Oxford Handbooks Online

Synesthesia, Cross-Modality, and Language Evolution

Christine Cuskley and Simon Kirby

Oxford Handbook of Synesthesia
Edited by Julia Simner and Edward Hubbard

Print Publication Date: Dec 2013
Subject: Psychology, Cognitive Neuroscience, Cognitive Psychology
Online Publication Date: Dec 2013 DOI: 10.1093/oxfordhb/9780199603329.013.0043

Abstract and Keywords

In this chapter we suggest that the origin of language, specifically the protolinguistic stage, was iconic rather than arbitrary, and fundamentally based on shared cross-modal associations. We provide evidence from natural language in the form of sound symbolism, distinguishing conventional sound symbolism from sensory sound symbolism. Sensory sound symbolism, or the presence of iconicity in natural language, is considered alongside psychological experiments in naming, and other investigations of cross-modal associations specifically involving linguistic sound. This evidence supports the idea that we can directly express a variety of sensory experiences through linguistic sound, and thus that language systems have the capacity to be iconic to a large extent. We outline a theory of sensory protolanguage emergence, considering how and why arbitrariness eventually became the dominant relationship between form and meaning. Finally, we consider the possibility of continuity between cross-modal transfer in other primates and more abstract cross-modal associations among humans, suggesting this may have scaffolded the emergence of symbolic communication.

Keywords: Cross-modality, iconicity, language, language evolution, protolanguage, ideophones

Introduction

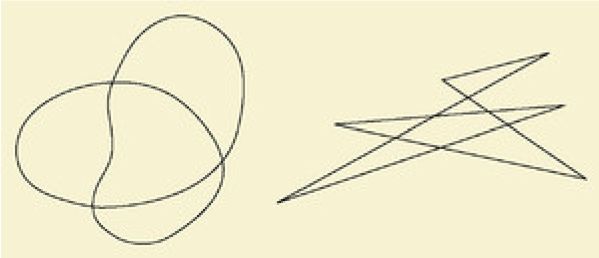
A central question for language evolution concerns the origin of linguistic symbols. Our species appears to be unique in possessing a learned system of arbitrary reference. Put differently, linguistic forms (words) are traditionally connected to their meanings only through convention, rather than by any goodness-of-fit; a rose, by another name, would still smell as sweet. The French word *pomme* expresses the concept “apple” equally as well as the Italian term *mela*. How can we explain the emergence of a system wherein the forms have no apparent connection to their meanings? This question is perhaps best framed by Harnad (1990) as the "symbol grounding problem." Harnad uses the example of a Chinese dictionary (Searle 1980): if an English speaker searches for the meanings of Chinese words, but where each word is defined only by other Chinese words, the English speaker cannot access any meanings. A symbol grounded only within a system of other symbols has no clear origin. How did this web of symbols begin?

In this chapter we will suggest a solution to the symbol grounding problem. Although words may be arbitrary in modern languages, we present a theory that words did indeed evolve, at least to some extent, based on a cross-modal goodness-of-fit: that is, in the origin of language, the sounds of words evoked the direct meaning of referents. Drawing on evidence from natural language systems as well as cross-modal associations, we will show that language users make implicit cross-modal associations between the sound quality of words and other sensory properties of the objects they denote. The primary evidence for this theory will draw on both natural language and experimental evidence, showing the potential for sound-to-meaning goodness-of-fit in language. Experimental evidence will focus specifically on the fact that humans have certain cross-sensory preferences for relating linguistic sound to other sensory dimensions possible (e.g., taste, (p. 870) shape), and that these a priori preferences may have guided the choice of word forms in language evolution. To this end, we will turn first to exactly how cross-modality can provide the basis for the emergence of such a system.

What cross-modality gives us

The lexicons on which languages are built are essentially systems of symbols. Symbol manipulation has been identified as a uniquely human faculty (Deacon 1997), as well as an essential feature of language, but there is no clear account of how our cognition came to support a symbol system that is seemingly disembodied: forms have no connections to their meanings. We will propose that shared sensory biases in the general population provided a basis for grounding language in our perceptual system and served to bootstrap a small lexicon which was initially less arbitrary than its evolutionarily later forms.

Cross-modal biases have been demonstrated experimentally across a range of modalities, including between taste and touch (Christensen 1980), pitch, and vowel quality (Crisinel and Spence 2009; Simner, Cuskley, and Kirby 2010), and touch and color (Simner and Ludwig 2009), among many others (see Part VII of this volume, as well as Calvert, Spence, and Stein 2004, for a review). The pervasiveness of cross-modal associations would suggest this might have been an important utility in an evolutionary sense. Connections between visual and motor areas, for example, would have allowed for complex motor actions guided by vision, such as seeing a branch in the visual field and preparing an appropriate grasp action in order to swing to it (Ramachandran 2004).



Click to view larger

Figure 43.1 Maluma and takete shapes. Reproduced from Köhler, Wolfgang, *Gestalt Psychology*. 1e ©1929, Liveright.)

This view, of cross-modality as part of the normal cognitive suite, puts synesthesia at one end of a spectrum of cross-modal abilities.[1] Synesthetes, like all people, make cross-modal connections, although these are active for synesthetes at a very different level. In other words, where non-synesthetes might make an intuitive cross-modal association between certain sounds and certain shapes, synesthetes would have a fundamentally different experience of actually "seeing" these colors in the literal perceptual sense. Moreover, the associations of synesthetes are stronger and more temporally stable than run-of-the-mill cross-modal associations. Although particular synesthetic perceptions are considered unique to any particular synesthete, there are definite observable trends, most notably in grapheme-color synesthetes. For example, the letter *A* is likely across synesthetes to be red above all other colors (Simner et al. 2005), and similar letter shapes evokes similar colors for grapheme-color synesthetes (Brang et al. 2011). Trends among synesthetes are also echoed by trends in cross-modality in the general population for grapheme-color associations (Simner et al. 2005), pitch-lightness associations (Ward, Huckstep, and Tsakanikos 2006), and touch-vision associations (Simner and Ludwig 2009). From this perspective, synesthetes can be viewed as super cross-modal associators, with unusually strong, stable, and specific cross-modal biases. (p. 871)

The idea that language may have non-arbitrary origins is not a particularly new one, although the notion of cross-modality as the basis for such a system is a newer idea. So-called "bow-wow" theories of language origins are some of the earliest on record, suggesting that language was built on direct imitations of sounds (see Atchison 2000 for a review). However, such theories have typically not been taken seriously by the field of linguistics, which emphasizes the fundamental arbitrariness of the sign (Saussure 1959), thus rejecting any natural connection between linguistic signs and their denotations in the world. Hockett's (1960) influential *Design Features of Natural Language*, particularly the feature of arbitrariness, further solidified this view. Any evolutionary perspectives eschewing arbitrariness—now considered a central and essential feature of language—were hardly going to be popular.

Despite this, Gestalt psychologists began experiments in motivated naming: exploring the possibility that certain names may better fit certain referents. The best known of these experiments was reported by Köhler (1929, 1947). In a simple forced choice task, subjects were given two nonsense words, *takete* and *maluma*, and two abstract shapes (shown in Figure 43.1), and asked to match the words to the shapes. Similar experiments have since been performed with children (Irwin and Newland 1940; Maurer, Pathman, and Mondloch

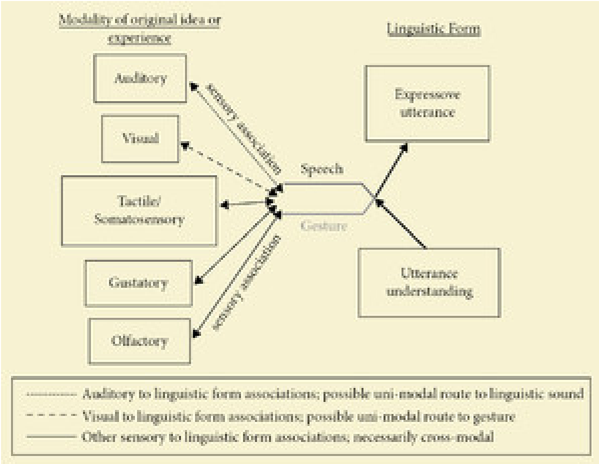
2006), as well as cross-culturally (Davis 1961) and with varying stimuli (Ramachandran and Hubbard 2001). The results of all of these experiments were rather striking: there appeared to be a strong and quantifiable bias in naming. In the vast majority of cases,[2] people assigned *maluma* to the curved figure and *takete* to the angular one, demonstrating a preference to map certain names to certain visual forms.

These experiments demonstrated that linguistic sounds have the potential for a "natural" connection to their referents. This connection is not limited to the uni-modal (i.e., within only the auditory modality) "bow-wow" variety of sound imitation, but holds for cross-sensory connections as well: the Köhler task demonstrated a common cross-sensory association between visual form and linguistic sound. Several authors hinted that this naming bias might be important in language (Köhler 1929, 1947; Werner 1957; Werner and Wapner 1952). Ramachandran and Hubbard (2001) were the first to (p. 872) explicitly relate the task to language evolution with their "synesthetic bootstrapping theory of language origins" (2001, 15), suggesting what shared visuo-motor mappings linked the articulation of certain linguistic sounds to certain visual properties. Specifically, they suggested there may be direct cross-activation between visual object shape and sound contours represented in the auditory cortex.

Cross-modal biases are pervasive, and biases to associate linguistic sounds with other sensory experiences would have been particularly useful for the evolution of spoken language. Linguistic cross-modal biases would allow the expression of our perceptions and experiences in any sensory domain through a single linguistic channel. Thus, we can expand upon antiquated "bow-wow" theories of language evolution, with a theory which allows both direct auditory imitation and cross-sensory imitation. Languages still use sounds to express concepts whose characteristics are auditory (e.g., *hiss* or *buzz*), but this also extends more widely to referents whose characteristics are not auditory. We can express various sensory experiences through a linguistic channel, and the fact that cross-modal biases are shared across all people would allow for the mutual understanding of an utterance formed on their basis (see Figure 43.2 which gives a schematic representation of how shared cross-sensory biases might have facilitated mutual comprehension in language evolution). In addition to this, recent research suggests that speech is not the only tool available to us in language. Language is a dual system, in which gestures play an integral role (Brown 2010; McNeill 2005; Tomasello 2008). Cross-modal biases, combined with a dual channel speech-gesture system, would have provided ample opportunity for an iconic system grounded in perception. In the emergence of language, we propose that these biases provided a perceptual grounding for a small lexicon, or *protolanguage* (see Arbib 2005; Bickerton 1990; Tallerman 2007; Wray 1998), which matured and expanded into the complex arbitrary systems we see today.

In the next section, we will review evidence for this theory as found in natural language systems, looking to modern languages for evidence of a connection between form and meaning.

Non-Arbitrariness in Modern Natural Language



Click to view larger

Figure 43.2 Figure shows what cross-modality affords us in theories of language evolution. Linguistic symbols as grounded in perceptual experiences, represented on the left (auditory, visual, tactile/somatosensory, gustatory, and olfactory). Shared cross-modal biases allow for common sensory associations between modalities, making the expression and understanding of varied aspects of experience possible through linguistic form. Shared associations are leveraged in the formation of an utterance as well as in utterance interpretation. Linguistic form is realized through a dual modality speech-gesture system; a view particularly relevant to language evolution, which likely occurred primarily in the context of face-to-face interactions. This dual system allows some ideas or events to be expressed through unimodal associations; most notably, gestures afford iconicity for visual events, and speech affords iconicity for auditory events.

Our theory generates a set of testable expectations. Firstly, we should expect to find some evidence of motivated forms in modern natural language—i.e., words that sound as if they fit what they denote, in a cross-sensory sense. But we quickly run up against the central tenet of linguistics discussed in the previous section: that word–meaning relationships are apparently arbitrary. However, a theory of sensory protolanguage—and the expectation that motivated forms should be present in modern language—does not wholly contradict the otherwise obvious arbitrariness of language. The majority (p. 873) of atomic elements in natural language (words or parts of words that cannot be further divided) remain arbitrary. However, there is also a subset of language which is not arbitrary, but sound symbolic, wherein sound actually encodes meaning in some way. Though sound symbolic aspects of language have been historically acknowledged, they have often been considered a peripheral phenomenon too rare to be functional (Hinton, Nichols, and Ohala 1994; Nuckolls 1999). Recently, however, as language typology expands beyond a more traditional focus on Indo-European languages to Austronesian and African language families as well as sign languages, non-arbitrariness has emerged as a universal in its own right (Perniss, Thompson, and Vigliocco 2010). Next, we will define the exact nature of the non-arbitrariness in question, and provide evidence of its pervasive presence in natural language, particularly in the form of sound symbolism. (p. 874)

Levels of non-arbitrariness

Arbitrariness is simply defined as a random relationship between the sound form of a word and its meaning, these holding only through convention among speakers. Thus, non-arbitrariness occurs where the sound forms of words do have an observable and regular relationship to their meanings. To understand where we can expect to find arbitrariness and non-arbitrariness, we have to understand how language systems are organized. Language operates on several distinct levels, which also form interacting areas of linguistic study (and these levels are shown in Figure 43.3). The phonological level involves the sounds relevant to a particular language: phonemes. Each language contains a finite number of phonemes, or sounds which contrast resulting in changes in meaning. For example, the sounds /p/ and /b/ are considered distinct phonemes in English, due to the different meanings for words such as *pat* and *bat* (/pæt/ and /bæt/). *Pat* and *bat* are known as a minimal pair, where a slight phonetic change (in this case, voicing—the vibration of the vocal cords is markedly delayed in the voiceless /p/, and more immediately present in the voiced /b/) results in an entirely different meaning.

Click to view larger

Figure 43.3 Figure shows the levels of language, forming interacting areas of linguistic study. Areas become more broad from left to right; with phonetics involving acoustic description well below the word level and pragmatics encompassing entire interactions. Meaning in a language traditionally enters at the morphological level. Definitions adapted from William O'Grady, Michael Dobrovolsky, and Francis Katamba, *Contemporary Linguistics: An Introduction* © 1996, Pearson Education Limited.

Yet, /p/ and /b/ do not actually encode meaning; there is nothing about /b/ that denotes a flying rodent, and nothing about /p/ that evokes the action of patting. Meaning truly enters the picture at the morphological level. Morphology is concerned with the smallest units of meaning in a language; while phonemes can provide a *contrast* in meaning, morphemes encode meaning itself. Not all morphemes are words, and many words contain multiple morphemes. For example, the word *bat* is a single morpheme that is also a word. The word *re-establishment*, however, contains three distinct morphemes: *re-* (to do again), *establish* (to erect or begin), and *-ment* (which carries information about the word's category: noun). The morphemes *re-* and *-ment* are (p. 875) not considered words, as they cannot stand alone like *establish* or *bat*, but can be re-used by attaching to other words.

Non-arbitrariness above the morpheme level is uncontroversial. The meaning of the word *re-establishment* is not arbitrary with respect to its form; it is composed of the meaning of its morphemic subparts. Wherever we find the morpheme *re-*, we find the meaning "to do again." While words are composed of the meanings of their morphemic subparts, morphemes are not composed of the meanings of their phonemic subparts; phonemes do not encode meaning. A finite set of phonemes carrying no meaning can be recombined in a variety of ways to create meaningful morphemes. This feature of language is part of what provides its capacity to express a virtually unlimited array of concepts.

One of the major questions in language evolution is how we came to combine meaningful subparts (such as morphemes) to create larger meaningful words and sentences. However, underlying this question is the mystery of how meaningful subparts emerged at all; the ability to create words from meaningful morphemes requires the emergence of meaningful morphemes as a pre-requisite. A sensory theory of protolanguage seeks to answer how we came to have meaningful morphemes at all, considering the subparts of those morphemes are theoretically meaningless. Our theory suggests that at some basic psychological level, certain phonemes *can* have a natural connection to meaning. This being the case, we would expect to see sound-meaning correspondences below the morpheme level in natural languages. The next section will discuss instances of sound symbolism, where we do in fact observe sound-meaning correspondences.

Sound symbolism

Hinton, Nichols, and Ohaha (1994) identify four distinct types of sound symbolism. *Corporeal sound symbolism* includes coughs, hiccups, and other "natural" noises which give information about the state of a speaker; largely extra-linguistic, these are not relevant for the current discussion. *Conventional sound symbolism* includes cases where phonemes are associated with features of meaning non-arbitrarily (more on this later). *Imitative sound symbolism* includes conventionalized onomatopoeia (*hiss*, *buzz*) as well as more directly imitative sounds (*ssss*, *zzzz*). *Synesthetic*[3]*sound symbolism* includes sounds designed to cross-modally imitate other sensory phenomena; as the sound of the word *takete* might be taken to "imitate" an angular form. Of most interest to a sensory theory of protolanguage is imitative and synesthetic sound symbolism. We will consider these together using a new term: *sensory sound symbolism*. Imitative sound symbolic forms are certainly more straightforward than their synesthetic counterparts because they represent an association within the auditory modality alone, rather than one that crosses sensory boundaries (e.g., a visual experience expressed as a linguistic sound). (p. 876) However, the two types of sound symbolism are unified, since linguistic symbols can be considered sensorily grounded in both cases (see Figure 43.2). More specifically, sensory sound symbolism constitutes a psychologically and linguistically natural connection, wherein form is motivated by meaning. It has been demonstrated using *cross-linguistic experiments*, in particular among a class of words called *ideophones* (which we define and describe in detail later).

Finally, sensory sound symbolism can be separated from what we might call *conventional sound symbolism*. Conventional sound symbolism is found in language in two forms: as significant *statistical correspondences* between form and meaning within the lexicon, and as *phonesthemes* (both defined and described in detail later). The following section will discuss both types of conventional and sensory sound symbolism in detail.

Conventional Sound Symbolism

Statistical correspondences

Where sound is somehow motivated by meaning below the morpheme level, we should expect measures of sound in a language system to correlate with measures of meaning. Put differently, words which sound similar should also share a significant similarity in meaning. Shillcock et al. (2001) examined the most frequent monomorphemic words in the British National Corpus (a large collection of spoken and written samples of English; 2001). Each word was compared to every other word on two measures: how similar the words were in terms of their sounds (phonologically), and how similar the words were in terms of their meanings (semantically). Shillcock and colleagues found a significant correlation between phonological and semantic similarity, indicating that words which sounded similar were more likely than chance to have similar meanings. Tamariz (2005) carried out a similar study with monomorphemic words in Spanish, likewise finding that phonologically close words were significantly more likely than chance to also be semantically close.

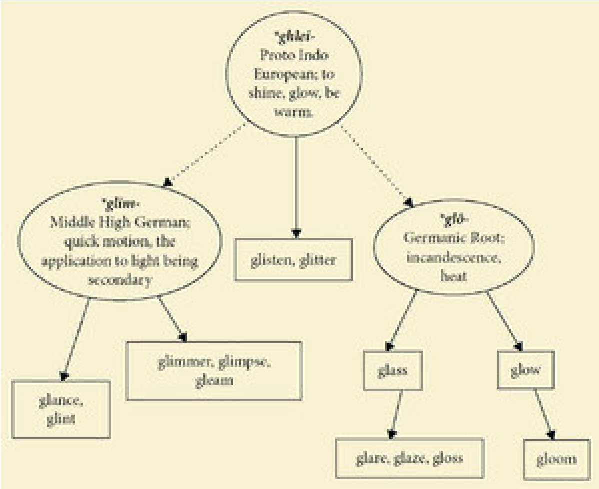
While this demonstrates that linguistic sound is somehow related to meaning below the morpheme level (since only monomorphemic words were considered), such a broad cross-lexical measure does not give any information about which particular sounds correspond to particular features of meaning, and why. These correspondences demonstrate a non-arbitrary relationship between form and meaning, but they do not reveal the basis of this relationship. They only fall out of the *system* writ large; a wide net must be cast in order to capture significant correspondences. Indeed, even though the correlation is highly significant, it accounts for a miniscule amount of the variance. This is as we would expect, given that any cross-lexical non-arbitrariness is far from obvious, and certainly not at the level of conscious awareness for most speakers of a language. Hence although sound and meaning correspondences are present and have significant strength, there is not sufficient evidence to suggest the exact nature of the correspondences, let alone (p. 877) whether they are motivated by cross-sensory associations.[4] Tamariz (2005) discusses the various pressures on a language system, suggesting that form-meaning correspondences may contribute to the overall systematicity in a language. This systematicity is useful for language learners; studies show that random systems are considerably more difficult for learners (Kirby, Cornish, and Smith 2008).

Monaghan, Christiansen, and Chater (2007) have taken a slightly different approach, and considered the matter cross-linguistically, specifically from the perspective of child language learners in English, Dutch, French, and Japanese. Rather than meaning distance per se, Monaghan and colleagues examined grammatical category information, contrasting open and closed class words[5] and nouns and verbs. In other words, they asked if open class words share more sound features with other open class words than with closed class words, performing similar analyses with nouns and verbs. Monaghan and colleagues found that words which shared class and grammatical categories shared phonological similarities in all four languages. Not all of these shared features were truly cross-linguistic; for example, while the presence of plosives (e.g., *p, d*) was a good cue for open classed words in English, the same did not necessarily hold for the other languages. However, some cues *were* shared across languages: overall length in phonemes was a good predictor of open class membership, the presence of bilabial phonemes (sounds produced using both lips; e.g., *m, b*) was a good predictor of nouns, and the presence of velar phonemes (sounds produced with the back of the tongue against the velum; e.g., *k, g*) was a good predictor of verbs (see also Farmer, Christiansen, and Monaghan 2006; Monaghan, Chater, and Christiansen 2005).

The cross-linguistic nature of these results is compelling, but the meaning metric used, though arguably more fine grained than Tamariz (2005) or Shillcock et al. (2001), is as much syntactic as it is semantic. Hence, correspondences observed between form and "meaning" appear to be driven by syntactic category rather than the individual lexical semantics themselves. Though these categories do have semantic features (nouns are more likely to be objects whereas verbs are more likely to be actions), they are still rather broad swathes demonstrating correspondences at the level of the system as a whole, rather than the lower phono-semantic level. Nonetheless, studies of statistical correspondences among monomorphemic words do demonstrate that there is non-arbitrariness in language, despite the historical view in linguistics that language is arbitrary. While these correspondences may represent sound symbolism (p. 878) that is more sensory than conventional, there is no definitive evidence to suggest this is the case.

Phonesthemes

Phonesthemes are a more specific sound-meaning correspondence directly observable in individual monomorphemic words. Phonesthemes are operationally defined as monomorphemic words sharing correspondences in sound and meaning beyond the level that would be predicted by chance alone (Bergen 2004). For example, a group of words in English beginning with *gl-* all have meanings denoting visual lightness (e.g., *glint, glimmer;* also see Figure 43.4). While the consonant cluster *gl-* is not considered a morpheme, it appears to have a non-arbitrary relationship to this visual meaning-feature. Phonesthemes have been identified in a diverse array of languages, including Austronesian languages (Blust 1988), Ojibwa (Rhodes 1981), and Swedish (Abelin 1999).



Click to view larger

Figure 43.4 Figure shows the possible historical relationship among *gl-* phonesthemes identified by Bergen (2004). Solid lines indicate an attested historical relationship documented in the *Oxford English Dictionary* (2011); dotted lines indicate a possible relationship between Proto Indo European and later Germanic forms. Although these reconstructions are not uncontroversial, we contend they are at least as likely or possible as “natural motivatedness” between the form *gl-* and the meaning “having to do with vision or light.”

Phonesthemes are compelling candidates for meaning-motivating-form at the phonemic level, but they may be problematic. First, phonesthemes are often identified (p. 879) by researchers through intuition, and are not objectively or independently verified (Drellishak 2007; Hutchins 1998). Where objective methods are used, results are mixed and confined to a small number of phonesthemes. Drellishak (2007) used three separate statistical approaches to the English lexicon, and uncovered only four phonesthemes in English: *sn-* (nose; snobbish), *st-* (firm; upright; linear), *-ing* (e.g., *bing, ting* for sounds), and *spr-* (to radiate out; elongated).[6] Bergen (2004) looked to the Brown Corpus (Kucera and Francis 1967) specifically for *gl-*, *sn-*, *sm-* and *fl-* phonesthemes. He found that all four clusters were more likely than chance to be used in the corpus with their phonesthetic meanings: for example, *gl-* was four times as likely to relate to vision as *sn-*.

In contrast, Hutchins (1998) compiled an exhaustive list of phonesthemes identified through intuition in the literature, resulting in over 145 phonestheme types. Using a subset of these, Hutchins tested whether nonsense words containing phonesthetic clusters were more likely to be associated with their phonesthetic meanings, rather than other meanings. Subjects chose the expected meaning well above chance levels, demonstrating a strong psychological basis for phonesthemes. In a study designed to examine online processing of phonesthemes, Bergen (2004) used a primed lexical decision task: subjects see a prime word followed by either a word or a non-word target, and are tasked with pressing a key indicating whether the target is a word or not. In such a task, the goal is to discern what effect the relationship between the prime and target has on decision times. In Bergen’s (2004) task, primes were either unrelated to targets (e.g., prime: *frill*, target: *barn*), related only phonologically (e.g., prime: *drip*, target: *drab*), related only semantically (e.g., prime: *cord*, target: *rope*), related pseudo-phonesthetically (sharing features of sound and meaning which do not occur significantly often in English, e.g., prime: *crony*, target: *crook*), or related fully phonesthetically (e.g., prime: *glitter*, target: *glimmer*). Phonesthetic primes resulted in the most significant facilitation to reaction times when paired with a phonesthetic target. These data demonstrate a psychological basis of phonesthemes even below the conscious level demonstrated by Hutchins (1998). Using Hutchins' exhaustive list of phonesthemes, Otis and Sagi (2008) returned to a corpus approach. Using the Project Gutenberg corpus (Lebert 2005) and latent semantic analysis (LSA; Landauer and Dumais 1997) to provide meaning distance, they found that ten word-initial phonestheme clusters shared significant semantic properties, including *gl-*, *sn-*, and *spr-* (Otis and Sagi 2008).

While phonesthemes demonstrate a strong form meaning relationship (i.e., a psychological bias as well as a statistical reality in some cases) there remains major doubt as to whether phonesthemes are naturally "motivated," in the sense of deriving from some shared cross-modal rule. Similar phonesthemes have not been demonstrated cross-linguistically, and thus could be the result of historical idiosyncrasies in a particular language. Put differently, although present in a variety of unrelated languages, phonesthemes do not necessarily manifest similarly across languages (e.g., *gl-* does not (p. 880) indicate light in most other languages with phonesthemes). In fact, many *gl-* phonesthemes identified by Bergen (2004)[7] share etymologies (e.g., both English and Swedish share this phonestheme and both are Germanic). In other words, rather than the form being cross-modally motivated by the meaning, at least in the case of *gl-*, the observed relationship may be the result of a particularly productive branch of words that goes back as far as Proto Indo-European (see Figure 43.4 for a possible depiction of the shared historic roots of *gl-*).

It is still possible that some phonesthemes are naturally motivated. However, historical relationships are at least as plausible, and must be definitively ruled out before declaring phonesthemes to be sensory sound symbolic. Even so, the problem is still not straightforward: the fact that certain sets of words deriving from the same proto-form survive in a language may be the result of a cross-modal goodness of phonological and semantic fit. Put differently, a natural goodness-of-fit between form and meaning may lead to that particular pairing of form and meaning being especially productive. Phonesthemes remain an interesting case of non-arbitrariness in language, but without further evidence, must be classed as conventional rather than sensory sound symbolism.

Sensory Sound Symbolism

Ideophones

Sensory sound symbolism is exemplified by an intriguing category of words known as *ideophones*. Also termed *expressives* (Diffloth 1994; Tufvesson, 2011) or *mimetics* (Imai et al. 2008), ideophones are sometimes grouped with phonesthemes (Nuckolls 1999), but are quite distinct in that native speakers consciously report ideophones as being linguistic forms sensorily motivated by their meaning. Dingemanse (2009) defines ideophones as "marked words which vividly depict sensory events." Onomatopoeia in English is an excellent example; *hiss* actually directly depicts the sound made by a snake. Though ideophones encompass onomatopoeia, they also go beyond it, depicting not only auditory events, but visual, emotional, and tactile events, among others. For example, in the Togo region of Ghana, *giligili* evokes circular shape, and *wúrúfúú* a fluffy texture for speakers of Siwu (Dingemanse 2011). Since the very form of ideophones evokes sensory experience for language users, they appear to be an instance of cross-modal goodness-of-fit underlying modern natural language. Ideophones have fascinated many linguists, with Frankis (1991) going so far as to refer to them as "the lunatic fringe of language" (17).

In some instances, ideophones occur in a strange kind of minimal pair. These minimal pairs are different from the usual *pat/bat* type discussed earlier; here, a change in sound is accompanied by a *specific* change in meaning; phonemes are directly encoding (p. 881) meaning, such that changes in phonemes evoke a systematic meaning change. For example, in Japanese, *kirakira* means "to glitter or sparkle," where *giragira* means "to glare or dazzle." The voicing distinction, in this case between /k/ and /g/, is changing the meaning in terms of magnitude; a small to an overwhelming amount of reflected light. The voicing distinctions present in Japanese ideophones are also trends in Siwu (Dingemanse 2011), demonstrating an interesting cross-linguistic effect between two genetically distant languages. In other ideophonic languages, vowel quality distinctions have a similar effect, as in Bahnar, where a change in vowel height (e.g., /i/ versus /a/) indicates a difference in size (Diffloth 1994). These cross-linguistic trends make for testable predictions. Common ideophonic features occurring between genetically distinct and geographically disparate languages would point to an overall cognitive bias, eliminating the historical explanations possible for phonesthemes. If ideophones truly depict sensory events, we should expect people who do not speak Japanese, for example, to have some insight into the meanings of Japanese ideophones. Recent experiments show that ideophones do provide clues to meaning even for naïve listeners.

Experiments with natural language

Iwasaki, Vinson, and Vigliocco (2007a) compared Japanese and English speakers' intuitions regarding Japanese ideophones for motion and laughing (e.g., *bura-bura*: strolling, *kusu-kusu*: giggle). Despite no knowledge of Japanese, the English speakers rated the meanings of Japanese ideophones similarly to the Japanese speakers. For the motion ideophones, voicing was identified as indicative of large size in both groups (e.g., *b* is larger than *p*). Similar judgments between Japanese and English speakers were also found for Japanese pain ideophones (Iwasaki, Vinson, and Vigliocco 2007b). Another approach to the cognitive bias underlying ideophones examines learnability. If ideophones are forms motivated by their meanings, the natural connection between form and meaning should make them more easy to learn than forms with no natural relationship to their meaning. Imai et al. (2008) found that 3-year-old children were able to more effectively learn the meaning of novel verbs when they were ideophonic. Nygaard, Cook, and Namy (2009) found similar results with English speaking adults; subjects learned the definitions of Japanese ideophones significantly quicker when they were paired with their actual meaning (*hayai*: fast), as opposed to the opposite meaning (*hayai*: slow) or a completely arbitrary meaning (*hayai*: blunt)

These studies show that people with no knowledge of a language genetically distant from their own are able to access the meanings of its words based solely on their form. Surprisingly, this is also true for non-ideophones. Tsuru and Fries (1933; as cited in Brown, Black, and Horowitz 1955) first investigated this with English speakers and Japanese words. Though Japanese has an extensive system of ideophones, Tsuru and Fries intentionally used description words that were not ideophonic in character. Their prediction was that although ideophones appear to be a special case of particularly evocative words, that there is information in all word forms—even those in (p. 882) non-ideophonic languages—about meaning. In Tsuru and Fries' (1933) initial study, subjects were provided with pairs of English words and their Japanese equivalents, for example *bird* and *worm* with *tori* and *mushi*, or *white* and *black* with *shire* and *kuro* (the pairs were presented in random order). With no prior knowledge of Japanese, subjects chose the correct Japanese equivalent of the English words significantly above chance levels (*tori* means BIRD and *shire* means WHITE). Similar results have been found for English speakers with a diverse array of languages, including Hungarian (Klank, Huang, and Johnson 1971), Polish, Chinese, and Czech (Brown, Black, and Horowitz 1955), Croatian (Maltzman, Morisett, and Brooks 1956), Hebrew (Brackbill and Little 1957), Hindi (Brown and Nuttall 1959), and Thai, Karnese, and Yoruba (Slobin 1968). While some of the languages (e.g., Japanese and Yoruba) have ideophone systems, many lack ideophones (e.g., Czech, Croatian, Polish). More recently, Berlin (1994) has run a similar experiment with English speakers and Huambisa, a Jivorian language spoken in Peru. Using bird and fish names, Berlin had subjects classify which member of a pair was a bird and which a fish. Over a list of 50 animal names, subjects have 58% accuracy at guessing the correct category, a statistically significant finding. Moreover, Berlin found that accuracy with some pairs was extremely high, up to 98%.

Tasks of this type can only be possible if English speakers are detecting features in the forms of the foreign words which encode meaning, even outside ideophonic systems. Quite independently, Berlin (2006) and Westermann (1927) have made suggestions regarding the nature of these specific sound-meaning connections, based on corpus research (see Table 43.1which shows their collection of sound symbolic items proposed in the literature). In the case of Huambisa, Berlin (1994) suggests that specific vowel and consonant contrasts play an important role, with high vowels (like /i/) being associated with the quick movement of birds, while lower vowels like /a/ are associated with the slower flowing movement of fish.[8] As a follow-up, Berlin (1994) examined a comprehensive corpus of bird and fish names in Huambisa, finding a preference for high vowels in bird names and low vowels in fish names.

Based on Jesperson's (1933) informal examination of several languages, finding that high vowels are often associated with not only with speed but small size (as in English *teeny*, *little*; Nuckolls, 1999; see also Ohala, 1984[9]), Berlin (2006) also examined several other South American languages for patterns of size sound symbolism. This time using only bird names, species were divided into those over 10 inches and those under. As expected, the names of smaller birds were significantly more likely to contain higher vowels. Berlin performed a similar analysis on words for *squirrel* and *tapir* in 25 languages, finding /i/ significantly more frequently in words for *squirrel* and /a/ more frequently in words for *tapir*. Berlin (2006) suggests that ethnozoological nomenclature is (p. 883) of special interest, as it often involves conscious creation of labels, and thus is a subset of language where sensory sound symbolism is likely particularly strong.

Table 43.1 Features of sound and their associated meanings in the sound symbolism literature. Adapted from Westerman (1927) as reported in Fox (1935) and Berlin (2006)

Sound	**Associated Meanings**
Front vowels, high tone, short vowels, voiceless consonants	Bright, quick, sharp, pleasant taste, pleasant smell, intensive color, energetic, fresh, exact, rapid, abrupt, short, agile, long, straight, hagged, skinny, sharp, thin, slender, angular, small
Back vowels, low tone, long vowels, voiced consonants	Large, voluminous, heavy, soft, dull, slow, tasteless, unpleasant smell, heavy, blunt, chaotic, slow, sluggish, flowing, awkward, smooth, spherical, short, stocky, fat, hefty, squat, rotund

Reproduced from D. Westerman. Laut, Ton und Sinn in westafrinaische Sudansprachen. In *Festschrift Meinhof*, 315–328 © 1927, Kommissionsverlag von L. Friedgerichsen & Co.

In another corpus approach, Ultan (1978) examined 136 languages, searching specifically for phonemes which indicated size contrasts. While only a about a third of the languages surveyed contained such relevant contrasts, of those that did over 85% exhibited the pattern predicted by Jesperson and confirmed by Berlin, with the high vowel denoting smaller size.[10] Woodworth (1991) extended a similar method to words denoting physical distance, examining deictics (e.g., *this*, *that*) and finding that higher vowels were present for proximal pronouns for 70% of the 26 languages surveyed. Traunmuller (1996) confirmed these results even more specifically, connecting the observed trends directly with the frequency of the second formant (an acoustic correlate of vowel quality).

Even in the absence of a formal ideophonic system, many languages contain clues to meaning at the phonemic level. Trends among specific vocabulary elements, be they animal names or deictic pronouns, demonstrate this well at a linguistic level: a given language system takes advantage of sound-meaning patterns below the morphological level. Experimentally, even naive subjects seem to be able to pick up on these correspondences in completely foreign and unrelated languages, suggesting not only a psychological reality, but a universality of certain sound symbolic patterns.

Linguistic Cross-Modality

Sensory sound symbolic patterns found in natural language can provide the basis for a different approach to finding evidence for natural goodness-of-fit between linguistic (p. 884) sound and meaning. Sound symbolic patterns found in natural language can be used to create novel linguistic stimuli testable against controlled perceptual stimuli, joining approaches from cross-modality more generally with the interest in linguistic sound symbolism. Various experiments have shown that people make common cross-modal associations between linguistic sounds and properties such as size (Newman 1933; Sapir 1929; Thompson and Estes, 2011), angularity (Maurer, Pathman, and Mondloch 2006; Ramachandran and Hubbard, 2001), and taste (Simner, Cuskley, and Kirby 2010). The following section reviews such studies.

Magnitude

Size vowel symbolism, the notion that some vowels are “smaller” than others, is of particular interest in studies of linguistic cross-modality. In fact, Jesperson’s (1933) informal investigation into sound symbolism in natural language was inspired by experimental work done by Sapir (1929). Sapir used sets of simple non-words, such as *mil* and *mal*, which contrasted only vowel quality. Subjects were asked to match these words with a large or small object. Sapir found that 80% of over 400 subjects matched /a/ with large objects and /i/ with small objects. Newman (1933) extended Sapir’s findings, showing that not only the absolute height of the vowels /i/ and /a/ affected judgments of size, but relative vowel height had a similar effect (see also Johnson 1967; Huang, Pratoomraj, and Johnson 1969).

More recently, Thompson and Estes (2011) used a different method incorporating consonants and improving Sapir’s traditional forced choice paradigm. Instead, a single abstract object was presented in the context of a known object, such as a cow, to indicate size, alongside a list of five nonsense words to choose from. The nonsense words were designed using the vowel qualities identified by Sapir (1929) combined with voiced and voiceless consonants, the latter (voiceless consonants) predicted to correlate to smaller size. Thompson and Estes (2011) found that subjects were significantly more likely to choose words containing “large” phonemes (e.g., /a/, /m/, /w/) for the large abstract objects, and conversely, more likely to pair “small” phonemes (/i/, /k/, /t/) with small objects.

Ohala (1994) has suggested a plausible mediation for magnitude sound symbolism called “the frequency code hypothesis.” Initially applying this to overall voice pitch (i.e., the *fundamental frequency* of the speech signal, known as F0), Ohala points out that in the natural world, smaller animals produce high pitched sounds while larger animals emit low pitched sounds, thus, we could conclude that the pitch of the speech signal underlies size sound symbolism. However, incidental variations in pitch in human speech are not a reliable indicator of speaker size (Fitch 2000). Intentional variations in pitch play a phonemic role in some languages in the form of tone (e.g., in Mandarin), but the use of tone is not pervasive, making the explanatory power of fundamental frequency as the underlying mechanism for size sound symbolism limited.

(p. 885) However, all languages do make use of vowel quality: the difference between vowels such as /i/, /a/ and /o/.[11] Vowel quality is acoustically determined by the relationship between resonant frequencies of F0 known as the first, second, and third *formants*. Not only does the relationship between formants underlie vowel quality, but the overall value of the formants (as long as their relative distance is maintained) is also an honest signal of speaker size in humans (Fitch 2000). Fisher-Jorgensen (1978) found that this relationship could be captured by subtracting the value of the first formant from the second. This gives a frequency representing the vowel contrasts relevant in magnitude sound symbolism: the larger the difference between second and first formant, the smaller the size (Ohala 1994). For example, there is a large difference between the second and first formant in the vowel /i/, which is more likely to be associated with small objects (as in /mil/ vs. /mal/). The frequency code can also be extended to consonants: voiceless consonants have higher frequencies than voiced consonants, and dental, alveolar, palatal and front velars have higher frequency than labials and back velars (Ohala 1994). Physical size can also be extended to other magnitude dimensions, including light and dark (Newman 1933), as well as the many other sensory dimensions (detailed in Table 43.1). In summary, the frequency code gives a mechanism for explaining a multitude of common associations made between magnitude and vowel quality.

Visual angularity

Parallel to the work on magnitude sound symbolism, there has also been a large body of work examining visual angularity and linguistic sound, after Köhler’s (1929, 1947) *takete/maluma* experiment (in which most people pair the word *takete* with an angular shape, and *maluma* with a rounded shape). The paradigm has been extended to include a variety of non-words, including *takete/uloomoo* (Davis 1961), *kiki/bouba* (Ramachandran and

Hubbard 2001), and *teetay*/*goga* (Maurer, Pathman, and Mondloch 2006), as well as varying visual stimuli.

Köhler's (1947) report of the phenomenon was directly inspired by Usnadze (1924) and Fischer's (1922) accounts of similar experiments. Rather than using a forced choice paradigm, Fischer had subjects rate the goodness-of-fit of non-words to shapes, finding that certain names were rated as fitting certain shapes exceptionally well across subjects. Usnadze (1924) used a more qualitative method, focusing on what caused reported goodness-of-fit between a name and a figure. Subjects either matched a name with a figure using an associative strategy, based on existing lexical knowledge (e.g., choosing the word *jage* for the jagged shape), or an intrinsic strategy, matching specific qualities of sound (e.g., vowel) to visual qualities of the shape (e.g., size). Interestingly, Usnadze reported that after a short time interval, subjects were more (p. 886) likely to make the same or similar sound-shape associations as their own previous trials if they had used an intrinsic rather than associative strategy on the initial trial. In a more extensive follow-up, however, Fox (1935) found that associative choices resulted in better recall.

Given the limited number of subjects in early studies, subsequent attempts focused on extending the developmental and cultural spread to examine both the innateness of naming biases as well as their cultural universality. Irwin and Newland (1940) were the first to examine shape-name biases with children and adults. Using a paradigm modeled after Köhler's, subjects from age 4 years to adults were asked to label shapes with aurally presented non-words designed by the experimenters to "fit" the test shapes. Irwin and Newland found that adults performed as expected, but this weakened as the age of subjects decreased, and responses did not differ from chance for 4- to 8-year-olds. Davis (1961) undertook the first cross-linguistic examination of the effect, with the goal of discerning whether the effect may be confined to speakers of Germanic languages. Davis tested both English schoolchildren as well as children in Tanganyika (found in modern Tanzania) who spoke Swahili (a Niger-Congo language) and Kitongwe (a Bantu language). Using Köhler's original visual stimuli and replacing the word *maluma* with *uloomu* (due to an apparent lexical confound in Kitongwe), Davis found the effect was robust for both the English and Tanganyika populations. However, Rogers and Ross (1975) subsequently reported random responses when testing the effect with the Songe in Papua New Guinea.

More recently, Ramachandran and Hubbard (2001, 2005) have anecdotally reported the experiment using the words *bouba* and *kiki*, reviving interest in the paradigm, especially with respect to language evolution. Maurer and colleagues examined the effect among 3-year-old children and adults, using a set of four pairs of shapes (with an angular and rounded shape within each pair) and four pairs of carefully designed non-words. Unlike previous manipulations, the non-words were purposefully designed rather than intuitively chosen to differ only in terms of vowel roundedness (e.g., the vowel in *bow* is rounded, and the vowel in *bay* is unrounded), with the prediction that rounded vowels would match with rounded shapes. Maurer, Pathman, and Mondloch (2006) found the effect among both children and adults, contrary to Irwin and Newland's (1940) earlier findings.

Connecting their results directly to Ramachandran and Hubbard's (2001) synesthetic bootstrapping theory, Maurer and colleagues suggest an articulatory mediation for their findings, wherein the rounding of the lips imitates the rounding of the shape.

Recent studies have increasingly moved towards more controlled stimuli from both linguistic and visual perspectives. Ahlner and Zlatev (2010) set out to use a larger set of non-words, using systematic combinations of front (e.g., /i/) and back (e.g., /u/) vowels with voiced sonorants (e.g., m, l) and voiceless obstruents (e.g., p, t). They found that adult Swedish speakers paired non-words with front vowels or voiceless obstruents with an angular shape, and this effect was particularly strong in non-words containing front vowels *and* voiceless obstruents together. Neilsen and Rendall (2011) also approached possible sound contrasts in terms of both vowel and consonant (p. 887) quality, as well as creating systematically varied round and angular shapes, finding that the quality of consonants was primarily responsible for non-word/angularity cross-modal associations.

Other recent approaches have deviated slightly from Köhler's original paradigm, choosing more implicit measures of associations. Westbury (2005) hypothesized that consonant articulation, in the form of stops (e.g., p, d, k) versus continuants (e.g., m, l, z), drives linguistic cross-modal associations involving visual angularity. Using a lexical decision task, both words and non-words were presented inside curved or angular frames. Westbury found that correct classification of non-words containing continuants was facilitated by curved frames, and correct classification of non-words containing stops was facilitated by angular frames. This was the first evidence of implicit sound symbolism for angularity, and the first study to show that a specific sound property (the continuant property of consonants) was related to angularity.

Another approach to angularity and linguistic sound looks at the learnability of sound-meaning mappings. Kovic and colleagues trained subjects on non-word-to-shape category associations, and then tested them using a timed categorization task (Kovic, Plunkett, and Westermann 2010). Subjects were trained to categorize figures with non-word labels that were either "sharp sounding" (e.g., /rɪf/) or "round sounding" (e.g., /mɔt/), in either a congruent (angular figures were *rifs*) or incongruent (angular figures were *mots*) condition. In the test phase, subjects heard the label followed by the presentation of an angular or curved figure and responded by pressing a "match" or "mismatch" button based on their training trials. Subjects who were trained in the congruent condition had significantly faster reaction times for both match and mismatch items.

Nielsen and Rendall (2012) used a similar paradigm, however, they approached the design of the non-words with the stop/continuant distinction established by Westbury (2005) in mind. Subjects were trained on form-meaning pairs which were either congruent (stop non-words were paired with angular shapes) or incongruent (stop non-words were paired with rounded shapes), and were then tested on these learned associations. Those who learned the incongruent vocabulary performed no different from chance, but subjects trained on the congruent condition performed significantly better than chance and significantly better than the incongruent group. This examination of learnability not only provides an implicit demonstration of cross-modal associations between linguistic form and angularity, but demonstrates that these biases might function to increase the learnability of a vocabulary system for language users.

Monaghan, Mattock, and Walker (2012) approached the learnability from a different angle, using the *cross-situational learning paradigm* (Monaghan and Mattock 2009; Yu and Smith 2007). In cross-situational learning, rather than training subjects on form-meaning pairs explicitly, a single word appears with multiple meanings over many training presentations—in this case, a single non-word appeared with multiple shapes. Subjects learned form-meaning pairs over successive trials by inference: for example, if the word *takete* appears with an angular shape over and over (even though another shape is always co-present), the form-meaning pair of *takete* + angular shape will be (p. 888) internalized. In other words, the non-word *takete* does not appear with the angular shape in isolation, but the angular shape reliably co-ocurrs with the word *takete* nonetheless. Monaghan and colleagues designed their non-word items carefully, utilizing the stop/continuant distinction as well as contrasting vowel quality (e.g., front /i/ or back /u/) in a separate condition. They found that form-meaning pairs were easier to learn where stops or front vowels were paired with angular items, and continuants or back vowels were paired with rounded items. These findings echo those of Neilsen and Rendall (2012), finding that sound symbolism facilitates learning even when more implicit strategies are necessary, as in cross-situational learning (see also Monaghan, Christiansen, and Fitneva 2011; Parault and Schwanenflugel 2006; Parault and Parkinson 2008).

Although the psychological evidence for sound symbolism regarding visual angularity is compelling, the mechanisms underlying the biases are unclear. The angularity of figures in all experiments is well controlled, but the linguistic stimuli vary, making it unclear what may be driving the effect from a linguistic perspective. Many investigations designed linguistic stimuli almost entirely based on intuition regarding goodness-of-fit with visual stimuli, making the effect rather circular (e.g., Davis 1961; Fox 1935; Irwin and Newland 1940; Kovic, Plunkett, and Westermann 2010; for further discussion see Westbury 2005). Some more recent investigations have chosen specific linguistic motivations for non-word stimuli, including vowel rounding (Maurer, Pathman, and Mondloch 2006) and manner of consonant articulation (Westbury 2005). However, neither of these studies held other phonetic factors in their linguistic stimuli constant, unlike Sapir's (1929) classic investigation, wherein *only* the vowel was changed to form an artificial minimal pair (e.g., *mil* versus *mal*). For example, in addition to manipulating vowel rounding, all of Maurer and colleagues' word pairs also alternate consonant voicing (e.g., *goga* contains voiced consonants and *teetay* voiceless). Though Westbury's (2005) consonants were for the most part carefully controlled, voicing varied systematically along with manner of consonant articulation, and vowel quality was not considered at all.

Linguistic cross-modality and taste

Most studies in linguistic cross-modality, even those directly examining natural language, have focused on the visual modality (e.g., size, angularity, lightness). Given the dominance of our visual system, and the excellent vision which defines primates (Jacobs 2009), the visual modality is a natural starting point. However, language encodes much more than the visual; if cross-modality were important in the evolution of language, we should expect it to connect linguistic sound not only to vision, but to all our sensory systems (as shown in Figure 43.2). Another reason to look beyond visual linguistic cross-modality is that the neurological condition of synesthesia is extremely varied, extending to taste in many cases (Day 2005). This is particularly (p. 889) relevant since cross-modal associations in synesthetes tend to mirror those in the general population.

Cross-modality and taste has been extensively studied, as the experience of eating is considered to be truly multi-modal, involving integration of smell, vision and sound (see Auvray and Spence 2008 for a review). A strawberry odor can amplify sweetness (Stevenson, Prescott, and Boakes 1999), and conversely, a sweeter solution will be rated as smelling fruitier (Verhagen and Engelen 2006). Discrimination between flavors is negatively affected when the color of a liquid does not match its taste (e.g., identifying a lime flavor in a red solution; Zampini et al. 2007), manipulating the sound produced when eating crisps can enhance perceived crunchiness (Zampini and Spence 2004), and color and label information can affect flavor perception (Shankar et al. 2009). However, the connection between taste and more abstract sensory experiences is less well studied.

Crisinel and Spence (2009) made the first empirical examination of taste and sound, asking subjects to categorize different bitter or sour tastes according to high or low pitch. Tastes were represented by names of foods associated with a particular taste quality, for example, coffee or tonic water for bitter and lime or vinegar for sour. Crisinel and Spence found that categorization accuracy was higher and response latency lower when high pitch was paired with sour items and low pitch with bitter items, demonstrating an apparent association between pitch and taste. A follow-up study (Crisinel and Spence 2010) confirmed these findings using additional implicit methods. However, these results may have been unduly influenced by the use of linguistic terms to mediate the experience of taste (i.e., cross-sensory mappings may have been based in part on the visual form of the words; see Simner, Cuskley, and Kirby 2010 for further detail). More recently, Gallace, Bochin, and Spence (2011) examined associations between linguistic sound and taste in greater depth. Rather than examining pitch, non-words derived from Kohler’s classic *takete*/*maluma* task were rated for goodness-of-fit with common food items. Each non-word pair (*takete*/*maluma*, *kiki*/*bouba*, and *ruki*/*lula*) provided the anchors for two ends of a scale, and food items were rated along each scale between the words. Brie, chocolate mousse, and blueberry jam were rated positively with the words *bouba* and *lula*, while cheddar cheese, mint chocolate, and crisps were rated positively with the words *takete* and *ruki* (see also Spence and Gallace 2011).

Simner, Cuskley, and Kirby (2010) aimed to examine how phonological features, rather than pitch or particular non-words, were associated with pure tastants. Pure tastants were chosen instead of linguistic taste terms such as *sweet*, *sour*, and, *bitter*, since the graphemic or phonological properties of lexical items may influence responses (see earlier). The use of specific linguistic sounds makes this study highly relevant for a sensory theory of protolanguage. Rather than examining implicit associations using a timed categorization task, or having subjects rate goodness-of-fit using an external scale, Simner and colleagues measured associations by having subjects respond to tastes (sweet, salty, bitter and sour, each at low and high concentrations) by choosing a sound (p. 890) directly. The sounds were presented in the form of four continua reflecting various qualities of speech salient in the sound symbolism literature: F1 vowel quality (roughly corresponding to vowel height), F2 vowel quality (roughly corresponding to vowel rounding), voicing continuity,[12] and overall spectral frequency.[13]

With this method, Simner, Cuskley, and Kirby found that subjects rated higher concentrations as being significantly lower vowels (e.g., /a/), more fronted vowels, and with higher spectral balance. These findings agree in part with previous notions regarding magnitude sound symbolism in particular (e.g., the frequency code hypothesis, Table 43.1): for example, low vowels indicate larger magnitude and also mapped to higher taste concentrations. Front vowels, however, are usually associated with small size, but were found in Simner and colleagues' study to map to high concentrations. This simply underscores the fact that there may not be a single rule governing linguistic cross-modal associations across all sensory systems. While the frequency code hypothesis might explain many of the magnitude associations underlying visual linguistic cross-modality, such an ethological explanation may not apply to taste.

Systematic associations were found not only for concentration, but also for taste quality. Sweet was a lower, more continuous vowel than salt, bitter and sour, and was significantly lower than sour in particular on the spectral balance scale. Interestingly, the sweet taste was involved in all of the significant differences regarding taste quality. From an evolutionary perspective, sweet was the only taste examined which always indicates a viable or attractive food source (along with umami; see Zhao et al. 2003) and would therefore greatly benefit from a naming convention that systematically derived from some shared cross-modal basis. Taken together, these results provide compelling evidence that linguistic cross-modality goes beyond the visual domain, increasing its explanatory power in terms of the evolution of language.

Conclusions: A Sensory Theory of Protolanguage Emergence

Many authors have debated the nature of protolanguage as being primarily gestural or musical (Arbib 2005; Mithen 2005); holistic or synthetic (e.g., Tallerman 2007; Wray 1998). Regardless of its nature, a detailed mechanism for the *emergence* of protolanguage remains a mystery, although many authors have argued that we should begin by looking to animal communication systems (e.g., see Fitch 2010). Few authors have proposed (p. 891) solutions which offer a way to ground protolanguage, and thus language itself, in systems predating the emergence of language. Our account of protolanguage places our sensory system firmly at the root of our linguistic system, providing embodiment for an otherwise disassociated network of symbols. The use of sensory sound symbolism would have allowed us to express and understand a variety of elementary concepts, from sharing the visual details of our surroundings to valuable information about food sources. While a sensory protolanguage may solve the issue of grounding our symbol system, it leaves many questions in language evolution unanswered. What gave us the impetus to share information at all? How did we get from a small iconic system to a virtually infinite arbitrary one, and why?

The first question remains one of the major mysteries not only of language evolution, but of human culture and cognition more generally. Many authors have considered in detail how and why we may have acquired such an impetus to share (see Dunbar 2003; Tomasello et al. 2005). Tomasello et al. (2005) point out that it would have required some degree of a theory of mind: the ability to represent others' states of belief and desires. Interestingly, a lack of theory of mind appears to correlate with an inability to perform common cross-modal associations (Ramachandran and Oberman 2006), perhaps indicating that the perspective-taking involved in theory of mind is also crucial for sharing cross-modal associations. Whatever the origins of this drive to share or its reasons for persisting, we would assume it predates the emergence of protolanguage. Given that the sort of information relevant for sharing would have to be mediated by our perceptual system, grounding our hopeful utterances in shared sensory experience would have been a reasonable starting point.

How and why might an iconic protolanguage have become arbitrary? Pressures on an expanding language system—a small protolanguage shifting to a full-fledged lexicon—would have resulted in the leap from iconic to arbitrary. There are two related advantages to an iconic system: increased learnability and decreased processing demands. The first, increased learnability, has been demonstrated most powerfully with ideophones (e.g., Imai et al. 2008; Nygaard, Cook, and Namy 2009) for both Japanese and English speakers; Yoshida and Smith (2006) have also shown that ideophonic words are used more frequently with children, when word learning is at its peak. Learnability is not trivial; in fact, it is considered a core property of human language (Hockett 1960)—a language must be learnable by its users in order to be useful. Iconic forms are easier to learn precisely because they are grounded in existing perceptual and cognitive systems, rather than symbols defined only by convention and essentially requiring rote memorization. This leads onto the second advantage: an iconic system requires less processing power. Having been more easily stored, the retrieval of an iconic form is more straightforward than retrieval of an arbitrary form. Iconicity adds a natural strength in the bond between form and meaning, making the retrieval of a meaning with only the form, or vice versa, more automatic. As a simplified example, if all forms denoting small objects contained an /i/, the task of trying to retrieve the meaning of a form such as *mil* is constrained to meanings involving small objects. In an arbitrary system, however—where *mil* might denote any type of object—the task of retrieving its (p. 892) meaning remains largely unrestrained, and the search for the meaning of a given form must be exhaustive.

Given these advantages, why would language ever become an arbitrary system? Gasser (2004) has examined the advantages of iconicity in a set of simulations using a simple feedforward network. The network learns iconic languages with much lower error; however, this advantage only persists while the iconic language is small. Arbitrariness is a response to pressure for a language to convey more meanings. Monaghan, Christiansen, and Fitneva (2010) have suggested this may be why some non-arbitrary form-meaning relationships hold between the relatively small number of categories in a system (e.g., nouns and verbs as per Monaghan, Christiansen, and Chater’s 2007 finding), but are not as apparent at finer levels of the lexicon (see also Tamariz 2005).

Finally, in order for a complete account of protolanguage emergence, there must be some attempt to demonstrate relevant continuity with our evolutionary relatives: other primates. Although complex tasks involving cross-modal association such as the *takete/maluma* task have not been tested with other primates, there is recent evidence that at least Chimpanzees engage in some cross-modal associations. Ludwig, Adachi, and Matsuzawa (2011) have shown that chimpanzees map high luminance to high pitch, a mapping well-documented in humans (Marks 1974). Chimpanzees have also shown impressive memory of numerical symbols in particular (Matsuzawa 2009), and Humphrey (2012) has suggested their success in symbolic memory tasks may be due to grapheme-color synesthesia, well-known to enhance the memory of human synesthetes (e.g., Yaro and Ward 2007). This would mean the cross-modal continuum mentioned earlier was likely present in our last common ancestor with chimpanzees, allowing it to be leveraged for language evolution in hominids.

Another, perhaps less abstract, cross-modal ability known as *cross-modal transfer* is found in both other primates and very young children. Cross-modal transfer provides an accurate expectation of an object’s properties in one modality having only experienced it in another. For example, we have expectations about what an object may feel like (e.g., sharp) based on how it looks (e.g., angular). This has obvious survival value; one can make effective decisions regarding multisensory interaction with novel objects via input from a single modality. Cross-modal transfer has been found in infants as young as 6 months old (Rose, Gottfried, and Bridger 1981), as well as in chimpanzees (Davenport, Rogers, and Russell 1973) rhesus monkeys (Cowey and Weiskrantz 1975), and even bushbabies (Ward, Yehle, and Doerflein 1970). Savage-Rumbaugh and colleagues have found more robust cross-modal transfer in language trained chimpanzees, including tasks that involve “not only cross-modal associations, but also the transformation of information from symbolic to representational modes” (Savage-Rumbaugh, Sevcik, and Hopkins 1988, 617). This finding would suggest not only a continuity of cross-modal abilities between humans and chimpanzees, but also that the availability of a small lexicon may enhance abilities in cross-modal association. In the course of protolanguage emergence, cross-modal abilities such as cross-modal transfer may have been the genesis of an iconic system. Learning and use of such a system may have expanded cross-modal transfer to cross-modal (p. 893) association, allowing for the more abstract associations commonly found among humans.

Of course, this theory leaves many questions regarding language evolution unanswered, and there are still gaps even regarding protolanguage emergence. The emergence of theory of mind is an evolutionary problem requiring continued interdisciplinary study. Although we can outline clear stages in the emergence of protolanguage and the move to a larger lexicon, an exact timeline remains difficult to pin down. Written records of modern language go back at least 6,000 years, and estimates of the emergence of language range from tens of thousands to hundreds of thousands of years ago. Lastly, while a theory based on sensory associations offers a comprehensive answer to the symbol grounding problem, there are many important aspects of language evolution it cannot hope to explain, such as the emergence of syntax. However, we hope to have shown that a consideration of cross-modal associations evident in both synesthesia and the general population more widely can shed light on how our species moved from a state with no symbolic communication to one which laid the foundations of language as we know it today.

References

Abelin, Åsa. 1999. Studies in sound symbolism. Doctoral dissertation, Göteborg University.

Ahlner, Felix, and Jordan Zlatev. 2010. Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies* 38:298–348.

Arbib, Michael A. 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28:105–167.

Atchison, Jean. 2000. *The Seeds of Speech: Language Origin and Evolution*. Cambridge: Cambridge University Press.

Auvray, Malika, and Charles Spence. 2008. The multisensory perception of flavor. *Consciousness and Cognition* 17:1016–1031.

Bauer, L. 1996. No phonetic inventory in evaluative morphology. *Studia Linguistica* 50 (2): 189–206.

Bergen, Benjamin K. 2004. The psychological reality of phonaesthemes. *Language* 80:290–311.

Berlin, Brent. 1994. Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In *Sound Symbolism*, ed. Leanne Hinton, Johanna Nichols, and John J. Ohala, 76–93. Cambridge: Cambridge University Press.

Berlin, Brent. 2006. The first congress of ethnozoological nomenclature. *Journal of the Royal Anthropological Institute* 12 : S23–S44.

Blust, Robert. 1988. *Austronesian Root Theory: An essay on the Limits of Morphology*. Amsterdam: John Benjamins.

Brackbill, Yvonne, and Kenneth B. Little. 1957. Factors determining the guessing of meanings of foreign words. *Journal of Abnormal and Social Psychology* 54:312–318.

Brang, David, Romke Rouw, Vilyanur S. Ramachandran, and Seana Coulson. 2011. Similarly shaped letters evoke similar colors in grapheme-color synaesthesia. *Neuropsychologia* 49:1355–1358.

(p. 894) British National Corpus, version 2 (BNC World). 2001. Distributed by Oxford University Computing Services on behalf of the BNC Consortium: ⟨**http://www.natcorp.ox.ac.uk/**⟩.

Brown, J. Erin. 2010. Coordinated multi-modal expression and embodied meaning in the emergence of symbolic communication. In *The Evolution of Language: Proceedings of the 8th International Conference*, ed. Andrew D. M. Smith, Marieke Schouwstra, Bart de Boer, and Kenny Smith, 375–376. London : World Scientific.

Brown, Roger W. 1958. *Words and things*. Glencoe, IL: The Free Press.

Brown, Roger W., Abraham H. Black, and Arnold E. Horowitz. 1955. Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology* 50:388–393.

Brown, Roger W., and Ronald Nuttall. 1959. Methods in phonetic symbolism experiments. *Journal of Abnormal and Social Psychology* 59:441–445.

Calvert, Gemma A., Charles Spence, and Barry E. Stein, eds. 2004. *The Handbook of Multisensory Processes*. Boston, MA : MIT Press.

Christensen, Carol M. 1980. Effects of taste quality and intensity on oral perception viscosity. *Perception and Psychophysics* 28:315–320.

Cowey, Alan, and Lawrence Weiskrantz. 1975. Demonstration of cross-modal matching in rhesus monkeys. Macaca mulatta. *Neuropsychologia* 13 (1): 117–120.

Crisinel, Anne-Sylvie, and Charles Spence. 2009. Implicit associations between basic tastes and pitch. *Neuroscience Letters* 464:39–42.

Crisinel, Anne-Sylvie, and Charles Spence. 2010. A sweet sound? Exploring implicit associations between basic tastes and pitch. *Perception* 39:417–425.

Davenport, R. K., C. M. Rogers, and I. Steele Russell. 1973. Cross-modal perception in apes. *Neuropsychologia* 11 (1): 21–28.

Davis, R. 1961. The fitness of names to drawings: a cross-cultural study in Tanganyika. *British Journal of Psychology* 52:259–268.

Day, Sean A. 2005. Some demographic and socio-cultural aspects of synesthesia. In *Synesthesia: Perspectives from Cognitive Neuroscience*, ed. Lynn C. Robertson and Noam Sagiv, 11–33. Oxford : Oxford University Press.

Deacon, Terrence. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain*. New York : W.W. Norton and Company.

Diffloth, Gérard. 1994. i: big, a: small. In *Sound Symbolism*, ed. Leanna Hinton, Johanna Nichols, and John J. Ohala, 107–114. Cambridge: Cambridge University Press.

Dingemanse, Mark. 2009. *Ideophones in unexpected places*. In Proceedings of Conference on Language Documentation and Linguistic Theory 2, ed. P. K. Austin, O. Bond, M. Charette, D. Nathan, and P. Sells, 83–97. London : SOAS.

Dingemanse, Mark. 2011. Ideophones and the aesthetics of everyday language in a West-African society. *Senses and Society* 6 (1): 77–85.

Drellishak, Scott. 2007. Statistical techniques for detecting and validating phonesthemes. Paper presented at the Linguistics Society of America Annual Meeting, Anaheim, CA, January.

Dunbar, Robin. 2003. The social brain: mind, language and society in evolutionary perspective. *Annual Review of Anthropology* 32:163–181.

Farmer, Thomas A., Morten H. Christiansen, and Padraic Monaghan. 2006. Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America* 103:12203–12208.

Fischer, S. 1922. Über das Entstehen und Verstehen von Namen. *Archiv für die gesamte Psychologie* 42:335–368.

(p. 895) Fisher-Jørgensen, Eli. 1978. On the universal character of phonetic symbolism with special reference to vowels. *Studia Linguistica* 32:80–90.

Fitch, W. Tecumseh. 2000. The evolution of speech: A comparative review. *Trends in Cognitive Sciences* 4 (7): 258–267.

Fitch, W. Tecumseh. 2010. *The Evolution of Language*. Cambridge: Cambridge University Press.

Fox, Charles W. 1935. An experimental study of naming. *American Journal of Psychology* 47:545–579.

Frankis, J. 1991. Middle English ideophones and the evidence of manuscript variants: Explorations in the lunatic fringe of language. In *Language Usage and Description: Studies Presented to N.E. Osselto on the Occasion of his Retirement*, ed. I. T. van Ostade, 17–25. Amsterdam: Rodopi.

Gallace, Alberto, Erica Bochin, and Charles Spence. 2011. On the taste of ‘Bouba’ and ‘Kiki’: An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience* 2 (1): 34–46.

Gasser, Michael. 2004. The origins of arbitrariness in language. In *Proceedings of the Cognitive Science Society Conference*, 434–439. Hillsdale, NJ: LEA.

Harnad, Steven. 1990. The symbol grounding problem. *Physica D* 42:325–336.

Hinton, Leanne, Johanna Nichols, and John J. Ohala. 1994. *Introduction: Sound symbolic processes*. In Sound Symbolism, ed. Leanne Hinton, Johanna Nichols, and John J. Ohala, 1–14. Cambridge: Cambridge University Press.

Huang, Yan-Huang, Sawat Pratoomraj, and Ronald C. Johnson. 1969. Universal magnitude symbolism. *Journal of Verbal Learning and Behavior* 8:155–156.

Humphrey, Nicholas. 2012. ‘This chimp will kick your ass at memory games—but how the hell does he do it?’ *Trends in Cognitive Science* 16 (7):353–355.

Hutchins, Sharon S. 1998. The psychological reality, variability, and compositionality of English phonesthemes. Doctoral dissertation, Emory University.

Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203:88–96.

Imai, Mutsumi, Sotaro Kita, Miho Nagumo, and Hiroyuki Okada. 2008. Sound symbolism facilitates early verb learning. *Cognition* 109 (1): 54–65.

Irwin, Francis W., and Elizabeth Newland. 1940. A genetic study of the naming of visual figures. *Journal of Psychology* 9:3–16.

Iwasaki, Noriko, David P. Vinson, and Gabriella Vigliocco. 2007. What do English speakers know about gera-gera and yota-yota? A cross-linguistic investigation of mimetic words for laughing and walking. *Japanese Language Education Around the Globe* 17:53–78.

Iwasaki, Noriko, David P. Vinson, and Gabriella Vigliocco. 2007. *How does it hurt, ‘kiri-kiri’ or ‘siku-siku’? Japanese mimetic words of pain perceived by Japanese speakers and English speakers*. In Applying theory and research to learning Japanese as a foreign language, ed. Masahiko Minimai, 2–19. Newcastle: Cambridge Scholars.

Jacobs, Gerald H. 2009. Evolution of color vision in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1531): 2957–2967.

Jesperson, Otto. 1933. *Symbolic value of the vowel 'i'*. In Linguistica: Selected papers of O. Jesperson in English, French and German, 283–303. Copenhagen, Denmark: Levin and Munksgaard.

Johnson, Ronald C. 1967. Magnitude symbolism of English words. *Journal of Verbal Learning and Verbal Behavior* 6 (4): 508–511.

Kirby, Simon, Hannah Cornish, and Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America* 105 (31): 10681–10686.

(p. 896) Klank, Linda J., Yan-Huang Huang, and Ronald C. Johnson. 1971. Determinants of success in matching word pairs in tests of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior* 10:140–148.

Köhler, Wolfgang. 1929. *Gestalt Psychology*. 1st ed. New York, NY: Liveright.

Köhler, Wolfgang. 1947. *Gestalt Psychology*. 2nd ed. New York, NY: Liveright.

Kovic, Vanja, Kim Plunkett, and Gert Westermann. 2010. The shape of words in the brain. *Cognition* 114:19–28.

Kucera, H., and W. N. Francis. 1967. *Computational Analysis of Present-day American English*. Providence, RI: Brown University Press.

Landauer, Thomas K., and Susan T. Dumais, 1997. A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review* 104:211–240.

Lebert, Marie. 2005. Project Gutenburg, from 1971 to 2005. ⟨**http://www.etudes-francaises.net/dossiers/gutenberg_Eng.htm**⟩.

Ludwig, Vera U., Ikuma Adachi, and Tetsuro Matsuzawa. 2011. Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans, *Proceedings of the National Academy of Sciences of the United States of America* 108:20661–20665.

Maltzman, Irving, Lloyd Morisett, and Lloyd O. Brooks. 1956. An investigation of phonetic symbolism. *Journal of Abnormal and Social Psychology* 53:249–251.

Matsuzawa, Tetsuro. 2009. Symbolic representation of number in chimpanzees. *Current Opinion in Neurobiology* 19:1–7.

Maurer, Daphne, Thanujeni Pathman, and Catherine J. Mondloch. 2006. The shape of boubas: Sound shape correspondences in toddlers and adults. *Developmental Science* 9:316 – 322.

McNeill, David. 2005. *Gesture and Thought*. Cambridge: Cambridge University Press.

Mithen, Steven. 2005. *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London : Weidenfeld & Nicolson.

Monaghan, Padraic, Nick Chater, and Morten H. Christiansen. 2005. The differential contribution of phonological and distributional cues in grammatical categorisation. *Cognition* 96:143–182.

Monaghan, Padraic, Morten H. Christiansen, and Nick Chater. 2007. The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology* 55:259–305.

Monaghan, Padraic, Morten H. Christiansen, and Stanka A. Fitneva. 2010. *Balancing arbitrariness and systematicity in language evolution*. In: A.D.M. Smith, M. Schouwstra, B. de Boer, and K. Smith (Eds.), The Evolution of Language: Proceedings of the 8th International Conference (pp. 465–466). London : World Scientific.

Monaghan, Padraic, Morten H. Christiansen, and Stanka A. Fitneva. 2011. The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General* 140:325–347.

Monaghan, Padraic, and Karen Mattock. 2009. Cross-situational language learning: The effects of grammatical categories as constraints on referential labelling. Proceedings of the 31st Cognitive Science Society Conference, 2226–2231. Mahwah, NJ: Lawrence Erlbaum.

Monaghan, Padriac, Karen Mattock, and Peter Walker. 2012. The role of sound symbolism in word learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 38 (5):1152–1164.

Neilsen, Alan, and Drew Rendall. 2011. The sound of round: Evaluating the role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology* 65:115–124.

(p. 897) Neilsen, 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):115–125.

Newman, Stanley S. 1933. Further experiments in phonetic symbolism. *American Journal of Psychology* 45:53–75.

Nuckolls, Janis B. 1999. The case for sound symbolism. *Annual Review of Anthropology* 28:225–252.

Nygaard, Lynn C., Alison E. Cook, and Laura L. Namy. 2009. Sound to meaning correspondences facilitate word learning. *Cognition* 112 (1): 181–186.

Ohala, John J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41:1–16

Ohala, John J. 1994. *The frequency code underlies the sound-symbolic use of voice pitch*. In Sound Symbolism, ed. Leanne Hinton, Johanna Nichols, and John J. Ohala, 325–347. Cambridge: Cambridge University Press.

O'Grady, William, Michael Dobrovolsky, and Francis Katamba. 1996. *Contemporary Linguistics: An Introduction*. Essex: Pearson Education Limited.

Otis, Katya, and Eyal Sagi. 2008. Phonoaesthemes: A corpora-based analysis. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, ed. B. C. Love, K. McRae, and V. M. Sloutsky, 65–70. Austin, TX: Cognitive Science Society.

Oxford English Dictionary. 2011. Oxford : Oxford University Press. ⟨**http://oed.com**⟩.

Parault, Susan J., and Paula J. Schwanenflugel. 2006. Sound-symbolism: A piece in the puzzle of word learning. *Journal of Psycholinguistic Research* 35:329–351.

Parault, Susan J., and M. Parkinson. 2008. Sound symbolic word learning in the middle grades. *Contemporary Educational Psychology* 33:647–671.

Perniss, Robin L. Thompson, and Gabriella Vigliocco. 2010. Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology* 1:1–15.

Ramachandran, Vilayanur S. (2004). *A brief tour of human consciousness*. New York : Pi Press.

Ramachandran, Vilayanur S. and Edward M. Hubbard. 2001. Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies* 8 (1): 3–34.

Ramachandran, Vilayanur S. and Edward M. Hubbard. 2005. Synaesthesia: A window into the hard problem of consciousness. In *Synaesthesia: Perspectives from Cognitive Neuroscience* ed. Lynn C. Robertson and Noam Sagiv, 127–189. New York : Oxford University Press.

Ramachandran, Vilayanur S., and Lindsay M. Oberman. 2006. Broken mirrors: a theory of autism. *Scientific American* 295:62–69.

Rhodes, R. 1981. On the semantics of Ojibwa verbs of breaking. In *Papers of the twelfth Algonquian conference*, ed. W. Cowan, 47–56. Ottawa, Canada: Carleton University Press.

Rogers, Susan K., and Abraham S. Ross. 1975. A cross-cultural test of the maluma-takete phenomenon. *Perception* 5 (2): 105–106.

Rose, Susan A., Allen W. Gottfried, and Wagner H. Bridger. 1981. Cross-modal transfer in 6-month-old infants. *Developmental Psychology* 17 (5): 661–669.

Sapir, Edward. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12:239–255.

Saussure, Ferdinand. 1959. *Course in general linguistics*. New York : Philosophical Library.

Savage-Rumbaugh, Sue, Rose A. Sevcik, and William D. Hopkins. 1988. Symbolic cross-modal transfer in two species of chimpanzees. *Child Development* 59 (3): 617–625.

Searle, John R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3 (3): 417:457.

(p. 898) Shankar, Maya U., Carmel A. Levitan, John Prescott, and Charles Spence. 2009. The influence of color and label information on flavor perception. *Chemosensory Perception* 2:53: 58.

Shillcock, Richard, Simon Kirby, Scott McDonald, and Chris Brew. 2001. *Filled pauses and their status in the mental lexicon*. In Disfluency in Spontaneous Speech (DiSS '01), 53–56. ⟨**http://www.isca-speech.org/archive_open/diss_01/dis1_053.html**⟩.

Simner, Julia, Christine Cuskley, and Simon Kirby. 2010. What sound does that taste? Cross-modal mappings across gustation and audition. Perception 39:553–569.

Simner, Julia, Jamie Ward, Monika Lanz, Ashok Jansari, Krist Noonan, Louise Glover, and David Oakley. 2005. Non-random associations of graphemes to colours in synaesthetic and normal populations. *Cognitive Neuropsychology* 2:2069–1085.

Simner, Julia, and Vera Ludwig. 2009. What colour does that feel? Cross-modal correspondences from touch to colour. Paper presented at the Third International Conference of Synaesthesia and Art, Granada, Spain, April.

Slobin, Dan I. 1968. Antonymic phonetic symbolism in three natural languages. *Journal of Personality and Social Psychology* 10 (3): 301–305.

Spence, Charles, and Alberto Gallace. 2011. Tasting shapes and words. *Food Quality and Preference* 22 (3): 290–295.

Stevenson, Richard J., John Prescott, and Robert A. Boakes. 1999. Confusing tastes and smells: How odours can influence the perception of sweet and sour tastes. *Chemical Senses* 24:627–635.

Tallerman, Maggie. 2007. Did our ancestors speak a holistic protolanuage? *Lingua* 117 (3): 579–604.

Tamariz, Monica. 2005. Configuring the phonological organization of the mental lexicon using syntactic and semantic information. In *Proceedings of the 27th Annual Conference* *of the Cognitive Science Society*, ed. Bruno G. Bara, Lawrence Barsalou, and Monica Bucciarelli, 2145–2150. Mahwah, NJ: Earlbaum.

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

Tomasello, Michael. 2008. *Origins of Human Communication*. Boston, MA : MIT Press.

Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28:675–691.

Traunmuller, Hartmut. 1996. Sound symbolism in deictic words. Paper presented at Fonetik 96, Swedish Phonetics Conference, Nässlingen, Sweden, May.

Tufvesson, Sylvia. 2011. Analogy-making in the Semai sensory world. *Senses and Society* 6 (1): 86–95.

Ultan, Russell. 1978. Size-sound symbolism. In *Universals of Human Language*, Volume 2, ed. Joseph H. Greenberg, Charles R. Ferguson, and Edith Moravcsik, 527–568. Stanford, CA: Stanford University Press.

Usnadze, D. 1924. Ein experimenteller Beitrag zum Problem der psychologischen Grundlagen der Namengebung. *Psychologische Forschung* 5:24–43.

Verhagen, Justus V., and Lina Engelen. 2006. The neurocognitive bases of human multimodal food perception: Sensory integration. *Neuroscience and Biobehavioral Reviews* 30:613: 650.

Ward, Jamie, Brett Huckstep, and Elias Tsakanikos. 2006. Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex* 42:264–280.

(p. 899) Ward, Jeanette P., Arthur L. Yehle, and R. Stephen Doerflein. 1970. Cross-modal transfer of a specific discrimination in the bushbaby (Galago senegalensis). *Journal of Comparative and Physiological Psychology* 73 (1): 74–77.

Werner, Heinz, and Seymour Wapner. 1952. Toward a general theory of perception. *Psychological Review* 59:324–338.

Werner, Heinz. 1957. *Comparative Psychology of Mental Development*. New York : International University Press.

Westbury, Chris. 2005. Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language* 93 (1): 10–19.

Westermann, D. 1927. Laut, Ton und Sinn in westafrikanische Sudansprachen. In *Festschrift Meinhof*, ed. F. Boas, O. Dempwolff, G. Panconelli-Calzia, A. Werner, and D. Westermann, 315–328. Hamburg: Kommissionsverlag von L. Friederichsen & Co.

Woodworth, Nancy L. 1991. Sound symbolism in proximal and distal forms. *Linguistics* 29:273–299.

Wray, Alison. 1998. Protolanguage as a holistic system for social interaction. *Language and Communication* 18:47–67.

Yaro, Caroline, and Jamie Ward. 2007. Searching for Shereshevskii: What is superior about the memory of synaesthetes? *Quarterly Journal of Experimental Psychology* 60:681: 695.

Yoshida, H., and L. Smith. 2006. Dynamic properties of form and meaning and children's learning of verbs. Paper presented at the 15th International Conference of Infant Studies, Kyoto, Japan, June.

Yu, Chen, and Linda B. Smith. 2007. Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science* 18:414–420.

Zampini, Massimiliano, and Charles Spence. 2004. The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *Journal of Sensory Studies* 19:347–363.

Zampini, Massimilliano, Daniel Sanabria, Nicola Phillips, and Charles Spence. 2007. The multisensory perception of flavor: Assessing the influence of color cues on flavor discrimination responses. *Food Quality and Preference* 18:975–984.

Zhao, Grace Q, Yifeng Zhang, Mark A. Hoon, Jayaram Chandrashekar, Isolde Erlenbach, Nicholas J. P. Ryba, and Charles S. Zuker. 2003. The receptors for mammalian sweet and umami taste. *Cell* 115:255–266.

Notes:

(1) This chapter will not go into detail regarding the nature of synesthesia and its various varieties, as this is well covered in the remainder of the volume.

(2) One notable exception, Rogers and Ross (1975), will be discussed in further detail later.

(3) Hinton and colleagues (1994) use the term synesthetic here in a slightly unconventional manner: they do not suggest these forms are confined to synesthetes or derive from their unique associations, but simply that there are shared cross-modal associations underlying some sound symbolic forms.

(4) For example, we could imagine a fictional lexicon where all round objects have the /e/ vowel present in the word *takete*. This language would be non-arbitrary, since there is a regular relationship between a particular vowel and roundedness—but it would not be naturally motivated. Instead, we know from Köhler's naming experiments that the word

takete fits better with angular objects than round ones. But non-arbitrariness can still logically be present even if it contradicts the natural motivatedness evident from Köhler's experiment.

(5) Open class words, or content words, form a category of words which can be readily added to, and includes all nouns and verbs. Closed class words, also known as function words, form a category to which new items are rarely added, including things like prepositions (e.g., *in*, *of*, *at*), determiners (e.g., *the*, *a*, *this*) and conjunctions (e.g., *and*, *but*).

(6) Drellishak (2007) did not confine his search to monomorphemic words or even content words, which may have adversely affected his results.

(7) Only monomorphemic roots are considered here; types identified by Bergen such as *glistening* and *glimmering* were considered instances of *glisten* and *glimmer* respectively.

(8) Vowel height distinctions are now well documented in many ideophone systems (Dingemanse 2011).

(9) Despite these examples, English appears to be an exception to these trends. Newman (1933) and Brown (1958) both failed to find a significant correlation among English words denoting size and vowel quality.

(10) See Diffloth (1994) and Bauer (1996) for a discussion of languages that do not conform to this pattern.

(11) Note that vowel quality is independent of voice pitch. For example, the word *teeny* can be sung in a high or low pitch.

(12) At one end of this scale, subjects heard a continuous vowel sound; as the value of the scale increased, periods of silence interrupted the continuous vowel resulting in pockets of voicing, mimicking words like kiki or takete.

(13) This was effected by changing all the formants of filtered white noise simultaneously from 0 to 5000 Hz, resulting in a scale from low-pitched white noise to high-pitched white noise.

Christine Cuskley

Christine Cuskley is a postdoctoral researcher at the Institute for Complex Systems (ISC-CNR) at La Sapienza University in Rome, Italy.

Simon Kirby

Simon Kirby is Professor of Language Evolution at the University of Edinburgh and co-founder of the Language Evolution and Computation Research Unit. He has pioneered the application of computational and mathematical modelling techniques to traditional issues in language acquisition, change, and evolution. In particular, he has developed an experimental and computational approach to cultural evolution called Iterated Learning, which treats language as a complex adaptive system operating on multiple interacting time-scales.