas and Kirby 2018). Self-domestication refers to the evolution of traits associated with domestication syndrome (including physical ones, like shorter muzzles and floppy ears in dogs, but also behavioral changes, such as increased prosociality).

Thomas and Kirby (2018) focus on a *conditions view* of domestication: what are the natural selective pressures created by domestication (self or otherwise)? Importantly, these are separate from artificial selection and selective breeding, and instead focus on the general environment of domestication, including “reduced living space, increased predictability of food and water supply, dietary changes, altered social structure, and greater availability of shelter ... [including] profound changes to an organism’s microclimate” (Thomas and Kirby, 2018, p. 15). Indeed, there is good evidence that the conditions of domestication induce similar genetic changes across diverse domesticated species (Theofanopoulou *et al.*, 2017). Primate self-domestication involved increased pressure for cooperation, selection against aggression, and greater prosociality in general (Hare, Wobber and Wrangham, 2012). Thomas and Kirby (2018) identify this changed phenotype in humans as leveraging vocal learning and cultural transmission combined with an emergent communicative intent to bolster the emergence of language.

It is important to acknowledge domestication as a gradual process, where a (sub)species becomes increasingly domesticated across generations. As such, it is difficult — and potentially misleading — to pinpoint an exact moment in time when humans became self-domesticated. Theofanopoulou *et al.* (2017) use genomic data and physical attributes to show that modern humans have genetic markers of domestication that are absent in Neanderthals. While this speaks to the relative domestication of humans and Neanderthals, it doesn’t preclude

the possibility of some self-domestication in earlier hominid species; indeed, there are some markers of self-domestication even in Bonobos (*Pan pansicus*) relative to common chimpanzees (*Pan troglodytes*, Hare, et al., 2012). Self-domestication likely intensified for *Homo sapiens* since the split from our last common ancestor with Neanderthals, approximately 550-765 thousand years ago (Meyer et al., 2016). In other words, humans are especially domesticated, but evidence shows our hominid relatives likely shared some of these traits. Neanderthals had a material culture not unlike contemporary humans (Hardy et al., 2020), which has been tied to neuroanatomical changes likely caused by self-domestication (Bruner and Gleeson, 2019). Neanderthals also share key features of vocal tract organization with humans that have been tied to complex vocalization (D’Anastasio et al., 2013; de Boer, 2009). Finally, recent evidence shows frequent and intense interaction between hominid species (Djakovic, Key and Soressi, 2022), including interbreeding (Villanea and Schraiber, 2019), making it unlikely that Neanderthals and humans had drastically different social organization.

Figure 1 shows a schematic of the relationships between the different aspects of the cross-modal cognitive suite, and a tentative suggestion as to how this relates to the split from our last common ancestor with our Neanderthal cousins. Regardless of when and to what extent self-domestication played a role in hominid evolution, self-domestication created significant pressure for humans to channel their experiences into communicative acts. This pressure then led to the expansion and exaptation of cross-modal transfer into the more unique human cross-modal cognitive suite, as this supported the learning and grounding of communicative systems. Exact dating of this process is difficult, and so it is impossible with current evidence to say whether (or to what extent) the cross-modal cognitive suite was exclusive to humans. However, it seems likely that these pressures somewhat affected the last common ancestor of *Homo sapiens* and Neanderthals more than 550 kya, or at least that Neanderthals responded to similar selective pressures in similar ways by undergoing some level of self-domestication and concurrent evolution of the cross-modal cognitive suite.

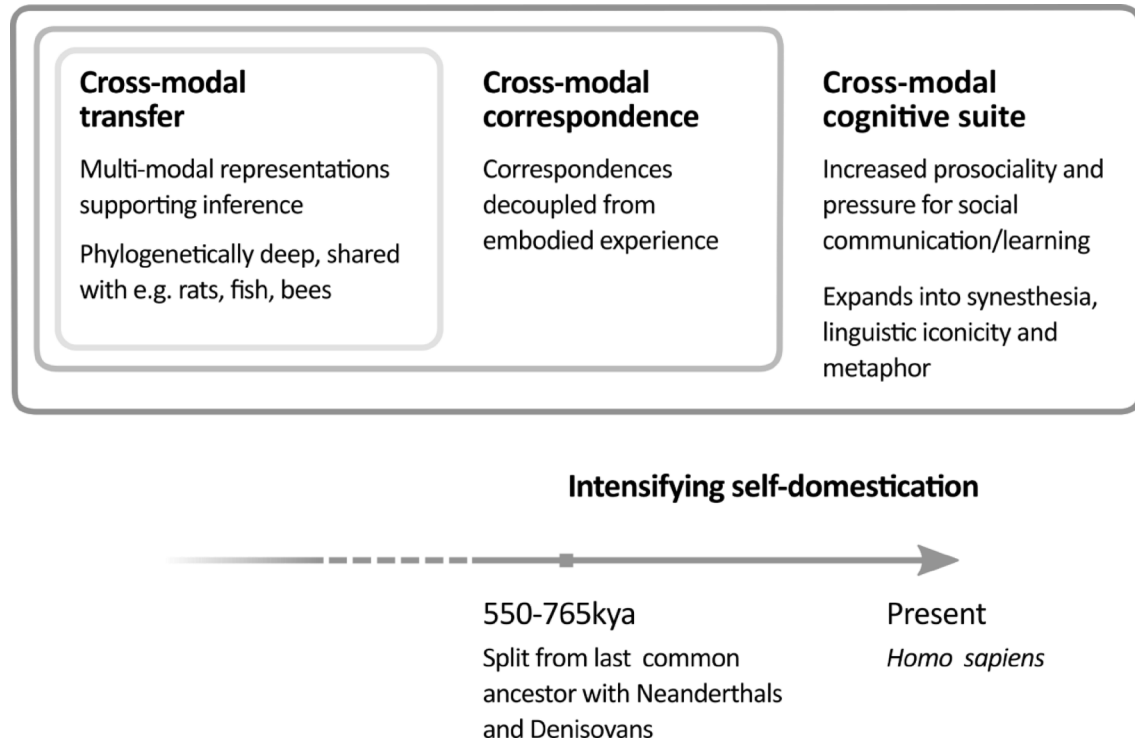


Figure 1: A sketch of the evolution of the cross-modal cognitive suite in the context of self-domestication, encompassing evolutionary precursors of cross-modal transfer and cross-modal correspondences.

The relationship between prosociality, self-domestication, and cross-modality plays an important role in the expansion to the cross-modal cognitive suite. Neurochemicals like serotonin (Popova *et al.*, 1991) are associated with the increased prosociality and cooperation at the core of self-domestication, and have also been shown to affect the cross-modal cognitive suite in humans. In synesthetes, the use of selective serotonin reuptake inhibitors (SSRIs) can disrupt synesthetic experience (Marek *et al.*, 2003; Brang and Ramachandran, 2008). In non-synesthetes, drugs which increase the release of serotonin in the brain, such as LSD and psilocybin, can induce temporary and genuine synesthetic experience (Luke *et al.*, 2022). These neurochemicals are found in higher levels in domesticated species, suggesting that they increase (partly) in support of prosocial behavior during domestication (Popova *et al.*, 1991). This indicates that the conditions of domestication not only provided a pressure for prosocial communicative acts, but also gave rise to neurochemicals that play a key role in the cross-modal cognitive suite. Sommer and colleagues (2022) measured prosociality alongside sensitivity to cross-modal correspondences and found that more prosocial individuals have stronger cross-modal correspondences.

There is further evidence for a relationship between self-domestication and the cross-modal cognitive suite in the brain. Bruner and Gleeson (2019) argue that self-domestication is specifically associated with expansion in the parietal lobe in both Neanderthals and humans (Bruner, 2018), including the IPS, which plays a key role in cross-modal cognition. While there is evidence of an increase in parietal brain volume in both Neanderthals and humans, the IPS specifically accounts for much of the variation between species – specifically, greater depth of the IPS in humans relative to Neanderthals (Bruner, 2010). Overall, during the time period when hominids would have begun the process of self-domestication (sometime around 550-765 thousand years ago), there is evidence of strong selection for behaviors associated with the IPS and the parietal lobe more generally, including “associative functions integrating visual, auditory, and spatial information with sensorimotor activity” (Coolidge *et al.*, 2015, p. 194; Holloway, 1996). This describes the core of the cross-modal cognitive suite underlying linguistic iconicity, which arguably played a key role in establishing lexical referential systems in the early stages of language evolution in terms of both grounding and learning (Cuskey and Kirby, 2013; Imai and Kita, 2014).

5. Conclusions

A growing body of work has suggested that linguistic iconicity played a key role in the evolution of language, particularly in its potential for grounding shared language in populations without an existing system, and for the role it plays in language learning. Work in this vein has extensively documented the proximate mechanisms and adaptive function of linguistic iconicity. This leaves a gap: where does the capacity for linguistic iconicity come from in our evolutionary history? This chapter has identified a key candidate for an evolutionary precursor to linguistic iconicity: cross-modal transfer. Cross-modal transfer is widespread in the animal kingdom, and involves using perceptual experience to make inferences; for example, pairing a high pitched sound with a smaller animal, even before you can see it.

In humans, self-domestication put selective pressure on this basic cross-modal system to expand and extend to cross-modal correspondences, which often have less obvious or direct mechanisms which connect modalities. For example, the suggested mechanisms behind associating more angular shapes with higher pitch involves a complex cross-modal chain. These kinds of cross-modal chains may well be the mechanism underlying any given cross-modal correspondence, but they indicate a step-change from more basic processes of single step cross-modal transfer, wherein mechanisms are more direct. This expansion of the cross-

modal suite entailed specific selection for linguistic cross-modal correspondences: our ability to readily associate multimodal communicative events (spoken, signed, or dual-channel) with perceptual experiences. In the context of self-domestication, more complex communication systems facilitated and intensified cooperation, putting further pressure on our cross-modal system to support communication.

This suite not only expanded to linguistic cross-modal correspondences, but also the related cross-modal phenomena of metaphor and synesthesia. Once a linguistic form is established via iconic grounding, the same cross-modal processes which facilitated perceptual grounding enabled metaphorical shifts and broadening, allowing language systems to expand (Cornejo *et al.*, 2009; Cuskley and Kirby, 2013; Elleström, 2017). The cross-modal cognitive suite not only allows us to readily *make* shared cross-modal correspondences, but also to *learn* them. The grounding function of iconicity gets a system started, but its advantages during learning are key for the continuation and expansion of a system. Synesthetes present a special case of “super” cross-modal associators, who have unusually stable, specific, and perceptually real cross-modal correspondences. While synesthetic correspondences show patterns, they are largely idiosyncratic in nature; but they may sow the seed for more widely shared cross-modal correspondences in a population. Because the ability to rapidly acquire new cross-modal mappings is part of the cross-modal cognitive suite, synesthetes may provide the spark a population needs to jumpstart a language system.

Language evolution is an inherently complex problem, with many facets beyond explaining how we came to have a complex shared vocabulary, including the function and acquisition of language structure, and the complex social cognition it entails. Nonetheless, understanding both the proximate and ultimate mechanisms underlying iconicity has strong explanatory power regarding the particularly human capacity for a large, open-ended system of shared symbols. Embedding iconicity within a wider cross-modal cognitive suite, and this within human self-domestication, begins to give us a clearer, more holistic picture of how and when language may have evolved.

References

- Adelman, James S., Zachary Estes, and Martina Cossu (2018). Emotional sound symbolism: Languages rapidly signal valence via phonemes. *Cognition* 175 (June): 122–30. <https://doi.org/10.1016/j.cognition.2018.02.007>.

- Ahlner, Felix, and Jordan Zlatev (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies* 38 (1/4): 298–348. <https://doi.org/10.12697/SSS.2010.38.1-4.11>.
- Arend, Isabel, Sarit Ashkenazi, Kenneth Yuen, Shiran Ofir, and Avishai Henik (2017). Automatic numerical-spatial association in synaesthesia: An fMRI investigation. *Neuropsychologia* 95 (January): 259–66. <https://doi.org/10.1016/j.neuropsychologia.2016.12.023>.
- Auvray, Malika, Ophelia Deroy, and Mohan Matthen (2015). How do synaesthetes experience the world? In *The Oxford Handbook of Philosophy of Perception*, 640–58. <https://doi.org/10.1093/oxfordhb/9780199600472.013.027>.
- Bankieris, Kaitlyn, and Julia Simmer (2015). What is the link between synaesthesia and sound symbolism? *Cognition* 136 (March): 186–95. <https://doi.org/10.1016/j.cognition.2014.11.013>.
- Barnett, Kylie J., Ciara Finucane, Julian E. Asher, Gary Bargary, Aiden P. Corvin, Fiona N. Newell, and Kevin J. Mitchell (2008). Familial patterns and the origins of individual differences in synaesthesia. *Cognition* 106 (2): 871–93. <https://doi.org/10.1016/j.cognition.2007.05.003>.
- Barrett, H. Clark (2012). A hierarchical model of the evolution of human brain specializations. *Proceedings of the National Academy of Sciences* 109 (supplement_1): 10733–40. <https://doi.org/10.1073/pnas.1201898109>.
- Bateson, Patrick, and Kevin N. Laland (2013). Tinbergen’s four questions: an appreciation and an update. *Trends in Ecology & Evolution* 28 (12): 712–18. <https://doi.org/10.1016/j.tree.2013.09.013>.
- Beaty, Roger E., Paul J. Silvia, and Mathias Benedek (2017). Brain networks underlying novel metaphor production. *Brain and Cognition* 111 (February): 163–70. <https://doi.org/10.1016/j.bandc.2016.12.004>.
- Benítez-Burraco, Antonio, Francesco Ferretti, and Ljiljana Progovac (2021). Human self-domestication and the evolution of pragmatics. *Cognitive Science* 45 (6): e12987. <https://doi.org/10.1111/cogs.12987>.
- Binkofski, Ferdinand Christoph, Juliane Klann, and Svenja Caspers (2016). On the neuroanatomy and functional role of the inferior parietal lobule and intraparietal sulcus, in Gregory Hickok and Steven L. Small (eds.), *Neurobiology of Language*. San Diego: Academic Press. 35–47. <https://doi.org/10.1016/B978-0-12-407794-2.00004-3>.
- Blasi, Damián E., Joseph Henrich, Evangelia Adamou, David Kemmerer, and Asifa Majid (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences* 26 (12): 1153–70. <https://doi.org/10.1016/j.tics.2022.09.015>.
- de Boer, Bart (2009). Acoustic analysis of primate air sacs and their effect on vocalization. *The Journal of the Acoustical Society of America* 126 (6): 3329–43. <https://doi.org/10.1121/1.3257544>.
- Bogart, Stephanie L., Allyson J. Bennett, Steven J. Schapiro, Lisa A. Reamer, and William D. Hopkins (2014). Different early rearing experiences have long-term effects on cortical organization in captive chimpanzees (*Pan Troglodytes*). *Developmental Science* 17 (2): 161–74. <https://doi.org/10.1111/desc.12106>.

- Bogart, Stephanie L., Jean-François Mangin, Steven J. Schapiro, Lisa Reamer, Allyson J. Bennett, Peter J. Pierre, and William D. Hopkins (2012). Cortical sulci asymmetries in chimpanzees and macaques: A new look at an old idea. *NeuroImage* 61 (3): 533–41. <https://doi.org/10.1016/j.neuroimage.2012.03.082>.
- Bozzi, P., and G. B. Flores D'Arcais (1967). Experimental study on intermodal relationships among expressive qualities. *Archivio Di Psicologia, Neurologia e Psichiatria* 28: 377–420.
- Brang, David, Edward M. Hubbard, Seana Coulson, Mingxiong Huang, and Vilayanur S. Ramachandran (2010). Magnetoencephalography reveals early activation of v4 in grapheme-color synesthesia. *NeuroImage* 53 (1): 268–74. <https://doi.org/10.1016/j.neuroimage.2010.06.008>.
- Brang, David, and Vilayanur S. Ramachandran (2008). Psychopharmacology of synesthesia; the role of serotonin 5-HT_{2A} receptor activation. *Medical Hypotheses* 70 (4): 903–4. <https://doi.org/10.1016/j.mehy.2007.09.007>.
- Brang, David, and Vilayanur S. Ramachandran (2011). Survival of the synesthesia gene: Why do people hear colors and taste words? *PLOS Biology* 9 (11): e1001205. <https://doi.org/10.1371/journal.pbio.1001205>.
- Bruner, Emiliano (2010). Morphological differences in the parietal lobes within the human genus: A neurofunctional perspective. *Current Anthropology* 51 (S1): S77–88. <https://doi.org/10.1086/650729>.
- Bruner, Emiliano (2018). Human paleoneurology and the evolution of the parietal cortex. *Brain, Behavior and Evolution* 91: 136–47. <https://doi.org/10.1159/000488889>.
- Bruner, Emiliano, and Ben T. Gleeson (2019). Body cognition and self-domestication in human evolution. *Frontiers in Psychology* 10. <https://www.frontiersin.org/article/10.3389/fpsyg.2019.01111>.
- Calvert, Gemma A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex* 11 (12): 1110–23. <https://doi.org/10.1093/cercor/11.12.1110>.
- Cohen Kadosh, Roi, Kathrin Cohen Kadosh, and Avishai Henik (2007). The neuronal correlate of bidirectional synesthesia: A combined event-related potential and functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience* 19 (12): 2050–59. <https://doi.org/10.1162/jocn.2007.19.12.2050>.
- Coolidge, Frederick L., Thomas Wynn, Karenleigh A. Overmann, and James M. Hicks (2015). Cognitive archaeology and the cognitive sciences, in Emiliano Bruner (ed.), *Human Paleoneurology*, Springer Series in Bio-/Neuroinformatics. Cham: Springer International Publishing. 177–208 https://doi.org/10.1007/978-3-319-08500-5_8.
- Cornejo, Carlos, Franco Simonetti, Agustín Ibáñez, Nerea Aldunate, Francisco Ceric, Vladimir López, and Rafael E. Núñez (2009). Gesture and metaphor comprehension: electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition* 70 (1): 42–52. <https://doi.org/10.1016/j.bandc.2008.12.005>.
- Cowey, Alan, and Lawrence Weiskrantz (1975). Demonstration of cross-modal matching in rhesus monkeys, *Macaca mulatta*. *Neuropsychologia* 13 (1): 117–20. [https://doi.org/10.1016/0028-3932\(75\)90057-3](https://doi.org/10.1016/0028-3932(75)90057-3).

- Crick, Francis C, and Christof Koch (2005). What is the function of the claustrum? *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1458): 1271–79. <https://doi.org/10.1098/rstb.2005.1661>.
- Cuskley, Christine (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics* 5 (1): 39–62. <https://doi.org/10.37693/pjos.2013.5.9651>.
- Cuskley, Christine, Mark Dingemanse, Simon Kirby, and Tessa M. van Leeuwen (2019). Cross-modal associations and synesthesia: Categorical perception and structure in vowel–color mappings in a large online sample. *Behavior Research Methods* 51 (4): 1651–75. <https://doi.org/10.3758/s13428-019-01203-7>.
- Cuskley, Christine, and Simon Kirby. 2013. Synesthesia, cross-modality, and language evolution, Julia Simner and Edward M. Hubbard (eds.), *The Oxford Handbook of Synesthesia*, Oxford University Press: Oxford. 869–99.
- Cuskley, Christine, Julia Simner, and Simon Kirby (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research* 81 (1): 119–30. <https://doi.org/10.1007/s00426-015-0709-2>.
- Ćwiek, Aleksandra, Susanne Fuchs, Christoph Draxler, Eva Liina Asu, Dan Dediu, Katri Hiovain, Shigeto Kawahara, et al. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* 377 (1841): 20200390. <https://doi.org/10.1098/rstb.2020.0390>.
- D’Anastasio, Ruggero, Stephen Wroe, Claudio Tuniz, Lucia Mancini, Deneb T. Cesana, Diego Dreossi, Mayoorendra Ravichandiran et al (2013). Micro-biomechanics of the Kebara 2 hyoid and its implications for speech in Neanderthals. *PLoS One* 8, no. 12: e82261.
- Davenport, R.K., C.M. Rogers, and I. Steele Russell (1973). Cross modal perception in apes. *Neuropsychologia* 11 (1): 21–28. [https://doi.org/10.1016/0028-3932\(73\)90060-2](https://doi.org/10.1016/0028-3932(73)90060-2).
- Demattè, M. Luisa, Daniel Sanabria, Rachel Sugarman, and Charles Spence (2006). Cross-modal interactions between olfaction and touch. *Chemical Senses* 31 (4): 291–300. <https://doi.org/10.1093/chemse/bjj031>.
- Deroy, Ophelia, and Charles Spence (2013). Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin & Review* 20 (4): 643–64. <https://doi.org/10.3758/s13423-013-0387-2>.
- Dingemanse, Mark, Marcus Perlman, and Pamela Perniss (2020). Construals of iconicity: Experimental approaches to form–meaning resemblances in language. *Language and Cognition* 12 (1): 1–14. <https://doi.org/10.1017/langcog.2019.48>.
- Djakovic, Igor, Alastair Key, and Marie Soressi (2022). Optimal linear estimation models predict 1400–2900 years of overlap between homo sapiens and neandertals prior to their disappearance from France and northern Spain. *Scientific Reports* 12 (1): 15000. <https://doi.org/10.1038/s41598-022-19162-z>.
- Dolscheid, Sarah, Shakila Shayan, Asifa Majid, and Daniel Casasanto (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science* 24 (5): 613–21. <https://doi.org/10.1177/0956797612457374>.
- D’Onofrio, Annette (2014). Phonetic detail and dimensionality in sound–shape correspondences: Refining the bouba–kiki paradigm. *Language and Speech* 57 (3): 367–93. <https://doi.org/10.1177/0023830913507694>.

- Eitan, Zohar, and Renee Timmers (2010). Beethovens last piano sonata and those who follow crocodiles: cross-domain mappings of auditory pitch in a musical context. *Cognition* 114 (3): 405–22. <https://doi.org/10.1016/j.cognition.2009.10.013>.
- Elleström, Lars (2017). Bridging the gap between image and metaphor through cross-modal iconicity: An interdisciplinary model, in Angelika Zirker, Matthias Bauer, Olga Fischer, and Christina Ljungberg (eds.). *Dimensions of Iconicity*. John Benjamins Publishing Company. 167–90.
- Ettlinger, G., and M. J. Jarvis. (1976). Cross-modal transfer in the chimpanzee. *Nature* 259 (5538): 44–46. <https://doi.org/10.1038/259044b0>.
- Ettlinger, G., and W. A. Wilson (1990). Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms. *Behavioural Brain Research* 40 (3): 169–92. [https://doi.org/10.1016/0166-4328\(90\)90075-P](https://doi.org/10.1016/0166-4328(90)90075-P).
- Gasser, Michael (2004). The origins of arbitrariness in language. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 26.
- Geschwind, Norman (1974). The development of the brain and the evolution of language, in Norman Geschwind (ed.) *Selected Papers on Language and the Brain*. Boston Studies in the Philosophy of Science. Dordrecht: Springer Netherlands. 86–104 https://doi.org/10.1007/978-94-010-2093-0_7.
- Gilbert, Sam J., and Paul W. Burgess (2008). Executive function. *Current Biology* 18 (3): R110–14. <https://doi.org/10.1016/j.cub.2007.12.014>.
- Gould, Stephen Jay, and Elisabeth S. Vrba (1982). Exaptation—a missing term in the science of form. *Paleobiology* 8 (1): 4–15. <https://doi.org/10.1017/S0094837300004310>.
- Grefkes, Christian, and Gereon R Fink (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy* 207 (1): 3–17. <https://doi.org/10.1111/j.1469-7580.2005.00426.x>.
- Hardy, Bruce L., M-H. Moncel, Céline Kerfant, Matthieu Lebon, Ludovic Bellot-Gurlet, and Nicolas Mélard (2020). Direct evidence of Neanderthal fibre technology and its cognitive and behavioral implications. *Scientific Reports* 10, no. 1: 1–9. <https://doi.org/10.1038/s41598-020-61839-w>.
- Hare, Brian, Victoria Wobber, and Richard Wrangham (2012). The self-domestication hypothesis: Evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour* 83 (3): 573–85. <https://doi.org/10.1016/j.anbehav.2011.12.007>.
- Hassemer, Julius, and Bodo Winter (2018). Decoding gestural iconicity. *Cognitive Science* 42 (8): 3034–49. <https://doi.org/10.1111/cogs.12680>.
- Hiraga, Masako (2004). *Metaphor and iconicity: A cognitive approach to analyzing texts*. Springer.
- Henrich, Joseph, Steven J. Heine, and Ara Norenzayan (2010). The Weirdest People in the World? *Behavioral and Brain Sciences* 33 (2–3): 61–83. <https://doi.org/10.1017/S0140525X0999152X>.
- Hinton, Leanne, Johanna Nichols, and John J. Ohala (1995). *Sound Symbolism*. Cambridge University Press.
- Hockett, Charles F. (1978). In search of Jove’s brow. *American Speech* 53 (4): 243–313. <https://doi.org/10.2307/455140>.

- Hodge, Gabrielle, and Lindsay Ferrara (2022). Iconicity as multimodal, polysemiotic, and plurifunctional. *Frontiers in Psychology* 13. <https://www.frontiersin.org/article/10.3389/fpsyg.2022.808896>.
- Holler, Judith, Linda Drijvers, Afrooz Rafiee, and Asifa Majid (2022). Embodied space-pitch associations are shaped by language. *Cognitive Science* 46 (2): e13083. <https://doi.org/10.1111/cogs.13083>.
- Holloway, Ralph L. (1996). Toward a synthetic theory of human brain evolution, in J.P. Changeaux and J. Chavallon (eds.), *Origins of the Human Brain*, Oxford, UK: Clarendon Press. 42-60.
- Imai, Mutsumi, and Sotaro Kita (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>.
- Jee, Hana, Monica Tamariz, and Richard Shillcock (2018). The substructure of phonics: the visual form of letters and their paradigmatic English pronunciation are systematically related. PsyArXiv preprint. <https://doi.org/10.31234/osf.io/n85mb>.
- Johansson, Niklas Erben, Andrey Anikin, and Nikolay Aseyev (2020). Color sound symbolism in natural languages. *Language and Cognition* 12 (1): 56–83. <https://doi.org/10.1017/langcog.2019.35>.
- Johansson, Niklas Erben, Jon W. Carr, and Simon Kirby (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>.
- Kantartzis, Katerina, Mutsumi Imai, and Sotaro Kita (2011). Japanese sound-symbolism facilitates word learning in english-speaking children. *Cognitive Science* 35 (3): 575–86. <https://doi.org/10.1111/j.1551-6709.2010.01169.x>.
- Kilian-Hütten, Niclas, Elia Formisano, and Jean Vroomen (2017). Multisensory integration in speech processing: Neural mechanisms of cross-modal aftereffects, in Maria Mody (ed.) *Neural Mechanisms of Language*. Innovations in Cognitive Neuroscience. Boston, MA: Springer US. 105–27 https://doi.org/10.1007/978-1-4939-7325-5_6.
- Koch, Christof, Marcello Massimini, Melanie Boly, and Giulio Tononi (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience* 17 (5): 307–21. <https://doi.org/10.1038/nrn.2016.22>.
- Köhler, Wolfgang (1929). *Gestalt Psychology*. New York, NY: Livewright.
- van Leeuwen, Tessa M. (2013). Individual differences in synesthesia, in Julia Simner and Edward M. Hubbard (eds.) *Oxford Handbook of Synesthesia*. Oxford, UK: Oxford University Press. 241–64 <https://doi.org/10.1093/oxfordhb/9780199603329.013.0013>.
- Lewkowicz, David J., and Gerald Turkewitz (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology* 16 (6): 597–607. <https://doi.org/10.1037/0012-1649.16.6.597>.
- Lilienfeld, Scott O., Kathryn C. Sauvigné, Steven Jay Lynn, Robin L. Cautin, Robert D. Latzman, and Irwin D. Waldman (2015). Fifty psychological and psychiatric terms to avoid: a list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases. *Frontiers in Psychology* 6. <https://www.frontiersin.org/article/10.3389/fpsyg.2015.01100>.

- Lockwood, Gwilym, Mark Dingemanse, and Peter Hagoort (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42 (8): 1274–81. <https://doi.org/10.1037/xlm0000235>.
- Ludwig, K. U., P. Sämman, M. Alexander, J. Becker, J. Bruder, K. Moll, D. Spieler, et al. (2013). A common variant in myosin-18b contributes to mathematical abilities in children with dyslexia and intraparietal sulcus variability in adults. *Translational Psychiatry* 3 (2): e229. <https://doi.org/10.1038/tp.2012.148>.
- Luke, David P., Laura Lungu, Ross Friday, and Devin B. Terhune (2022). The chemical induction of synaesthesia. *Human Psychopharmacology: Clinical and Experimental*, e2832. <https://doi.org/10.1002/hup.2832>.
- Mameli, Matteo, and Patrick Bateson. 2011. An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366 (1563): 436–43. <https://doi.org/10.1098/rstb.2010.0174>.
- Marek, Gerard J., Linda L. Carpenter, Christopher J. McDougle, and Lawrence H. Price (2003). Synergistic action of 5-HT_{2A} antagonists and selective serotonin reuptake inhibitors in neuropsychiatric disorders. *Neuropsychopharmacology* 28 (2): 402–12. <https://doi.org/10.1038/sj.npp.1300057>.
- Marks, Lawrence E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance* 13 (3): 384–94. <https://doi.org/10.1037/0096-1523.13.3.384>.
- Maurer, Daphne, Laura C. Gibson, and Ferrinne Spector. 2013. Synesthesia in infants and very young children, in Julia Simner and E. M. Hubbard (eds.) *Oxford Handbook of Synesthesia*, Oxford, UK: Oxford University Press. 46–62 <https://doi.org/10.1093/oxfordhb/9780199603329.013.0003>.
- Maurer, Daphne, Thanujeni Pathman, and Catherine J. Mondloch (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science* 9 (3): 316–22. <https://doi.org/10.1111/j.1467-7687.2006.00495.x>.
- Merriam, Alan P., and Valerie Merriam (1964). *The Anthropology of Music*. Chicago, IL: Northwestern University Press.
- Meyer, Matthias, Juan-Luis Arsuaga, Cesare de Filippo, Sarah Nagel, Ayinuer Aximu-Petri, Birgit Nickel, Ignacio Martínez, et al. (2016). Nuclear DNA sequences from the middle pleistocene Sima de los Huesos Hominins. *Nature* 531 (7595): 504–7. <https://doi.org/10.1038/nature17405>.
- Mondloch, C. J., and D. Maurer (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience* 4 (2): 133–36. <https://doi.org/10.3758/CABN.4.2.133>.
- Murgiano, Margherita, Yasamin Motamedi, and Gabriella Vigliocco (2021). Situating language in the real-world: The role of multimodal iconicity and indexicality. *Journal of Cognition* 4 (1): 38. <https://doi.org/10.5334/joc.113>.
- Nielsen, Alan KS, and Mark Dingemanse (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech* 64 (1): 52–72. <https://doi.org/10.1177/0023830920914339>.
- Nielsen, Alan, and Drew Rendall (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of*

- Experimental Psychology/Revue Canadienne de Psychologie Expérimentale* 65 (2): 115–24. <https://doi.org/10.1037/a0022268>.
- Nielsen, Alan and Drew Rendall (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition* 4 (2): 115–25. <https://doi.org/10.1515/langcog-2012-0007>.
- Occhino, Corrine, Benjamin Anible, Erin Wilkinson, and Jill P. Morford (2017). Iconicity is in the eye of the beholder: How language experience affects perceived iconicity. *Gesture* 16 (1): 100–126. <https://doi.org/10.1075/gest.16.1.04occ>.
- Ohala, John J. (1995). The frequency code underlies the sound-symbolic use of voice pitch. In Leanne Hinton, Johanna Nichols and John J. Ohala (eds.), *Sound Symbolism*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511751806.022>. 325–47
- Ohtake, Yuka, and Etsuko Haryu (2013). Investigation of the Process Underpinning Vowel-Size Correspondence. *Japanese Psychological Research* 55 (4): 390–99. <https://doi.org/10.1111/jpr.12029>.
- Ortega, Gerardo, and Aslı Özyürek (2020). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods* 52 (1): 51–67. <https://doi.org/10.3758/s13428-019-01204-6>.
- Over, Ray, and N. J. Mackintosh (1969). Cross-modal transfer of intensity discrimination by rats. *Nature* 224 (5222): 918–19. <https://doi.org/10.1038/224918a0>.
- Ozturk, Ozge, Madelaine Krehm, and Athena Vouloumanos (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology* 114 (2): 173–86. <https://doi.org/10.1016/j.jecp.2012.05.004>.
- Parise, Cesare V., Katharina Knorre, and Marc O. Ernst (2014). Natural auditory scene statistics shape human spatial hearing. *Proceedings of the National Academy of Sciences* 111 (16): 6104–8. <https://doi.org/10.1073/pnas.1322705111>.
- Parise, Cesare V., and Charles Spence (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research* 220 (3): 319–33. <https://doi.org/10.1007/s00221-012-3140-6>.
- Parkinson, Carolyn, Peter Jes Kohler, Beau Sievers, and Thalia Wheatley (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception* 41 (7): 854–61. <https://doi.org/10.1068/p7225>.
- Perlman, Marcus, Rick Dale, and Gary Lupyan (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2 (8): 150152. <https://doi.org/10.1098/rsos.150152>.
- Perniss, Pamela, and Gabriella Vigliocco (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.
- Pleyer, Michael, and Stefan Hartmann (2019). Constructing a consensus on language evolution? Convergences and differences between biolinguistic and usage-based approaches. *Frontiers in Psychology* 10 (November): 2537. <https://doi.org/10.3389/fpsyg.2019.02537>.

- Popova, Nina K., Ninel N. Voitenko, A. V. Kulikov, and D. F. Avgustinovich (1991). Evidence for the involvement of central serotonin in mechanism of domestication of silver foxes. *Pharmacology Biochemistry and Behavior* 40 (4): 751–56. [https://doi.org/10.1016/0091-3057\(91\)90080-L](https://doi.org/10.1016/0091-3057(91)90080-L).
- Purves, Dale, George J. Augustine, David Fitzpatrick, William C. Hall, Anthony-Samuel LaManita, McNamara, and S. Mark Williams, eds. (2004). *Neuroscience*. 3rd ed. Sunderland, MA: Sinauer Associates.
- Ramachandran, Vilayanur S (2004). *A Brief Tour of Human Consciousness: From Impostor Poodles to Purple Numbers*. Pi Press.
- Rasenberg, Marlou, Asli Özyürek, Sara Bögels, and Mark Dingemanse (2022). The primacy of multimodal alignment in converging on shared symbols for novel referents. *Discourse Processes* 59 (3): 209–36. <https://doi.org/10.1080/0163853X.2021.1992235>.
- Rose, Susan (2013). From hand to eye: findings and issues in infant cross-modal transfer, in David J. Lewkowicz and Robert Lickliter (eds.), *The Development of Intersensory Perception: Comparative Perspectives*, Psychology Press, 265–84.
- Rothen, Nicolas, Anil K. Seth, Christoph Witzel, and Jamie Ward (2013). Diagnosing synaesthesia with online colour pickers: Maximising sensitivity and specificity. *Journal of Neuroscience Methods* 215 (1): 156–60. <https://doi.org/10.1016/j.jneumeth.2013.02.009>.
- Rouw, Romke, and H. Steven Scholte (2010). Neural basis of individual differences in synesthetic experiences. *Journal of Neuroscience* 30 (18): 6205–13. <https://doi.org/10.1523/JNEUROSCI.3444-09.2010>.
- Rouw, Romke, H. Steven Scholte, and Olympia Colizoli (2011). Brain areas involved in synaesthesia: A review. *Journal of Neuropsychology* 5 (2): 214–42. <https://doi.org/10.1111/j.1748-6653.2011.02006.x>.
- Sadaghiani, Sepideh, Joost X. Maier, and Uta Noppeney (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *Journal of Neuroscience* 29 (20): 6490–99. <https://doi.org/10.1523/JNEUROSCI.5437-08.2009>.
- Sadaghiani, Sepideh, and Jonathan Wirsich (2020). Intrinsic connectome organization across temporal scales: New insights from cross-modal approaches. *Network Neuroscience* 4 (1): 1–29. https://doi.org/10.1162/netn_a_00114.
- Saito, Daisuke N., Tomohisa Okada, Yusuke Morita, Yoshiharu Yonekura, and Norihiro Sadato (2003). Tactile–visual cross-modal shape matching: A functional MRI study. *Cognitive Brain Research* 17 (1): 14–25. [https://doi.org/10.1016/S0926-6410\(03\)00076-4](https://doi.org/10.1016/S0926-6410(03)00076-4).
- Sapir, Edward (1929). A study in phonetic symbolism. *Journal of Experimental Psychology* 12 (3): 225–39. <https://doi.org/10.1037/h0070931>.
- Sawada, Kazuhiko (2020). Cerebral sulcal asymmetry in macaque monkeys. *Symmetry* 12 (9): 1509. <https://doi.org/10.3390/sym12091509>.
- Schmidt, Gwenda L., and Carol A. Seger (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition* 71 (3): 375–86. <https://doi.org/10.1016/j.bandc.2009.06.001>.

- Schumacher, Sarah, Theresa Burt de Perera, Johanna Thenert, and Gerhard von der Emde (2016). Cross-modal object recognition and dynamic weighting of sensory inputs in a fish. *Proceedings of the National Academy of Sciences* 113 (27): 7638–43. <https://doi.org/10.1073/pnas.1603120113>.
- Searle, J. R. (2004). *Mind: A Brief Description*. New York, Oxford University Press.
- Shayan, Shakila, Ozge Ozturk, Melissa Bowerman, and Asifa Majid (2014). Spatial metaphor in language can promote the development of cross-modal mappings in children. *Developmental Science* 17 (4): 636–43. <https://doi.org/10.1111/desc.12157>.
- Sidhu, David M., and Penny M. Pexman (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review* 25 (5): 1619–43. <https://doi.org/10.3758/s13423-017-1361-1>.
- Simner, Julia (2007). Beyond perception: Synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Sciences* 11 (1): 23–29. <https://doi.org/10.1016/j.tics.2006.10.010>.
- Simner, Julia (2012). Defining synaesthesia. *British Journal of Psychology* 103 (1): 1–15. <https://doi.org/10.1348/000712610X528305>.
- Simner, Julia, and Duncan A. Carmichael (2015). Is synaesthesia a dominantly female trait? *Cognitive Neuroscience* 6 (2–3): 68–76. <https://doi.org/10.1080/17588928.2015.1019441>.
- Simner, Julia, Christine Cuskley, and Simon Kirby (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*, 39(4): 553–569. <https://doi.org/10.1068/p6591>
- Simner, Julia, and Robert H. Logie (2008). Synaesthetic consistency spans decades in a lexical–gustatory synaesthete. *Neurocase* 13 (5–6): 358–65. <https://doi.org/10.1080/13554790701851502>.
- Simner, Julia, Jamie Ward, Monika Lanz, Ashok Jansari, Krist Noonan, Louise Glover, and David A. Oakley (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology* 22 (8): 1069–85. <https://doi.org/10.1080/02643290500200122>.
- Slobodenyuk, Nadiya, Yasmina Jraissati, Ali Kanso, Lama Ghanem, and Imad Elhajj (2015). Cross-modal associations between color and haptics. *Attention, Perception, & Psychophysics* 77 (4): 1379–95. <https://doi.org/10.3758/s13414-015-0837-1>.
- Slocombe, B. G., Duncan A. Carmichael, and Julia Simner (2016). Cross-modal tactile–taste interactions in food evaluations. *Neuropsychologia*, Special Issue: Synaesthesia and Multisensory Processes, 88 (July): 58–64. <https://doi.org/10.1016/j.neuropsychologia.2015.07.011>.
- Solvi, Cwyn, Selene Gutierrez Al-Khudhairy, and Lars Chittka (2020). Bumble bees display cross-modal object recognition between visual and tactile senses. *Science* 367 (6480): 910–12. <https://doi.org/10.1126/science.aay8064>.
- Sommer, Kees, Christine Cuskley, Simon Kirby, Tessa Verhoef (2022). How Self-Domestication and Prosociality May Shape Cross-Modal Language. *Proceedings of the Joint Conference on Language Evolution (JCoLE)*. doi:10.17617/2.3398549.
- Speed, Laura J., Hannah Atkinson, Ewelina Wnuk, and Asifa Majid (2021). The sound of smell: Associating odor valence with disgust sounds. *Cognitive Science* 45(5): e12980.

- Spence, Charles (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics* 73 (4): 971–95. <https://doi.org/10.3758/s13414-010-0073-7>.
- Spence, Charles, and Ophelia Deroy (2012). Crossmodal correspondences: Innate or learned? *i-Perception* 3 (5): 316–18. <https://doi.org/10.1068/i0526ic>.
- Spike, Matthew (2017). The evolution of linguistic rules. *Biology & Philosophy* 32 (6): 887–904. <https://doi.org/10.1007/s10539-018-9610-x>.
- Squire, Larry R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, Multiple Memory Systems, 82 (3): 171–77. <https://doi.org/10.1016/j.nlm.2004.06.005>.
- Stone, Ruth M. (1981). Toward a Kpelle conceptualization of music performance. *The Journal of American Folklore* 94 (372): 188–206. <https://doi.org/10.2307/540124>.
- Tallinen, Tuomas, Jun Young Chung, John S. Biggins, and L. Mahadevan (2014). Gyrification from constrained cortical expansion. *Proceedings of the National Academy of Sciences* 111 (35): 12667–72. <https://doi.org/10.1073/pnas.1406015111>.
- Theofanopoulou, Constantina, Simone Gastaldon, Thomas ORourke, Bridget D. Samuels, Angela Messner, Pedro Tiago Martins, Francesco Delogu, Saleh Alamri, and Cedric Boeckx (2017). Self-domestication in homo sapiens: Insights from comparative genomics. *PLOS ONE* 12 (10): e0185306. <https://doi.org/10.1371/journal.pone.0185306>.
- Thomas, James, and Simon Kirby (2018). Self domestication and the evolution of language. *Biology & Philosophy* 33 (1): 9. <https://doi.org/10.1007/s10539-018-9612-8>.
- Thompson, Patrick D., and Zachary Estes (2011). Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology* 64 (12): 2392–2404. <https://doi.org/10.1080/17470218.2011.605898>.
- Tinbergen, Nikolaas (1963). On aims and methods of ethology. *Zeitschrift Für Tierpsychologie* 20 (4): 410–33. <https://doi.org/10.1111/j.1439-0310.1963.tb01161.x>.
- Tomson, Steffie N., Nili Avidan, Kwanghyuk Lee, Anand K. Sarma, Rejnal Tushe, Dianna M. Milewicz, Molly Bray, Suzanne M. Leal, and David M. Eagleman (2011). The genetics of colored sequence synesthesia: Suggestive evidence of linkage to 16q and genetic heterogeneity for the condition. *Behavioural Brain Research* 223 (1): 48–52. <https://doi.org/10.1016/j.bbr.2011.03.071>.
- Vigliocco, Gabriella, Pamela Perniss, and David Vinson (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 20130292. <https://doi.org/10.1098/rstb.2013.0292>.
- Villanea, Fernando A., and Joshua G. Schraiber (2019). Multiple episodes of interbreeding between Neanderthal and modern humans" *Nature Ecology & Evolution* 3(1): 39-44. <https://doi.org/10.1038/s41559-018-0735-8>.
- Walker, Peter, James Gavin Bremner, Marco Lunghi, Sarah Dolscheid, Beatrice D. Barba, and Francesca Simion (2018). Newborns are sensitive to the correspondence between auditory pitch and visuospatial elevation. *Developmental Psychobiology* 60 (2): 216–23. <https://doi.org/10.1002/dev.21603>.

- Ward, Jamie, Clare Jonas, Zoltan Dienes, and Anil Seth (2010). Grapheme-colour synaesthesia improves detection of embedded shapes, but without pre-attentive “pop-out” of synaesthetic colour. *Proceedings of the Royal Society B: Biological Sciences* 277 (1684): 1021–26. <https://doi.org/10.1098/rspb.2009.1765>.
- Westbury, Chris. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language* 93(1): 10-19. <https://doi.org/10.1016/j.bandl.2004.07.006>
- Westbury, Chris, Geoff Hollis, David M. Sidhu, and Penny M. Pexman (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>.
- Woodin, Greg, Bodo Winter, Marcus Perlman, Jeannette Littlemore, and Teenie Matlock (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>.
- Zbikowski, Lawrence M. (1998). Metaphor and music theory: Reflections from cognitive science. *Music Theory Online* 4 (1): 1–8.
- Ziegler, Johannes C., Ludovic Ferrand, and Marie Montant (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory & Cognition* 32 (5): 732–41. <https://doi.org/10.3758/BF03195863>.