COMPARATIVE COGNITION & BEHAVIOR REVIEWS

Imitating Sounds: A Cognitive Approach to Understanding Vocal Imitation

Eduardo Mercado III
University at Buffalo, The State University of New York

James T. Mantell
St. Mary's College of Maryland

Peter Q. Pfordresher
University at Buffalo, The State University of New York

Vocal imitation is often described as a specialized form of learning that facilitates social communication and that involves less cognitively sophisticated mechanisms than more "perceptually opaque" types of imitation. Here, we present an alternative perspective. Considering current evidence from adult mammals, we note that vocal imitation often does not lead to learning and can involve a wide range of cognitive processes. We further suggest that sound imitation capacities may have evolved in certain mammals, such as cetaceans and humans, to enhance both the perception of ongoing actions and the prediction of future events, rather than to facilitate mate attraction or the formation of social bonds. The ability of adults to voluntarily imitate sounds is better described as a cognitive skill than as a communicative learning mechanism. Sound imitation abilities are gradually acquired through practice and require the coordination of multiple perceptual-motor and cognitive mechanisms for representing and generating sounds. Understanding these mechanisms is critical to explaining why relatively few mammals are capable of flexibly imitating sounds, and why individuals vary in their ability to imitate sounds.

Keywords: mimicry; copying; social learning; singing; emulation; imitatible; convergence; imitativeness

In his seminal text, *Habitat and Instinct*, Lloyd Morgan (1896, p. 166) describes two general kinds of imitation: instinctive imitation and intelligent or voluntary imitation. The examples he provides of intelligent imitation mostly involve reproducing sounds—a child copies words used by his companions, a mockingbird imitates the songs of 32 other bird species, a jay imitates the neighing of a horse, and so on. In fact, most of the examples of "imitation proper"

that Morgan provides consist of birds reproducing the sounds of other species. Similarly, Romanes (1884) focuses almost exclusively on reports of birds imitating songs, music, and speech in his discussion of imitation. These classic portrayals of vocal reproductions as providing the best and clearest examples of imitation stand in stark contrast to current psychological discussions of imitation, which often classify examples such as those given by Romanes and Morgan as non-imitative performances that merely resemble actual imitation (Byrne, 2002; Heyes, 1996). When did phenomena that were once considered archetypal examples of voluntary imitation transform into a footnote of modern cognitive theories? Was some discovery made that fundamentally changed our scientific understanding of the processes underlying vocal imitation? Have psychologists or biologists succeeded in explaining what vocal imitation is and how it works to the point where little can be gained from further study? Or, have theoretical assumptions led scientists to underestimate the cognitive mechanisms required for an individual to be able to flexibly imitate sounds?

In the present article, we attempt to identify what exactly vocal imitation entails, and to assess whether current explanatory frameworks adequately account for this ability, including its apparent rarity among mammals. Past theoretical and empirical considerations of vocal imitation have often focused on the ability of birds to learn songs or reproduce speech (Kelley & Healy, 2010; Margoliash, 2002; Nottebohm & Liu, 2010; Pepperberg, 2010; Tchernichovski, Mitra, Lints, & Nottebohm, 2001), especially during development, or on the sophisticated ways in which birds interactively copy songs (Akcay, Tom, Campbell, & Beecher, 2013; Molles & Vehrencamp, 1999; J. J. Price & Yuan, 2011; Searcy, DuBois, Rivera-Caceres, & Nowicki, 2013). In contrast, vocal imitation by mammals other than humans has received little attention. When mammalian vocal imitation has been discussed, it typically has been described as a vocal learning mechanism, because of its presumed involvement in vocal repertoire development (Janik & Slater, 1997, 2000; Tyack, 2008). Although imitation clearly has an important role in learning, imitation by definition involves performance (via reproduction) and thus can exhibit varying degrees of success. Moreover, the effectiveness of imitation is itself an index of learning (e.g., we say a tennis player has reached expertise in serving when he or she can demonstrate the coordination exhibited by a professional). Thus, we argue here that vocal imitation by adult mammals is better viewed as performance of a learned skill, and that a closer examination of those species and individuals that have acquired this skill to a high degree can clarify the mechanisms that underlie vocal imitation abilities. Currently, the only mammals that have clearly demonstrated the ability to voluntarily imitate sounds are primates (particularly humans) and cetaceans (whales and dolphins). The main goals of this article are to reassess the available evidence on vocal imitation in these two groups and to provide new perspectives on how to better integrate future investigations of vocal imitation phenomena.

The paper is divided into six sections. In the first two sections, we consider alternate conceptualizations of vocal imitation (both historical and modern) that have different theoretical implications for the origin and role of vocal imitation. These alternate frameworks function as hypotheses against which we compare the literature summarized in subsequent sections. In section three, we discuss possible constraints on vocal imitation with respect to sounds that are imitatible by human and non-human primates, and also consider the degree to which the vocal motor system, as opposed to other motor systems, is attuned to the imitation of sound. Section four evaluates past reports that cetaceans, a group of mammals famous for their vocal flexibility, are capable of imitating sounds. Consideration of evidence from both primates and cetaceans leads to the proposal that sound imitation may serve a critical role for spatial perception and the coordination of actions (section five), in contrast to other accounts, which focus on its role in the development of social communication. Finally, in the sixth section we discuss possible mechanisms for vocal imitation, highlighting an existing computational model of speech learning and imitation that may provide an integrative theoretical framework for conceptualizing the representational mechanisms underlying the sound imitation abilities of mammals.

I. What Is Vocal Imitation?

Over the past century, researchers have used varying terminology to describe animals' reproduction of sounds. In some cases, the same term has been used to describe different classes of phenomena. In others, different terms have been applied to the same phenomenon. For instance, the terms *vocal mimicry* and *vocal copying* have often been used as either synonyms for *vocal imitation*, or as a way to distinguish particular kinds of imitative or non-imitative vocal processes (Baylis, 1982; Morgan, 1896; Witchell, 1896). Marler (1976a) distinguished cases in which vocal production is modified as a result of auditory experience (*vocal learning*) from cases in which an individual produces sounds of a novel morphology by imitating previously experienced sounds (*vocal imitation*). Similarly, in their description of vocal developmental processes in bottlenose dolphins (*Tursiops truncatus*), McCowan and Reiss (1995) distinguished vocal learning, which they suggest occurs mainly during development, from *vocal mimicry*, which they describe as an imitative process that contributes

to vocal learning (see Wickler, 2013, for a discussion of how the term *mimicry* might best be applied to sound production). To avoid potential confusion, we provide a glossary detailing our use of terminology (Table 1).

There is general consensus that vocal imitation must involve some attempt (intentional or incidental) to match an auditory event with the vocal motor system. The nature of this ability, however, has been a point of debate. Early on, Thorndike (1911) rejected the proposals of Morgan (1896) and Romanes (1884) that vocal reproductions by birds were examples of imitation. Thorndike seemed to believe that vocal imitation required less sophisticated mental capacities than other kinds of imitation. He claimed that the ability to copy sounds was a specialized capacity possessed by a few select bird species, and that, "we cannot . . . connect these phenomena with anything found in mammals or use them to advantage in a discussion of animal imitation" (1911, p. 77). Many researchers have subsequently endorsed Thorndike's characterization of vocal imitation, either explicitly or implicitly (Byrne & Russon, 1998; Galef, 1988; Heyes, 1994; Shettleworth, 1998). For example, Tyack and Clark (2000) described the vocal imitation abilities of cetaceans as "the most unusual specialization in cetaceans." In contrast, several experimental psychologists have argued that vocal imitation abilities are not specialized at all, but simply reflect basic mechanisms of conditioning (reviewed by Baer & Deguchi, 1985; Kymissis & Poulson, 1990). Some researchers describe vocal imitation as a specialized social learning mechanism that enables individuals to rapidly acquire new communicative signals (Bolhuis, Okanoya, & Scharff, 2010; Janik & Slater, 2000; Kelley & Healy, 2011; Sewall, 2012; Tyack, 2008), whereas others classify all instances of vocal reproduction as non-imitative phenomena (Byrne, 2002; Galef, 1988; Heyes, 1996; Zentall, 2006). In the following, we critically consider each of these approaches to explaining what vocal imitation is, noting their strengths and limitations.

Is Vocal Imitation an Outcome of Instrumental Conditioning?

Instrumental (operant) conditioning is a learning process in which the consequences of an action determine its future likelihood of occurring (Domjan, 2000; Immelmann & Beer, 1989). Miller and Dollard (1941) suggested that apparently copied actions (including vocal acts) might in some cases only match by coincidence, having been reinforced independently of any similarities in performance. In such situations, some apparent cases of vocal "imitation" can be viewed as an instance of instrumental conditioning, referred to as *matched-dependent behavior*. For example, one could train a dog to produce whining sounds whenever it hears the cries of a baby. The dog's whines might be acoustically similar in certain respects to the baby's cries, but these similarities are coincidental; the dog might just as easily have been trained to bark or to open a door whenever it heard the cries. Though the trained behavior may match the discriminative stimulus, it is not the degree of match per se that leads to reinforcement. Miller and Dollard distinguished matched-dependent behavior from *copying*, in which the presence of reinforcement is contingent on the successfulness of matching. Learning to sing a melody by matching the sounds produced by an instructor would be an example of this kind of vocal copying. The teacher uses feedback to reinforce correct matches and to punish mismatches. The main difference between matched-dependent behavior and copying in Miller and Dollard's framework is that a copier directly compares his acts (or their outcomes) with those of a target to evaluate their similarity, such that the level of detected similarity becomes a cue controlling behavior. A commonality across matched-dependent behavior and copying is that changes in vocal behavior are described as reflecting reinforcement histories alone, and thus do not require any specialized learning mechanisms.

Miller and Dollard's explanation for acts of vocal imitation (construed as copying or matched-dependent behavior) continues to be endorsed by some psychologists (Heyes, 1994, 1996). For instance, Heyes (1994, p. 224) suggested that, "copying is virtually synonymous with vocal imitation." This interpretation of vocal imitation rests on four major assumptions: (1) a vocalizing individual initially produces sounds at random, after which a subset are rewarded; (2) all a vocalizing individual needs to be able to do to reproduce a sound is recognize similarities between produced sounds and previously perceived sounds; (3) mismatches between an internally stored model of a previously experienced sound and percepts of produced sounds drive instrumental conditioning (see also the discussion of auditory template matching in section six); and (4) such mismatches correspond to errors in production. Like those of many previous researchers, the examples of vocal copying provided by Heyes focus mainly on song learning and speech reproduction by birds.

Mowrer (1952, 1960) similarly proposed that vocal imitation, even in humans, was a consequence of instrumental conditioning rather than a specialized ability. He suggested that for such conditioning to occur, a sound produced by a model initially had to be established as a secondary reinforcer by being associated with pleasant outcomes. Later, a "babbling" individual (e.g., a parrot or human infant) might occasionally make a similar sound. Assuming that the vocalizing individual generalized from its past experiences with the secondary reinforcer, hearing the self-produced sound would reinforce the immediately

preceding vocal act. The more similar to the original sound the babbled sound was, the more reinforcing it should be, leading to a kind of autoshaping or successive approximation in which the vocalizing individual is differentially self-rewarded based on how closely it produces copies of the original sound. By this account, vocal imitation is simply an automatic, trial-and-error process that depends on initial rewards from another organism to establish certain sounds as secondary reinforcers. Mowrer thus describes vocal imitation phenomena as the result of latent learning about associations between sounds and rewards. He notes that sounds can only maintain their efficacy as secondary reinforcers if they are occasionally supplemented by external social reinforcements. Thus, as in Miller and Dollard's (1941) explanation of copying, Mowrer claimed that feedback from a teacher is critical for vocal imitation to occur.

Baer and colleagues (1967) later showed that explicitly trained vocal imitation in children immediately generalized to novel sounds. They suggested that topographical similarity between a performed act and a perceived act could become a conditioned reinforcer, which could lead to generalized imitation across different stimuli (see also Garcia, Baer, & Firestone, 1971; Gewirtz & Stingle, 1968; Zentall & Akins, 2001). Their proposal parallels Miller and Dollard's (1941) claim that recognition of similarity is a critical component of vocal copying, and again requires reinforcement of imitative vocal acts by a teacher. In Baer and colleagues' generalized imitation framework, a vocalization is imitative if it occurs after a vocal act demonstrated by another individual, and if the form of the model's vocalization determines the form of the copier's vocalization. The proposal that generalized vocal imitation can be viewed as a consequence of operant conditioning has received some support from recent studies of the role of vocal imitation in speech development by children (Poulson, Kymissis, Reeve, Andreators, & Reeve, 1991; Poulson, Kyparissos, Andreatos, Kymissis, & Parnes, 2002).

Collectively, past theoretical analyses of vocal imitation by experimental psychologists have often focused on establishing that this phenomenon can be viewed as an outcome of instrumental conditioning with few if any unique characteristics. These accounts generally do not explain why vocal imitation abilities are absent in most mammals. Given that mammals are quite capable of being instrumentally conditioned, some researchers have suggested that the rarity of vocal imitation abilities in mammals reflects limitations in vocal control (Arriaga & Jarvis, 2013; Deacon, 1997; Fitch, 2010; Mowrer, 1960). However, this explanation remains speculative (Lieberman, 2012), and others have suggested that what is missing are mechanisms that make it possible for an organism to adaptively adjust existing

vocal control mechanisms. For example, Moore (2004) hypothesized that an organism must possess specialized imitative learning mechanisms beyond those necessary for instrumental conditioning before vocal imitation becomes possible (see also Subiaul, 2010). Thus, Thorndike's (1911, p. 77) view that vocal imitation abilities are "a specialization removed from the general course of mental development," has resurged in recent years and is currently the dominant view among biologists studying vocal imitation.

Is Vocal Imitation a Specialized Type of Vocal Learning?

As noted above, Marler (1976a) defined vocal learning as a process whereby vocal production is modified as a result of auditory experience. More recently, this term has been used to refer to any type of learning that involves vocal systems (Arriaga & Jarvis, 2013). Several reviews of vocal learning by mammals suggest that it represents a specialized form (or actually several different forms) of motor learning (Egnor & Hauser, 2004; Janik & Slater, 1997, 2000; Jarvis, 2013; Sewall, 2012; Tyack, 2008). Within modern vocal learning taxonomies, vocal imitation is often described as a particular type of vocal learning called *vocal production learning* (Fitch, 2010; Tyack, 2008) or *production learning* (Byrne, 2002; Janik & Slater, 2000). Vocal production learning, defined as the ability to modify features of sounds based on auditory inputs, has been distinguished from *contextual learning*, which is said to consist of learning how to use or comprehend sounds (Figure 1). Janik and Slater (2000) divided vocal production learning into three "forms" depending on which articulators were thought to be involved. Tyack (2008) identified over a dozen forms based on how

Figure 1. Taxonomy proposed by Janik and Slater (2000) in which vocal learning is distinguished from contextual learning and subtypes of vocal learning are associated with different effectors. In this framework, vocal imitation of amplitude and duration features involves respiratory learning, imitation of frequency contours or pitch involves phonatory learning, and imitation of timbre involves articulatory learning.
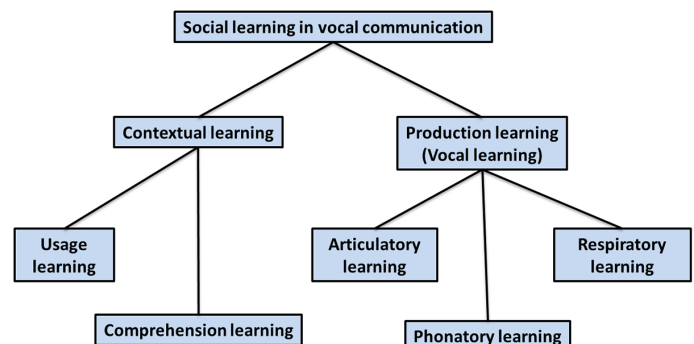
animals used the sounds (e.g., vocal production learning involving sounds used for echolocation). The three kinds of vocal production learning proposed by Janik and Slater correspond to acoustic features controlled by the vocalizing individual, including: (1) duration and amplitude; (2) pitch or frequency modulation; and (3) relative energy distribution or timbre. They hypothesize that modifying the duration and amplitude of a sound represents the simplest form of vocal learning (because these features depend mainly on respiratory control), and that modifying frequency components through control of vocal systems requires more advanced mechanisms (Janik & Slater, 1997, 2000). They also suggest that, because of its rarity, the ability to copy (i.e., imitate) novel sounds is the most advanced form of vocal learning.

A simple way of thinking about the distinction between contextual learning and production learning proposed by Janik and Slater (2000) is that contextual learning determines *when* animals produce the sounds they know how to make, whereas production learning determines *what* sounds they know how to make. For instance, situations in which animals respond to hearing certain sounds by producing similar sounds (e.g., dogs that bark when they hear barking or infants that cry when they hear crying) would not qualify as vocal imitation or vocal learning by these criteria (Andrew, 1962). These cases would meet the criteria for contextual learning, however, because sound usage is context dependent. Such phenomena are typically referred to as instances of *vocal contagion* (Piaget, 1962).

Byrne (2002), following the terminology proposed by Janik and Slater (2000), describes instances of vocal contagion as a kind of contextual learning[1] in which heard sounds prime particular vocal acts, a process that he refers to as *response facilitation*. He describes vocal production learning as a potentially more interesting case because it includes the generation of new vocal acts and therefore requires more than just response facilitation. Nevertheless, he suggests that such vocal acts are not imitative, because in some cases only the outcomes of the actions are reproduced (see also Morgan, 1896). For instance, a mynah bird reproducing speech sounds cannot replicate the speech acts of a human, because the bird does not use the same vocal organs to produce sounds (however, see Beckers, Nelson, & Suthers, 2004; Patterson & Pepperberg, 1994).

Vocal imitation of novel sounds often has been touted as the clearest evidence of vocal production learning (Fitch, 2010; Janik, 2000; Tyack, 2008). The basic logic underlying

past emphasis on the imitation of novel sounds is that if a vocalization is not novel, then one cannot be sure that imitation actually occurred. The origins of this criterion can be traced to Thorpe (1956, p. 135), who proposed that, *"By true imitation is meant the copying of a novel or otherwise improbable act or utterance, or some act for which there is clearly no instinctive tendency."* Herman (1980) was one of the first to suggest that copying novel sounds requires more sophisticated cognitive mechanisms than modifying features of existing vocalizations (see also Baylis, 1982). He noted that many mammalian species can be trained to adjust their existing vocalizations into new forms or usage patterns (Adret, 1993; Johnson, 1912; Koda, Oyakawa, Kato, & Masataka, 2007; Molliver, 1963; Myers, Horel, & Pennypacker, 1965; Salzinger, 1993; Salzinger & Waller, 1962; Schusterman, 2008; Schusterman & Feinstein, 1965; Shapiro & Slater, 2004), whereas few species show any ability or inclination to copy novel sounds. Hearing individuals vocalize in ways that resemble the vocalizations of other species forcefully suggests that one has witnessed an imitative act. However, as originally noted by both James (1890) and Thorndike (1911), observations of an organism producing a novel action that resembles human actions, however precisely, does not provide strong evidence that the organism is imitating. Conversely, the fact that production of familiar vocalizations can potentially be attributed to mechanisms other than imitation does not provide strong evidence that those vocalizations are not truly imitative. Such ambiguities severely limit the usefulness of current taxonomical approaches for describing and understanding vocal imitation processes.

Limitations of the Vocal Learning Framework

Problems with defining vocal imitation. Past emphasis on specifying criteria for reliably identifying instances of vocal imitation have led researchers to focus almost exclusively on situations in which similarities between a produced sound and other environmental sounds seem unlikely to have occurred by chance (e.g., when the sound is novel and acoustically complex). However, the fact that human observers perceive an animal's vocalizations as strikingly similar to a salient environmental sound (e.g., electronic sounds, speech, or melodies), either through subjective impressions or quantitative acoustic analyses, is no more evidence that vocal imitation has occurred than the fact that certain photos of the surface of Mars look like a face is evidence that aliens reconfigured the landscape into that shape. Videos showing examples of cats and dogs producing vocalizations that are aurally comparable to the phrase, "I love you," are now commonplace, and yet few if any scientists would view these as evidence that these pets are imitating human speech.

---

[1] By biologists' definition of learning, vocal contagion is a kind of learning because auditory inputs lead to a change in behavior. Psychologists would instead classify vocal contagion as an elicited behavior or a reflexive action.

This is because many different mechanisms can lead to the production of atypical sounds. A novel vocalization might be a seldom-used part of an individual's repertoire, the result of some combination of previously learned vocal acts, or an aberration resulting from atypical genetics, diseases, or neural deficits. If an elephant produces a sound that resembles that of trucks (Poole, Tyack, Stoeger-Horwath, & Watwood, 2005) or speech (Stoeger et al., 2012), it remains possible that these sounds are ones that a small number of elephants infrequently make, independently of whether they have ever heard trucks or speech. Alternatively, vocalizations may have been modified through differential reinforcement to more closely resemble those of environmental sounds (which would represent a case of contextual learning using the taxonomy shown in Figure 1). Consequently, the novelty criterion does not reliably differentiate vocal imitation, vocal learning, or contextual learning.

In contrast, human research benefits from experimenters' ability to explicitly instruct human participants to intentionally imitate sound sequences (e.g., Mantell & Pfordresher, 2013). In these experiments, researchers assume that participants are following instructions and earnestly attempting to imitate sounds. This assumption implies that both accurate and inaccurate reproductions of any sound sequence (novel or familiar) are viewed as valid attempts at vocal imitation. A second branch of human vocal imitation research exploits the tendency for human speech patterns to align with (become more similar to) previously experienced speech stimuli (e.g., Goldinger, 1998). In these studies, experimenters instruct their subjects to perform a vocal task such as word naming without actually telling them to imitate sounds. Researchers assume that when an individual produces speech with features similar to those produced by a speaker s(he) has been recently exposed to, then this performance is indicative of spontaneous vocal imitation. Human vocal imitation research thus uses contextual factors as criteria for identifying imitative acts rather than idiosyncratic features of vocalizations.

A second criterion that has occasionally been used to exclude vocal performances from involving imitation is that an action (vocal or otherwise) can only be considered imitative if the specific movements of a model are replicated (e.g., Byrne, 2002). An oddity of this criterion is that if a dolphin were to copy sounds produced by a sea lion, then this would not count as imitation, because dolphins and sea lions have different sound producing organs. However, if a second dolphin copied the first dolphin's "barking," then this would count as imitation, because the imitator shares the same vocal organs as the model, and thus would likely replicate the sound producing movements of the model (see Wickler, 2013, for a more extensive critique of such distinctions). The logical consequence of this exclusionary criterion is that cross-species vocal imitation is impossible, because "imitation" requires identical physiological production constraints. However, defining imitation in this way does little to clarify what humans and other animals are doing when they seem to be copying sounds they have heard. Furthermore, it creates a false dichotomy between vocal imitation and other more visible forms of sound imitation (e.g., imitating a percussive rhythm).

Problems with equating learning and performance. A basic assumption underlying the claim that imitation of novel sounds provides the clearest evidence of vocal production learning is that, because the organism is producing an "otherwise improbable" sound that it has not been observed to produce before, it must have gained the ability to do so through its auditory experience (i.e., it must have learned how to produce the sound by hearing it). It is clear, however, that the individual imitating the sound had the vocal control mechanisms necessary for producing the novel sound prior to ever hearing that sound. Hearing the novel sound merely set the occasion for the individual to express an already present capacity. By analogy, a person who has never seen a motorcycle before and sits on one for the first time does not spontaneously acquire the motor control needed to sit on a motorcycle simply by seeing someone else sit on one. Instead, the person generalizes existing sitting skills to a novel object. Similarly, an organism that copies features of a novel sound it has heard is applying existing vocal production skills to a novel auditory object. For example, upon hearing a tugboat horn, a child may successfully reproduce the long, low, shifting, spectrotemporal pattern on her first try, at least in relative pitch terms. There is no reason to think that experience with an unfamiliar percept somehow endows an observer with previously unavailable motor control abilities (Galef, 2013). Consequently, reproduction of novel sounds does not provide clear evidence of vocal production learning, and such performances actually might be better viewed as evidence of contextual learning, because it is the context that determines when the individual reproduces a novel sound.

It is important to note that psychologists' use of the term *learning* differs from how this term is used colloquially and by biologists in that psychologists view learning as a long-lasting change in the mechanisms of behavior that results from past experiences with particular stimuli and responses (Domjan, 2000). In contrast, biologists' definition of *learning* as "behavioral changes effected by experience" (Immelmann & Beer, 1989), makes no distinction between short- and long-term changes, emphasizes changes in actions rather than changes in mechanisms, and makes no attempt to specify why an action changed. Psychologists

do not consider changes in behavior as strong evidence of learning, because many experiences such as fatigue, hunger, pain, injury, motivation, and drunkenness can also produce changes in behavior. Furthermore, numerous experiments have shown that learning can occur without any overt changes in an organism's behavior. For these reasons, experimental psychologists have drawn a distinction between learning and *performance*—performance refers to what organisms do. How an organism performs often reflects past learning, but there is not a one-to-one mapping between performance and learning. Because biologists do not make a similar distinction, some phenomena they might classify as vocal learning do not meet psychologists' criteria for learning. In the following, we use the term *learning* in the psychological sense, but use the phrase "vocal learning" in the sense preferred by biologists (see Table 1).

From a psychological perspective, production of novel sounds provides no evidence that learning has occurred. In fact, the learning that enables an individual to reproduce a particular sound may occur long before the novel sound is actually produced. This is not to say that vocal imitation of familiar or novel sounds never plays a role in vocal learning. Certainly, copying of sounds can afford many opportunities for learning that would not otherwise be available, especially in young children. Nevertheless, it is important to recognize that not only does vocal learning occur in the absence of vocal imitation (reviewed by Schusterman, 2008), but vocal imitation can also occur without involving any new learning. These facts are clearly problematic for any taxonomy that defines vocal imitation as a kind of learning.

Synthesis

The two approaches to explaining vocal imitation described above—defining vocal imitation either as an outcome of instrumental conditioning or as a kind of learning (the vocal learning framework)—parallel more general frameworks for describing imitation. For instance, in evaluating strategies for defining imitation, Heyes (1996) identified three basic solutions: the essentialist solution, the positivist solution, and the realist solution. The essentialist solution is a definition-by-exclusion strategy in which researchers classify different imitation-like phenomena using specific criteria in an attempt to identify what is truly an imitative act. The vocal learning taxonomical framework is an example of this strategy. Limitations of this approach are that classifications are only as good as the demarcation criteria that are developed, and that defining vocal imitation by exclusionary criteria does not clarify what vocal imitation actually entails. The positivist solution involves selecting an operational definition for what will be called vocal imitation. The instrumental conditioning framework

qualifies as this type of strategy because it focuses less on differentiating vocally imitative acts from other acts, and more on identifying the conditions that lead to vocal reproductions. Finally, there is the realist solution, which focuses on explaining behavior in terms of theories about mental processes that yield testable hypotheses. Cognitive accounts of vocal imitation by adult humans exemplify this approach.

II. Vocal Imitation Is a Cognitive Process

One reason that vocal imitation often has been described as a learning mechanism is that comparative studies have focused on its role in vocal development (Baer, Peterson, & Sherman, 1967; Marler, 1970; McCowan & Reiss, 1997; Mowrer, 1952; Nottebohm & Liu, 2010; Subiaul, Anderson, Brandt, & Elkins, 2012; Tyack & Sayigh, 1997). In particular, there have been extensive comparisons between song learning by birds and speech learning by humans (Bolhuis et al., 2010; Doupe & Kuhl, 1999; Jarvis, 2004, 2013; Lipkind et al., 2013; Marler, 1970)[2]. From this perspective, vocal imitation provides a way for naïve youngsters to acquire the communicative abilities of mature adults. In fact, some researchers have argued that copying of sounds outside the natural repertoire may be a functionless evolutionary artifact (Garamszegi, Eens, Pavlova, Aviles, & Moller, 2007; Lachlan & Slater, 1999). Although vocal imitation abilities can be an important component of vocal development, the most versatile vocal imitators are adult humans (Amin, Marziliano, & German, 2012; Majewski & Staroniewicz, 2011; Revis, De Looze, & Giovanni, 2013). Furthermore, humans invariably achieve expertise in vocal imitation abilities well after learning to produce speech sounds. The most capable human vocal imitators perform copying feats that few adults can replicate. One could even argue that highly developed communication skills are a prerequisite for the highest levels of proficiency in vocal imitation, because professional imitators (e.g., impersonators, actors, singers) often receive detailed verbal feedback from instructors and peers over several years.

Viewing Vocal Imitation as a
Component of Auditory Cognition

Cognitive psychologists' conceptualization of vocal imitation by adult humans differs dramatically from that proposed by biologists and comparative psychologists for vocal imitation by non-humans. In particular, the emphasis in cognitive studies of vocal imitation is on how sounds and

---

2 Pepperberg (2005) noted, however, that in many cases it is unclear whether birdsong learning actually involves vocal imitation (see also Marler, 1997).

vocal acts are perceived, how links between percepts and actions contribute to performance, and how mental representations of events contribute to these processes. From this perspective, studies of vocal imitation in adults can be viewed as part of the field of *auditory cognition*, which focuses on understanding how mental representations and cognitive processes enable the understanding and use of sound.

In some respects, the cognitive approach to describing vocal imitation represents a return to Morgan's (1896) portrayal of imitation. Recall that Morgan divided imitation into two types: instinctive and voluntary. As an example of instinctive vocal imitation, he described a scene in which a chick comes across a dead bee and gives an alarm call, which leads a second nearby chick to give a similar alarm call. Today, the latter part of this scenario would be described as a case of vocal contagion. Morgan contrasted this kind of reflexive vocal matching with voluntary imitation, which he also refers to as conscious, intentional, or intelligent imitation. He noted that voluntary imitation is not independent of instinctive imitation, but rather builds on it (see also Romanes, 1884). Notably, frameworks that describe vocal imitation as either instrumental conditioning or vocal learning make no distinction between reflexive and voluntary imitation. This distinction is common in cognitive studies of human vocal imitation, however, and has recently also been revisited in discussions of motor imitation. For example, Heyes (2011) distinguishes between two "radically different" types of imitation: a complex, intentional type that individuals can use to acquire novel behaviors (voluntary imitation), and a simple, involuntary variety that involves duplicating familiar actions (referred to as automatic imitation). Cognitive psychologists have also drawn a distinction between overt imitative acts, which involve the observable, physical reproduction of sound, and covert imitation, which involves the unobservable, mental, or subvocal reproduction of sounds or actions (Pickering & Garrod, 2006; Wilson & Knoblich, 2005). These distinctions have important implications for understanding what vocal imitation is, and for identifying the cognitive processes that make vocal imitation possible.

Automatic Imitation Suggests Vocal Imitation Frequently Goes Unnoticed

Automatic vocal imitation has been studied extensively by speech researchers and has been observed at multiple levels of processing, including syntactic, prosodic, and lexical alignment in conversation (Garrod & Pickering, 2009; Gregory & Webster, 1996; Levelt & Kelter, 1982; Neumann & Strack, 2000; Pickering & Branigan, 1999; Shockley, Richardson, & Dale, 2009). Automatic imitation is modulated by social factors such as gender (Namy, Nygaard, & Sauerteig,

2002), personal closeness (Pardo, Gibbons, Suppes, & Krauss, 2012), attitude toward the interlocutor (Abrego-Collier, Grove, & Sonderegger, 2011), conversational role (Pardo, Jay, & Krauss, 2010), model attractiveness (Babel, 2012), and even sexual orientation (Yu et al., 2011). Talkers apparently imitate both visual and auditory components of observed speech (Legerstee, 1990; R. Miller, Sanchez, & Rosenblum, 2010). Automatic vocal imitation processes may occur relatively continuously without any awareness by the vocalizing individual (or others) that they are occurring.

One common way of generating automatic vocal imitation in the laboratory is to have talkers listen to and then intentionally repeat just-heard speech (a task called *shadowing*). Listeners can voluntarily replicate speech with a delay as short as 150 ms (Porter & Lubker, 1980). Shadowing could be viewed as a case of rapid vocal imitation, but is more often described as word repetition. Rapid production of just-heard words supports the notion that perceived sounds may be automatically converted into articulatory commands (Skoyles, 1998). When a talker produces shadowed words in ways that are more similar to the just-heard words than to his or her spontaneous speech, then this is viewed as evidence that the talker has automatically imitated features of the just-heard words (Fowler, Brown, Sabadini, & Weihing, 2003; Honorof, Weihing, & Fowler, 2011; Kappes, Baumgaertner, Peschke, & Ziegler, 2009; Mitterer & Ernestus, 2008; Nielsen, 2011; Shockley, Sabadini, & Fowler, 2004). Goldinger (1998) found that immediately shadowed words were more likely to be judged by external evaluators as matching the just-heard sound than versions produced after a four second delay. He also found that when talkers shadowed uncommon words, their reproductions were more likely to be judged as matching the just-heard sound than when they shadowed common words. Similar effects have been observed in tasks in which talkers replicated unique word features that were encountered up to a week previously (Goldinger & Azuma, 2004; Nielsen, 2011). These findings suggest that the effects of automatic vocal imitation mechanisms on speech production may persist for long periods. Further evidence that experienced sounds may involuntarily affect vocal production comes from the earworm phenomenon, wherein a person involuntarily mentally or overtly rehearses a catchy tune that was previously encountered (Beaman & Williams, 2010; Halpern & Bartlett, 2011; Williamson et al., 2012).

People voluntarily shadow words when they are instructed to repeat them in laboratory studies, but there are cases in which individuals involuntarily shadow recently heard sounds in their environment, referred to as *echolalia*. Echolalia is commonly seen in people with autism and is also associated with several other disorders (Fay, 1969;

Schuler, 1979; van Santen, Sproat, & Hill, 2013). It can involve either immediate or delayed reproduction of relatively complex sequences of speech sounds (Prizant & Rydell, 1984) or non-vocal sounds (Fay & Coleman, 1977; Filatova, Burdin, & Hoyt, 2010), and is often viewed as a contributing factor to dysfunctional language learning (Eigsti, de Marchena, Schuh, & Kelley, 2011). To date, detailed acoustic comparisons between heard speech and echolalic speech have not been performed, so the fidelity with which repeated sounds are copied is unclear.

Collectively, past studies of automatic vocal imitation demonstrate that humans sometimes reproduce features of previously experienced sounds without intending to do so and without being aware that they are copying heard features. Because automatic vocal imitation is often not apparent to the vocalizing individual and can occur after a significant delay, it may be more prevalent than is currently recognized. How automatic imitation relates to voluntary vocal imitation is a key question that researchers have grappled with for over a century.

Covert Imitation Suggests That Vocal Imitation May Enhance Perceptual Processing

Virtually all past discussions of vocal imitation assume that it is a process that primarily serves to enable an individual to produce certain sounds by reference to sounds previously heard. A recent alternative perspective is that imitative abilities may instead (or additionally) facilitate the prediction of future events (Grush, 2004; Hurley, 2008; Wilson & Knoblich, 2005). This perspective assumes that individuals are better able to perceive the actions of conspecifics if they can construct mental simulations of ongoing acts (including vocal acts) that occur in parallel with the perception of those acts. These mental simulations would be available to the individual perceiving the acts, but would not be evident in the observer's behavior.

Covert vocal imitation is described as an automatic process in which a sound is represented, at least in part, in terms of the motor acts necessary to re-create the sound. The suggestion is that vocal imitative processing is not a rare event (as suggested by frameworks that only consider production of novel sounds to be evidence of imitation), but is instead a routine component of auditory processing. Echolalia is often interpreted as evidence that auditory processing normally engages an imitative process that would naturally lead to overt imitative acts if not for being actively inhibited (Fay & Coleman, 1977; Grossi, Marcone, Cinquegrana, & Gallucci, 2012). From this perspective, what is rare is for an organism to produce overt actions that reveal these representational processes—overt vocal imitation then becomes analogous to "thinking out loud." Wilson

and Knoblich (2005) suggest that vocal imitation serves not to enable the acquisition of new sounds, but rather as a perceptual process that uses "implicit knowledge of one's own body mechanics as a mental model to track another person's actions in real time" (p. 463). The advantage of such processing is that a listener can potentially fill in missing or ambiguous information and infer the trajectory of likely actions in the near future. In section five, we consider in more detail how such mental simulations may specifically contribute to audiospatial perception by cetaceans.

Voluntary Imitation Suggests That Vocal Imitation Can Be Consciously Controlled

Piaget (1962) was one of the first psychologists to collect empirical evidence that automatic vocal imitation abilities in human infants may provide a foundation for the later development of voluntary vocal imitation abilities. He strongly argued that vocal imitation was not an evolutionarily specialized ability. In fact, Piaget starts his book on imitation by stating that, "Imitation does not depend on an instinctive or hereditary technique . . . . the child learns to imitate" (1962, p. 5). Piaget proposed six successive stages in the development of voluntary vocal imitation in children: (1) vocal contagion, (2) interactive copying of sounds, (3) systematic rehearsal of sounds in the repertoire, (4) exploratory copying of novel sounds, (5) increased flexibility at imitating novel events, and (6) deferred imitation. Studies of vocal development in parrots led Pepperberg (2005) to suggest that parrots progress through similar stages of imitative development. She described three levels of vocal imitation proficiency, starting with the involuntary copying of sounds, followed by intentional production of copied sounds, which in some cases develops into more sophisticated, creative sound production including the recombination of familiar segments into new sounds.

Relatively few researchers have theorized about the mechanisms or functions of vocal imitation in adult humans. Donald (1991) described *vocal mimesis* by adults as differing from vocal imitation in that it involves the invention of intentional representations as well as "the ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic" (p. 168). He noted that vocal reproduction can serve communicative purposes, but may also function simply to represent an event to oneself. In his framework, vocal mimesis allows for the self-cued recall of previously perceived sounds, as well as the control of how those sounds might be transformed during reproductions; vocal acts that were initially involuntary (e.g., laughing) can be explicitly recalled and used intentionally, for instance in reenactments of past episodes or when acting out a scene. Donald proposed that the cognitive basis of vocal mimesis
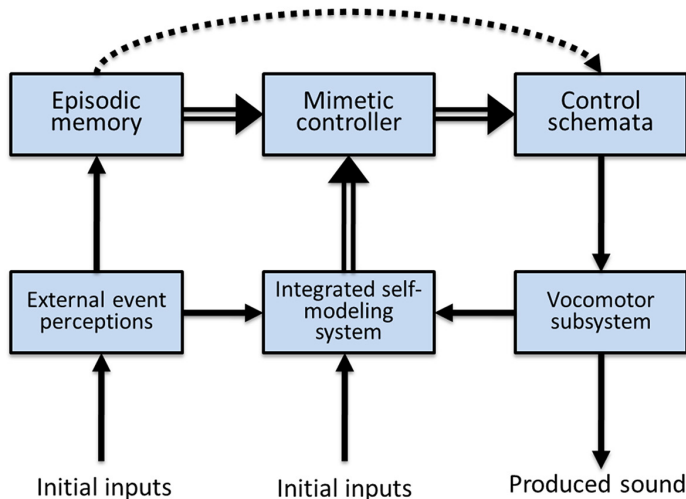
Figure 2. Donald's (1991) qualitative model of vocal mimesis in adult humans. The mimetic controller integrates episodic representations with outputs from self-representational systems to control how sounds are produced and to compare external events with self-produced actions.

involves a combination of episodic memory abilities and "an extended conscious map of the body and its patterns of action, in an objective event space; and that event space must be superordinate to the representation of both the self and the external world" (p. 189). He describes the main outputs of this system as consisting of self-representations and episodic memories. Thus, his proposed mimetic system (Figure 2) builds on and encompasses an episodic memory system, which some describe as one of the most advanced cognitive systems in adult humans (Tulving, 2002).

The idea that episodic memory representations play a key role in vocal reproduction has also been discussed in relation to speech shadowing tasks (Goldinger, 1998). Goldinger proposed that each word exposure generates a memory trace that resonates with previously encoded traces of the same word. When there are fewer past traces in memory (i.e., the word is uncommon), resonance with the current word presentation is weak. As a result, the unique vocal characteristics of the just-heard word are more likely to be retained in the mental representation that drives the shadowing production plan. Unlike traditional descriptions of episodic memory, which assume that such memories are consciously accessed, Goldinger's proposal implies that such memories may also automatically shape vocal production. Essentially, the idea is that memories of recent auditory episodes may continuously modulate how a listener vocalizes.

The capacity of adult humans to voluntarily copy sounds is best viewed as a cognitive skill that requires refined perceptual-motor control and planning abilities. *Cognitive skills* are abilities that an organism can improve through practice

or observational learning that involve judgments or processing beyond what is involved in performing perceptual-motor responses (Anderson, 1982; Mercado, 2008; Rosenbaum, Carlson, & Gilmore, 2001). Relevant cognitive processes that may contribute to an adult's vocal imitation skills include conscious maintenance and recall of past auditory or vocal episodes, selective attention to subcomponents of experienced and produced sounds, identification of specific goals of reproducing certain acoustic features, and awareness of possible benefits that can be attained through successful sound reproduction. From a cognitive perspective, an imitative vocal act is a memory-guided performance rather than a learning mechanism, and an individual's ability to flexibly perform such acts will depend strongly on how that individual mentally represents both sounds and sound producing actions (Roitblat, 1982; Roitblat & von Fersen, 1992).

The most impressive vocal imitation abilities of adult humans involve voluntary, highly experience-dependent skills that are more reminiscent of soccer skills than of learning mechanisms. Soccer players can all walk, run, and judge the consequences of their motor acts, but these abilities are insufficient to make someone a professional soccer player. Similarly, the ability to make sounds, recognize similarities between sounds, and remember sounds are all necessary for vocal imitation, but these abilities do not make a person a professional impersonator. It would not make sense to say that a toddler uses his soccer abilities to learn how to walk, and it may similarly be questionable to say that a toddler uses his vocal imitation abilities to learn how to talk. What the toddler does in both cases is learn how to flexibly control his or her actions based on past experiences. S(he) gradually learns to voluntarily run and kick in strategically advantageous ways and also gradually learns to voluntarily produce sound features based on memories of past percepts and actions.

Synthesis

Past attempts to understand the nature of vocal imitation reflect the ways in which this phenomenon has been used as an explanatory construct. Psychologists have often noted the important role that vocal imitation may play in language learning, and consequently have emphasized how the availability and guidance of adult speakers may contribute to learning when infants copy their examples. Biologists have also stressed how vocal imitation can facilitate the vocal learning of communicative signals. Consequently, it is perhaps only natural that researchers have traditionally described vocal imitation as a learning mechanism. In contrast, it seems less likely that an adult human shadowing speech in a laboratory, or humming a tune while exiting a concert, is doing so to learn how to speak or to hum.

Identifying when vocal imitation abilities are used provides hints about what those abilities may be for, but those hints may be misleading when only a subset of the relevant contexts are considered or when those abilities are difficult to observe. Without understanding the mechanisms that underlie sound imitation, and without any ability to monitor those mechanisms, it is simply not possible to definitively identify instances in which vocal imitation is occurring.

What then is vocal imitation? Clearly, different fields offer different ways of answering this question. Historically, animal learning researchers have described vocal imitation as the generalization of a conditioned response that is acquired through a supervised learning process. In this framework, acquisition of vocal imitation abilities (and consequent vocal communicative capacities) is subserved by general mechanisms of associative learning, rather than adaptively specialized vocal learning mechanisms. Animal behavior researchers, in contrast, have treated vocal imitation as a highly specialized adaptation that serves primarily to increase the flexibility with which animals can expand or customize their vocal repertoire. In this context, vocal imitation is the learning mechanism. Finally, cognitive psychologists construe vocal imitation as a consequence of multiple voluntary and involuntary representational processes. From the cognitive perspective, vocal imitation may help an organism learn, but this capacity can also be enlisted when no learning or vocalizing is occurring.

In the following, we use the term *vocal imitation* to refer to the *vocal reenactment of previously experienced auditory events*, essentially endorsing the framework developed by cognitive researchers studying vocal imitation in adult humans. Moreover, we claim that vocal imitation is a complex cognitive ability that involves coordinating action and perception. As such, vocal imitation can both be learned and in turn facilitate learning. The strength of this definition, and the cognitive approach more generally, is that it encompasses voluntary and automatic imitation, including covert imitation, and gives a clearer sense of the scope of cognitive processes that may contribute to vocal imitation abilities. A potential weakness of this definition is that it does not provide specific criteria for distinguishing imitative vocal acts from those that are non-imitative. As history has repeatedly shown, however, identifying such demarcation criteria is a formidable task, made all the more difficult by an incomplete understanding of the mechanisms underlying vocal imitation. Taxonomical distinctions may be useful for classifying different vocal phenomena, but it is less clear that they provide a viable framework for understanding what vocal imitation is or how it works. Instead, we focus on understanding how past experiences with various sounds enable some organisms to reproduce them. Because cognitive psychologists have studied vocal imitation most extensively in primates (primarily humans), we first consider the factors that determine when primates imitate sounds, as well as the features of sounds that primates are most likely to reproduce.

III. Sound Imitation by Human and Non-Human Primates

When considering the factors that constrain an individual's ability to imitate sounds (or likelihood of doing so), a key question is: what makes a sound more or less imitatible? The answer to this question may vary across species and even within and across individuals of the same species. Wilson (2001b) defined imitatible stimuli as those for which an individual's body can engage in an activity in which its configuration and movement can be mapped onto the configuration and movement of the stimulus, even if the mapping is not perfect and only applies to a limited set of properties of the stimulus. Most humans can easily imitate at least some speech sounds, but all other primates cannot. This has generally been interpreted as evidence that humans have unique capacities for imitating sounds. It remains possible, however, that sounds exist that at least some non-human primates might easily imitate, but that humans would find difficult or impossible to imitate. In the following, we suggest that a primate's ability to imitate a particular sound depends, at least in part, on how the individual represents the sound and sound producing actions.

Imitating Sounds Non-Vocally

If vocal imitation is defined as the vocal reenactment of previously heard events, then *sound imitation* can be viewed as a generalization of this ability that includes both vocal and non-vocal reenactments. Past emphasis on understanding how vocal imitation enables individuals to learn to produce novel vocalizations has distracted attention away from instances in which organisms use non-vocal motor acts to reproduce sounds. According to Wilson's (2001b) definition of imitatible stimuli, any sound-producing movements of an individual's body that can be mapped onto features of heard sounds can potentially make that sound imitatible. Thus, it is important to consider all available sound producing body movements when evaluating the imitatibility of a sound for a particular species.

There have been several anecdotal reports of animals non-vocally reproducing environmental sounds such as percussive knocking (e.g., Witchell, 1896). This phenomenon has only recently been studied scientifically, however. Moore (1992) reported that a parrot (*Psittacus erithacus*) reproduced knocking sounds by drumming its head on objects after repeatedly observing a person hitting on

a door. He later described this behavior as an instance of percussive mimicry, which he argued was a more sophisticated ability than vocal reproduction. Most reports of sound imitation by non-human primates involve non-vocal sound production (for rare exceptions, see Kojima, 2003; Masataka, 2003). Marshall, Wrangham, and Arcadi (1999) observed that chimpanzees (*Pan troglodytes*) exposed to a male that produced a Bronx cheer[3] as part of his pant-hoot call subsequently began using this sound in their own calls. A captive orangutan (*Pongo pygmaeus × Pongo abelii*) independently learned to whistle and was able to match the duration and number of whistles produced by a human model (Lameira et al., 2013; Wich et al., 2009; see Figure 3a). Though performed with the mouth, whistling is a non-vocal motor act requiring fine control of lip positions and airflow. Most recently, infant chimpanzees have been shown to adopt particular non-vocal sound production techniques (kisses, lip smacks, Bronx cheers, teeth clacking) as attention-getting signals based on the techniques modeled by their mothers (Taglialatela, Reamer, Schapiro, & Hopkins, 2012). Chimpanzees also can be trained to produce such non-vocal sounds, suggesting that their ability to voluntarily generate novel sounds is more flexible than previously thought (Hopkins, Taglialatela, & Leavens, 2007; Russell, Hopkins, & Taglialatela, 2012).

The non-vocal sound imitation abilities of humans are often taken for granted in music education (Drake, 1993; Drake & Palmer, 2000; Palmer & Drake, 1997). For instance, a teacher may ask students to clap the rhythm of a song that they are learning to sing, or ask them to copy a demonstrated percussive pattern on various instruments. Conversely, most music students take ear-training classes that involve having to produce visually presented musical intervals vocally (called "sight-singing"), with the assumption that this ability will facilitate non-vocal reproduction of music. A musician that reproduces the melodic sequence produced by a singing bird or fellow musician when she plucks strings, presses piano keys, or uses air to make a reed vibrate, is also imitating the sounds non-vocally (Clarke, 1993; Clarke & Baker-Short, 1987). Many musicians learn to play songs "by ear," which involves transforming heard sounds into the motor acts required to reproduce them (McPherson & Gabrielsson, 2002; Woody & Lehmann, 2010). Musicians and non-musicians can readily imitate the intonation patterns of sentences by moving a stylus on a tablet (d'Alessandro, Rilliard, & Le Beux, 2011). It is not clear anecdotally, either among human or non-human primates, that there is anything special about non-vocal reproduction

---

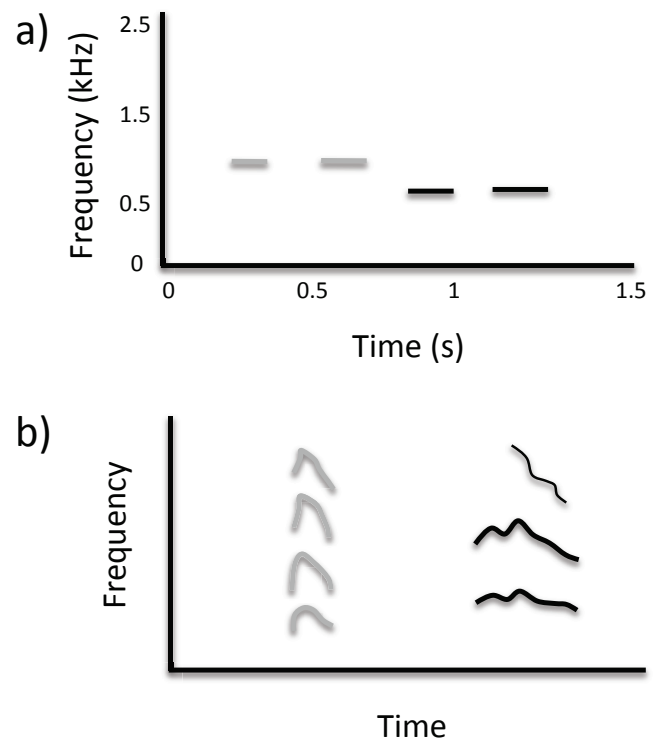[3] an unvoiced sound produced by placing the tongue between the lips and blowing

Figure 3. (a) Non-vocal sound imitation by an orangutan (adapted from Wich et al., 2009; Figure 2). Gray lines show spectrographic contours of whistles produced by a human, and black lines show the contours of subsequent whistles produced by the orangutan in which the number, timing, and duration of sounds are similar to features present in the target sequence. (b) Spontaneous vocal production by an infant chimpanzee (black lines show spectrographic contour and harmonics) with acoustic features similar to those of a preceding environmental sound (gray lines), indicative of vocal imitation (adapted from Kojima, 2003; Figure 9-2).

of sounds relative to vocal imitation. The individuals appear to be reproducing sounds based on past experiences, regardless of whether the reenactment is produced through the voice or through some other means. In fact, the perceptual and cognitive demands appear to be comparable: the individual perceives a sound and then uses that sound as a guide for controlling motor acts that generate a similar event.

It is possible, however, that vocal and non-vocal sound imitation involve qualitatively different mechanisms. For instance, Moore (2004) argues that the parrot's capacity for copying sounds percussively requires adaptations beyond those necessary for vocal imitation. In human studies, some have suggested that processing of different auditory events (e.g., melodies versus speech) may involve separate underlying mechanisms (Peretz & Coltheart, 2003; Zatorre & Baum, 2012; Zatorre, Belin, & Penhune, 2002), whereas others argue that there may be significant

overlap (Mantell & Pfordresher, 2013; Patel, 2003; C. Price & Griffiths, 2005). Evidence supporting the view that vocal and non-vocal sound imitation can involve separate mechanisms was recently reported by Hutchins and Peretz (2012). In their study, participants who were classified as either accurate or poor-pitch singers matched pitch either vocally or manually by using a slider. The slider was used so that participants could continuously control pitch, as is the case for vocal pitch control, thus somewhat equating demands of pitch control across distinct effector systems. They found that pitch-matching errors in poor-pitch singers were voice specific. In other words, poor-pitch singers successfully matched pitch using the slider, but not using their voice. These results suggest that an individual's ability to reproduce a pitch depends on the specific movements and associated feedback involved in matching the pitch.

For primates, sounds are imitatible when they are encoded in such a way that the stored representation of that sound enables the listener to voluntarily generate motor acts that produce phenomenological features present within the originally experienced sound. Note that by this criterion, any sound that a human hears is potentially imitatible, because the listener should be able to at least approximate the duration of the heard sound through some sound producing action. It is less clear which sounds would qualify as imitatible for other primates. Based on the currently available evidence, non-vocal sounds produced with the mouth seem to be relatively easy for chimpanzees and orangutans to reproduce, whereas vocal sounds are relatively easy for humans to imitate. Given that some sounds, such as those produced by conspecifics, will be easier to reproduce than other sounds, findings regarding which sounds (or features of sounds) are most imitatible can provide important clues about the factors that constrain imitation capacities within and across species.

Variations in the Imitatibility of Sounds

If sound imitation depends on adaptively specialized auditory-motor processing, then the sound features that should be easiest for an organism to imitate should be those present within functional vocalizations produced by conspecifics. Recent studies of humans provide some support for the hypothesis that vocal imitation is facilitated for natural vocalizations. For instance, matching of pitch is more accurate with a human voice timbre than a synthetic vocal timbre (Lévêque, Giovanni, & Schön, 2012; R. Moore, Estis, Gordon-Hickey, & Watts, 2008) or with a complex tone (Hutchins & Peretz, 2012; Watts & Hall, 2008). Adults also match pitch better when the vocal range of the target is closer to their range, as when female imitators match a female voice (H. E. Price, 2000). Similarly, children match pitch better when matching a child's

voice, and are better at matching pitch for a female than a male adult voice, given the greater similarity of female voice formants and pitch to a child's voice (Green, 1990).

Mantell and Pfordresher (2013) recently explored differences in the vocal imitation of pitch within two cognitive domains: music (song) and language (speech). We summarize the results of this study here as a paradigmatic example of how vocal imitation can be influenced by stimulus structure, and of how the fidelity of imitations can be quantitatively assessed. According to the modular model of audition proposed by Peretz and Coltheart (2003), pitch processing occurs in domain-specific, information-encapsulated modules (Fodor, 1983) separate from speech processing. In a direct test of this framework, Mantell and Pfordresher compared the accuracy with which people intentionally imitated the pitch-time contents of spoken sentences and sung melodies. They created speech and song stimuli that matched in word content, pitch contour (the pattern of rising and falling pitch), pitch range, and syllable/note timing. The difference between the speech and song targets was that each note of the sung targets conformed to diatonic, musical tonal rules. Mantell and Pfordresher reasoned that if the pitch processing system underlying vocal imitation was truly modular, phonetic information should not influence imitative performance. Thus, the critical experimental factor was the presence or absence of phonetic information in the target sequences. They created wordless versions of all of the speech and song stimuli by synthesizing the pitch-time contents of each of the worded sequences. The wordless

Figure 4. Pitch contours (shown as black dots) extracted from an adult human's vocalizations when the individual was instructed to imitate a target vocal sequence compared with spectral and temporal features of the target sequence (gray lines).
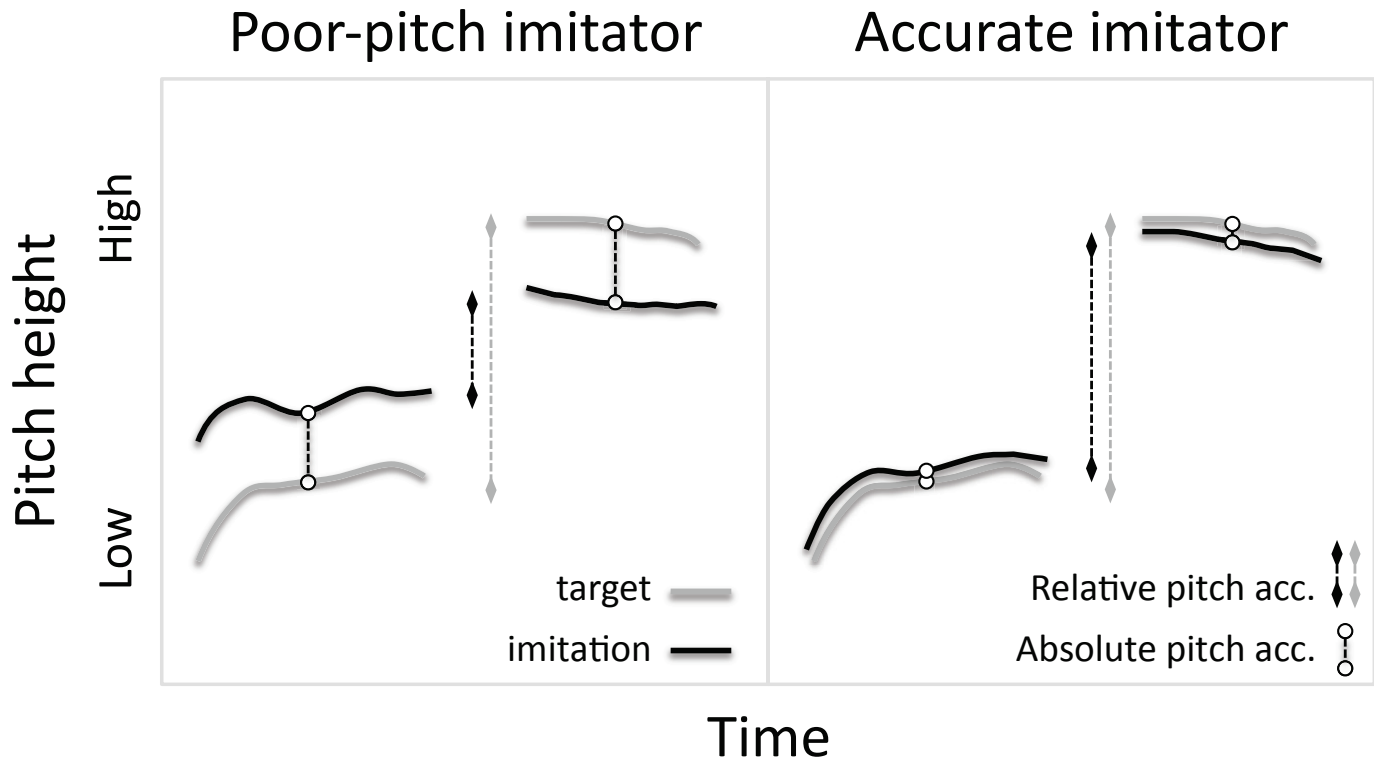
Poor-pitch imitator   Accurate imitator
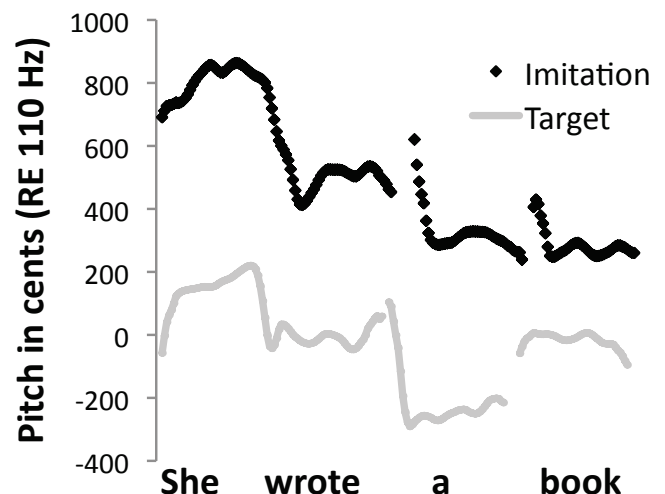
Figure 5. Poor-pitch imitators (left) produce vocalizations that do not match the target sounds in absolute or relative pitch, whereas typical adult humans (right) match both spectral features.

versions sounded like hummed versions of the sentences and songs, but they lacked all acoustic-phonetic identification cues. Imitation accuracy was gauged by directly comparing the target sequence with a temporally aligned imitative production (Figure 4) and by calculating two different quantitative measures of similarity (Figure 5). The first measure, mean absolute error, assessed the accuracy with which each imitative production matched the pitch content of the target. The second measure, pitch correlation, scrutinized the accuracy with which each imitative production tracked the relative (rising and falling) pitch-time contour of the target.

The critical finding of this study was that the presence of phonetic information in both the target and the imitative production reliably improved pitch accuracy. Thus, subjects imitated worded speech and song sequences more accurately than they imitated wordless speech and song sequences, despite the fact that the wordless versions were acoustically simpler (e.g., they lacked complex acoustic-phonetic spectral information). This finding is contrary to predictions afforded by a modular framework of music and speech processing, because if musical pitch processors are encapsulated to speech, then pitch processing should occur independently and unhindered (or not facilitated) by any parallel phonological processes. It also contradicts the proposal that imitation of spectrotemporal contours is inherently more difficult than

imitation of other acoustic features (Janik & Slater, 2000). Mantell and Pfordresher further found that participants varied in their accuracy at imitating absolute and relative features of target sequences (see also Dalla Bella, Giguere, & Peretz, 2007; Pfordresher & Brown, 2007). Specifically, participants imitated the absolute pitch within songs better than the absolute pitch in sentences, but imitated the relative pitch-time contours of speech and song equally well.

In Mantell and Pfordresher's (2013) study, participants imitated recordings of vocalizations and also synthesized versions of these recordings, making it possible to examine whether they adjusted the resonant properties of their vocal tract in order to imitate the timbre of targets. The synthesized recordings featured a timbre that resembled a human voice, but that differed considerably from the timbre of vocal recordings. Analyses of the long-term average spectra during imitations (Figure 6) suggested that participants adjusted their own vocal resonances in order to imitate the timbre of each target, even though this was not necessary according to instructions, which simply focused on the imitation of pitch content. As illustrated in Figure 4, participants also naturally matched the temporal structure of heard sequences, which was also not specifically requested in the instructions. Thus, when humans voluntarily imitate speech or song sequences, they spontaneously imitate multiple acoustic
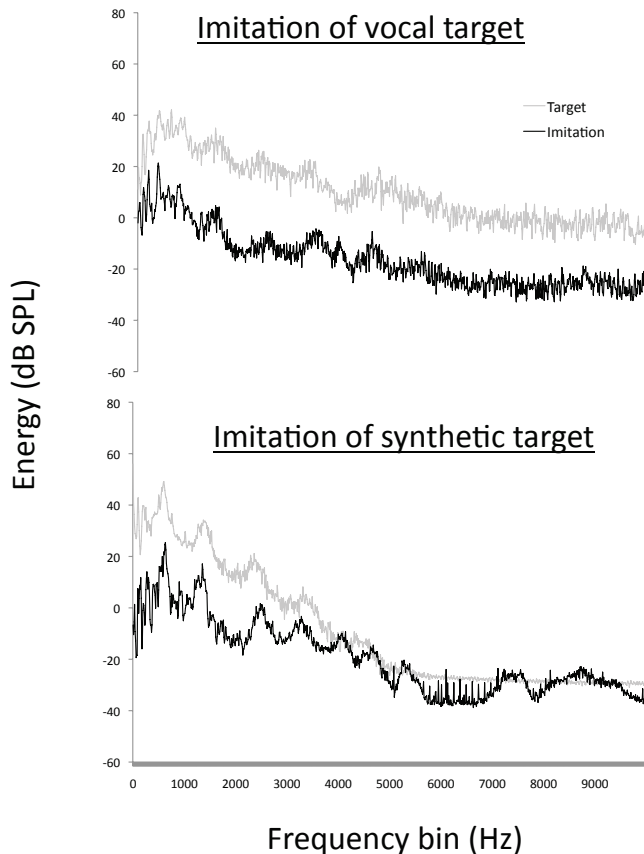
Figure 6. Long-term average spectra showing that adult humans spontaneously match the timbre of target sound sequences when targets are either natural or synthetic.

features of the sequences. Interestingly, when an orangutan imitated whistle sequences produced by a human (Wich et al., 2009), it also spontaneously matched the duration and temporal spacing of target sequences (Figure 3a), suggesting that this propensity is not limited to human imitators.

What Makes a Sound Imitatible?

As noted above, a basic question surrounding the imitatibility of sounds concerns whether, or to what degree, organisms have evolved dedicated systems that are specialized for imitating certain sound features. The imitatibility of sounds is not simply based on whether the acoustic properties of individual sounds resemble those of natural vocalizations. People are able to vocally reproduce melodies presented on a piano as well as those that are sung, and infant chimpanzees sometimes imitate environmental sounds (Figure 3b, Kojima, 2003). The complexity of a target sequence can strongly limit its imitatibility. At a cognitive level, different kinds of target sequences represent different auditory

domains and may, according to some theories, be processed by different cognitive modules. Take for instance the difference between a sung melody on the syllable "la" versus a spoken sentence. Both are auditory sequences, but each is complex in its own way. Because the former sequence is heard as "musical," it may be processed differently from the latter sequence. Such putative separation across domains may therefore influence imitatibility and, consequently, many human studies focus on the structural complexity of rhythmic, melodic, and phonic combinations rather than on the relative difficulty of producing individual sounds.

An important ancillary consideration when evaluating the imitatibility of sounds is the flexibility of vocal production by the imitator. Obviously, an individual who can imitate a wide range of inputs must be able to engage in flexible vocal motor control. Flexibility in pitch range increases dramatically during childhood, and thus may play a large role in the development of pitch matching abilities in singing (Welch, 1979). Similarly, poor-pitch singers, who exhibit a general deficit of vocal imitation, also exhibit an apparent lack of flexibility in vocal imitation (Pfordresher & Brown, 2007). Poor-pitch singers also show a larger advantage for matching pitch from recordings of their own voice, in contrast to matching the vocal pitch of other singers, than do more accurate singers (R. Moore et al., 2008; Pfordresher & Mantell, 2014). Finally, when transferring from the imitation of one sequence to another, poor-pitch singers show a greater tendency to perseverate the previously imitated pitch pattern than do more accurate singers (Wisniewski, Mantell, & Pfordresher, 2013). Interestingly, this apparent lack of flexibility in poor-pitch singers does not appear to be based on vocal motor control in that poor-pitch singers exhibit similar pitch range and ability to control a sustained pitch as accurate singers (Pfordresher & Brown, 2007; Pfordresher & Mantell, 2009). Instead, their inflexibility seems to result from dysfunctional vocal imitation abilities.

Even when considering only the performance of adult humans, there is no fixed scale of most-to-least imitatible sounds or sound features. Nevertheless, it may be possible to generate a gross scale of different properties associated with sounds being more or less imitatible. For instance, sound features that are imperceptible or sounds (and sequences) with complex, aperiodic, novel acoustic structures are typically more difficult to imitate, whereas sounds that are routinely self-generated tend to be the easiest to reproduce. Interestingly, this scale is the inverse of the criteria that biologists have developed for identifying instances of vocal imitation. Specifically, production of highly imitatible sounds is generally considered to be the least compelling behavioral evidence of vocal imitation, whereas production of novel complex sounds (which are often less imitatible)

is currently considered to be the most compelling evidence. Consequently, the sounds that an individual is most likely to be proficient at imitating are also the sounds that scientists are least likely to consider relevant to studies of vocal imitation. In fact, in the taxonomy of vocal learning abilities proposed by Janik and Slater (2000), some sounds are inherently impossible to imitate; by their definitional criteria, an individual cannot imitate any sound that is already within the individual's vocal repertoire. This constraint arises from the fact that they view vocal imitation as a learning mechanism. If vocal imitation is viewed as vocal reenactment, however, then individuals can potentially imitate any sound. This includes their own vocalizations, a process referred to as *self-imitation* (Pfordresher & Mantell, 2014; Repp & Williams, 1987; Vallabha & Tuller, 2004).

Studies of intentional vocal imitation in humans are beginning to shed new light on how sound imitatibility varies within and across individuals. They have yet to reveal, however, why sound imitatibility varies. If a person is particularly good at imitating a family member's voice that is similar to his or her own, is this because the person possesses an adaptively specialized module that is tuned to the specific features of sounds produced by relatives? Is it because shared genetics have led to similar vocal organs? Or, is it because the person aspires to be like that family member and has practiced copying particular mannerisms of their role model's vocal style over many years? To a large extent, the imitatibility of a sound depends on what resources the listener brings to bear for perceiving, encoding, and producing sounds. A clearer understanding of the physical and mental mechanisms relevant to increasing the imitatibility of sounds can be gained by examining those individuals who have reached the highest levels of performance—professional imitators.

Expertise in Sound Imitation

If, as we claim, voluntary imitation of sounds is a cognitive skill, then it should be possible to improve imitation abilities with training. However, if sound imitation is more of an innate capacity, then individual variations in ability should be less dependent on experience. Earlier claims that vocal imitation involves feedback-based error correction (Heyes, 1996; N. E. Miller & Dollard, 1941) predict that the fidelity with which particular sounds are imitated should increase incrementally as the number of comparisons between produced vocalizations and remembered targets increases. However, studies of the vocal imitation of pitch in singing have not shown any improvements across repeated trials in which participants attempted either to match the same pitch vocally (Hutchins & Peretz, 2012), or to repeatedly imitate the same spoken or sung sequence

(Wisniewski et al., 2013). Likewise, efforts to enhance pitch imitation accuracy by having participants sing along with the correct sequence (auditory augmented feedback) have yielded mixed results and may even degrade the performance of poor-pitch singers (Hutchins, Zarate, Zatorre, & Peretz, 2010; Pfordresher & Brown, 2007; Wang, Yan, & Ng, 2012; Wise & Sloboda, 2008). It is clear anecdotally that individuals can improve their vocal imitation abilities through instruction and practice. However, simply relying on error correction based on auditory feedback may not suffice. More successful methods of augmented feedback involve showing the singer a graphical display of the imitated and target pitches as on-screen icons, with changes to sung pitch influencing the spatial proximity of these displays (Hoppe, Sadakate, & Desain, 2006).

Anecdotally, evidence that learning experiences can strongly determine sound imitation abilities comes from the performances of professional musicians, who often train and practice for decades to achieve the control necessary to produce particular sound qualities (e.g., features such as vibrato or breathiness). Often, musical training focuses on teaching students how to produce higher quality sound sequences. This generally means the student must learn to reproduce the features of sounds commonly produced by more proficient musicians. The fact that many professional musicians spend several hours a day performing exercises to maintain and enhance their musical skills attests to the important contributions of practice to their ability to flexibly and accurately reproduce sounds in a prescribed way.

A second domain in which imitative skills appear to be refined through practice is the learning or copying of non-native languages. Much of the difficulty in learning a new language relates to learning to produce speech sounds to match some pre-established standard. The ability to imitate foreign languages varies considerably across individuals (Golestani & Zatorre, 2009; Reiterer et al., 2011), and is predicted by levels of articulatory flexibility and working memory capacity (Reiterer, Singh, & Winkler, 2012). Professional actors may learn to reproduce a wide range of dialects or even foreign languages that they do not speak when performing dialogue. What exactly are second language learners or professional actors learning in these situations? In part, they seem to be learning which features of speech sounds and vocal gestures they need to copy. Importantly, speakers do not need to learn the necessary adjustments for each word within a language, but can immediately apply what they have learned to many novel words and sentences. In some cases, subtle distinctions in speech sounds may be extremely difficult for a non-native speaker to imitate (Ingvalson, Holt, & McClelland, 2012; Lim & Holt, 2011; Reiterer et al., 2011), again attesting

to the important role that experience plays in an adult's ability to vocally imitate, even when the sounds being imitated are the naturally occurring speech of other humans.

Learning language or musical skills might depend more on developing expertise in particular perceptual-motor acts or on gaining knowledge about symbols and rules than on improving sound imitation abilities. Some entertainers have more explicitly developed expertise in sound imitation, however, including professional impersonators, tribute artists, and vocalists described as beatboxers. These performers all specialize in reproducing speech or musical sound sequences. For example, beatboxers interleave imitations of both percussive and vocal elements of electronically or acoustically generated sound sequences, often using novel modes of sound production to capture key features of the musical sequences being imitated. Like other musicians, these expert sound imitators gain their unique skills through extensive directed practice and performance.

Recently, researchers have found that professional speech impersonators match the general pitch of the fundamental, temporal variations in the fundamental, speaking rate, prosody, formant structure, and the timbre of model speakers (Amin et al., 2012; Eriksson, 2010; Eriksson & Wretling, 1997; Majewski & Staroniewicz, 2011; Revis et al., 2013; Zetterholm, 2006). Impersonators match the timing of speech sounds at the sentence or prosodic level rather than at the word level (Eriksson & Wretling, 1997; Liberman & Mattingly, 1985; Revis et al., 2013), and vary considerably in their ability to match particular features. Compared to amateurs, professional impersonators are more aware of differences between their vocalizations and those of a target, and are better able to emphasize features that are likely to be salient to listeners (Revis et al., 2013). Interestingly, expert impersonators, like caricaturists, often exaggerate features of copied sounds such that imitations judged to be most accurate by listeners generally do not exactly match the acoustic features of the model (Majewski & Staroniewicz, 2011; Zetterholm, 2006). In fact, when amateur impersonators imitated models, the acoustic properties of their imitations more closely matched the speech of models, but listeners nevertheless judged these attempts as worse copies of the models than those produced by professionals (Majewski & Staroniewicz, 2011). These acoustic experiments show that expert vocal imitators copy and adjust sounds along multiple acoustic dimensions in parallel, and can do so even when producing novel speech sequences that incorporate speech sounds/words that differ from those of the model.

Collectively, evidence from expert imitators suggests that the enhanced sound imitation abilities of adult humans reflect a protracted learning process that can extend over decades. This raises questions about whether differences in imitation abilities across species might reflect differences in training histories rather than (or in addition to) differences in adaptive specializations. A related possibility is that constraints on sound imitation in non-humans may reflect differences in cognitive plasticity across species (Mercado, 2008), such that even with comparable training histories and the same cognitive mechanisms available, some species may be better able to acquire the cognitive skills necessary for flexible sound imitation. Neither of these accounts requires that humans possess any adaptively specialized "extra parts" to account for cross-species differences in vocal imitation abilities.

Synthesis

Past assessments of the sound imitation abilities of non-human primates have been unequivocally dismissive. For instance, Hauser (2009, p. 304) states that, "monkeys and apes . . . show no evidence for vocal imitation. There is no capacity (and it has been fifty years of intensive looking by primatologists), absolutely no evidence for vocal imitation." Although there is evidence that adult non-human primates may modify their vocalizations so that they are more similar to those of other individuals within a group (reviewed by Egnor & Hauser, 2004; Owren, Amoss, & Rendall, 2011), referred to as *vocal convergence* or *vocal matching*, it is unclear whether such convergence is the result of learning or genetics. Although non-human primates do not vocally imitate sounds to the same extent as humans, they do have some capacity to represent a subset of sounds in ways that enable them to non-vocally imitate those sounds. If flexible sound imitation abilities are cognitive skills learned through practice, as the evidence from adult humans suggests, then a non-human primate might need significant pedagogical guidance over many years before flexible sound imitating abilities are evident (see also Pepperberg, 1986). It seems clear, nevertheless, that non-human primates only rarely overtly imitate sounds in naturalistic contexts. Consequently, studies of imitation in monkeys and apes may be less informative than studies of other mammals that more regularly imitate sounds. Few mammals other than humans have shown the ability to voluntarily imitate sounds, which has led some researchers to suggest that vocal imitation requires unique, human-specific neural and cognitive processing mechanisms. By examining the situations faced by those rare mammalian species that are known to naturally voluntarily imitate sounds, one can potentially gain insights into the representational demands that might lead to the kinds of internal processes that would provide an organism with flexible sound imitation abilities. Identifying similarities and differences between the imitation abilities of humans and non-humans can thus

provide important clues about the nature of the mechanisms that determine imitative proficiency and proclivity.

IV. Sound Imitation by Whales and Dolphins

The only mammalian order that includes multiple species with the apparent ability to flexibly imitate sounds is cetaceans. In the following section, we review the evidence for sound imitation abilities in cetaceans in detail, considering not only the strengths and weaknesses of this evidence, but also how it compares to findings from human research. Cetaceans provide a particularly important test bed for examining the origins of imitative abilities as well as the mechanisms that underlie such abilities, because although they have diverged in many ways from terrestrial mammals, they seem to possess cognitive capacities that are similar in certain respects to those of humans (Herman, 1980; Marino et al., 2007; Mercado & DeLong, 2010). For instance, bottlenose dolphins are the only mammals other than humans that have demonstrated the ability to voluntarily imitate both seen and heard actions (Herman, 1980, 2002; Kuczaj & Yeater, 2006; Yeater & Kuczaj, 2010). Humpback whales are the only non-human mammals that continuously and collectively restructure their vocal repertoire throughout their adult lives (Guinee, Chu, & Dorsey, 1983). Given that researchers are severely limited in their ability to observe and conduct experiments with cetaceans, the prevalence of observations indicative of cetacean sound imitation abilities is noteworthy. The following subsections focus on the sound producing abilities of belugas (*Delphinapterus leucas*), orcas (*Orcinus orca*), humpback whales (*Megaptera novaeangliae*), and bottlenose dolphins, four cetacean species often described as vocal imitators. Evidence suggestive of vocal learning and imitative abilities has been reported for other cetacean species (DeRuiter et al., 2013; May-Collado, 2010; Rendell & Whitehead, 2003), and thus the four species emphasized here are best viewed as a sample of convenience.

Flexible Sound Production Mechanisms
May Enhance Imitative Capacities

Vocal flexibility is a key aspect of vocal imitation and may be a prerequisite for vocal imitation abilities (Arriaga & Jarvis, 2013; Deacon, 1997; Fitch, 2010; Mowrer, 1960). Like most mammals, cetaceans can produce sounds using both internal organs and other body parts, referred to as vocalizations/phonations and percussive sounds respectively. Some researchers have questioned using the term *vocalization* to describe cetacean sounds because, unlike terrestrial mammals, most cetaceans do not appear to use vocal folds to produce sounds (Cranford et al., 2011; however, see

Reidenberg & Laitman, 1988, 2007). For those cetaceans that produce sounds nasally rather than vocally (which includes belugas, orcas, and dolphins), it would be more accurate to say that they possess nasal imitation abilities. As noted above, the term *sound imitation* avoids such complications because it does not specify how sounds are reproduced.

Cetacean vocalizations have traditionally been classified first by suborder (i.e., baleen whales vs. toothed whales), and then either by the acoustic features of the sounds perceived by the investigator, or in terms of proposed functional classes. The vocalizations of toothed whales have been classified into three aurally defined categories: clicks, whistles, and burst-pulse sounds. Clicks are often associated with echolocation, whereas whistles and burst-pulse sounds are often associated with communication (Herman & Tavolga, 1980; Janik, 2009a). Baleen whale vocalizations have often been described as being very different from toothed whale vocalizations and as much more difficult to classify (Edds-Walton, 1997). Distinctions have been drawn between calls and songs (Clark, 1990), and between different kinds of calls (e.g., moans, cries, grunts, and pulse trains). Edds-Walton (1997) categorized baleen whale sounds into three functional/contextual categories: contact calls, winter (breeding) vocalizations, and social sounds.

Popper and Edds-Walton (1997) suggested that the vocalizations of both toothed and baleen whales could be collectively classified into three discrete categories based on their acoustic features: tonal or narrow-band whistles or moans, pulsed sounds, and broadband clicks. However, other analyses suggest that these three subjective categories represent points along a continuum of pulsed vocalizations, with clicks corresponding to low-rate pulse trains, pulsed sounds to medium-rate pulse trains, and "whistles" to high-rate pulse trains (Killebrew, Mercado, Herman, & Pack, 2001; Mercado, Schneider, Pack, & Herman, 2010; Murray, Mercado, & Roitblat, 1998). In this context, the term *whistle* is a misnomer, because the mechanism of sound production is the same as that of clicks and pulsed sounds, namely vibrating membranes. This interpretation has recently been experimentally confirmed in bottlenose dolphins (Madsen, Jensen, Carder, & Ridgway, 2012). Observational evidence shows that cetaceans can continuously modulate sounds along this pulse-rate continuum, much as professional human singers do when producing pitches across a wide range (Mercado et al., 2010; Murray et al., 1998). In other words, the vocal repertoire of cetaceans is graded rather than discrete, and vocal control in cetaceans is generally comparable to that of human singers; click trains are analogous to vocal fry, pulsed sounds share features of sung vowels, and whistles are comparable to the sounds a soprano might produce when singing in the whistle register.

A key difference between how human singers typically vocalize and the ways that cetaceans vocalize is that some cetacean species can control two independently vibrating sources simultaneously (Cranford et al., 2011). In this way, the vocal flexibility of sound production is greatly increased for cetaceans relative to other mammals, and is more comparable to the dual syringeal production mechanisms used by many singing birds. Much less attention has been given to studies of percussive sounds made by cetaceans (e.g., rhythmic tail slapping), so little is known about how flexibly cetaceans might use these sound producing modes. In terms of vocal range, cetaceans as a group are unmatched among mammals. A humpback whale can produce sounds lower than any human male singer as well as sounds higher than the highest pitches sung by professional sopranos (Mercado et al., 2010). Dolphins can also produce a wide variety of tonal sounds as well as ultrasonic clicks (Au, 1993). The various species described below differ in their specific vocal skills and repertoires, but all show greater flexibility in sound production than any mammal other than humans. Thus, unlike non-human primates, there is no question about whether cetaceans have the dexterity necessary to imitate many acoustic features.

Speech-Like Sound Production by Belugas

The repertoire of vocal sounds produced by belugas has been evaluated both in captive animals (Vergara & Barrett-Lennard, 2008) and wild populations (Chmelnitsky & Ferguson, 2012), and is historically considered to be one of the most varied of all cetaceans (Fish & Mowbray, 1962; Schevill & Lawrence, 1949). Like most toothed whales, they produce a wide range of pulsed sounds, many of which have been described as whistles or pulsed calls. In contrast to many other toothed whales, the graded structure of beluga sounds has been consistently noted in past studies (Chmelnitsky & Ferguson, 2012; Karlsen, Bisther, Lydersen, Haug, & Kovacs, 2002; Sjare & Smith, 1986). Belugas appear to be able to produce two independent calls simultaneously (Chmelnitsky & Ferguson, 2012), consistent with reports from other highly vocal toothed whales. Like most cetaceans, belugas are thought to vocalize primarily to echolocate or to socially communicate. However, assessments of the functionality of sounds (other than click trains) have been limited mainly to observational studies in which different sound types were correlated with different social contexts (Panova, Belikov, Agafonov, & Bel'kovich, 2012; Sjare & Smith, 1986).

The social structure within groups of belugas appears to be fluid. They sometimes form large groups and there are indications that their sound repertoire varies with context and number of individuals present (Panova et al., 2012).

There are no reports of belugas imitating sounds in the wild, but such behavior would be virtually impossible to detect. It is also unclear how easily sound imitation by belugas in captivity would be to identify. Nevertheless, there are at least two published reports of captive belugas producing speech-like sounds without explicit training. The first was a report of an adult male that was heard to produce his name: Logosi (Eaton, 1979). This beluga was described as being particularly interested in human visitors, spending much of his time near viewing windows. He was also described as producing sounds that resembled the "sound of human voices heard underwater." Some listeners described these sounds as resembling Russian, Chinese, or garbled voices. A more recent report (Ridgway, Carder, Jeffries, & Todd, 2012) describes recordings of a second beluga spontaneously producing sounds that were "as if two people were conversing in the distance, just out of range for our understanding." The temporal patterning of sound production was also found to be comparable to speech. Trainers were able to teach the beluga to "speak" on cue, so that the sound production mechanisms used could be examined more closely. When the beluga produced speech-like sounds, atypical modes of sound production were observed in which the beluga sequentially inflated two vocal sacs. More naturalistic evidence of sound imitation during development was reported for a captive beluga calf that appeared to adopt new modes of call production after being exposed to the novel calls of an adult male that was introduced into his environment (Vergara & Barrett-Lennard, 2008). Researchers have speculated that belugas may copy the sounds of conspecifics to facilitate individual and group recognition or possibly to maintain social bonds (Janik & Slater, 1997).

One interesting feature of speech-like sound production by belugas is that humans speak in air, but the beluga's auditory system is adapted for receiving sounds underwater. Consequently, it is difficult to know whether differences between speech-like sounds produced by belugas and those produced by humans reflect limitations in their ability to reproduce sounds, or correspond to the effects of distortion caused either by impedance mismatches at the air–water interface, or because the beluga heard the speech with its head out of the water. In other words, a beluga might be accurately replicating the sounds that it experienced and still sound like it was producing distorted or garbled speech. This ambiguity highlights the fact that the similarity of two sounds is observer dependent; two sounds that are "different" to one observer (or species) might be "the same" to another, or vice versa.

Orca Sound Matching:
Imitation of Familiar Vocalizations?

Orcas, commonly referred to as killer whales, are the largest species of dolphin. Their vocal repertoire is similar in many respects to that of belugas, except that they produce relatively more intermediate pulse-rate calls than the higher pulse-rate "whistles" typical of belugas. Orcas also have been recorded producing two types of sound simultaneously (referred to as biphonic calls) more consistently than have belugas (Filatova et al., 2012; P. J. O. Miller, Shapiro, Tyack, & Solow, 2004). Much of the interest in orca vocalizations comes not from any particularly unusual features of their calls or their usage, but from the fact that stable social groups of orcas use a shared repertoire of sounds that is so consistent that recordings of particular sounds can be used to identify particular families of orcas (Deecke, 1998; Filatova et al., 2010; Ford, 1991; Weib, Symonds, Spong, & Ladich, 2011). The predictability with which groups of orcas use certain sets of sounds with recognizable acoustic features has led many researchers to conclude that orcas within each group use a discrete library of 7–17 calls that is adopted by convention (Ford, 1991; Kremers, Lemasson, Almunia, & Wanker, 2012; Rendell & Whitehead, 2001; Strager, 1995). Field observations indicate that orcas use their sounds differentially depending on the social context (Ford, 1989; Hoelzel & Osborne, 1986; Thomsen, Franck, & Ford, 2002). Some of the call types appear to be shared across social groups, and overlap in repertoires has been used as an index of the relationships between distinct groups (Riesch, Ford, & Thomsen, 2006; Yurk, Barrett-Lennard, Ford, & Matkin, 2002). Longitudinal analyses of call variations within particular groups suggest that the features of sounds within each group's repertoire are being gradually modified over time, and that modifications are constrained in such a way that the differences in sounds used across groups are not increasing over time (Deecke, 1998; Grebner et al., 2011). It has been suggested that just as researchers can identify families of orcas from their call repertoire, the orcas themselves may use calls as signifiers of family membership. However, there is currently no evidence that orcas use sounds in this way. Recent studies suggest that orcas will match the calls of other orcas that they hear in vocal exchanges (P. J. O. Miller et al., 2004; Weib et al., 2011).

As with belugas, observations suggesting that orcas have the capacity to imitate sounds have mostly been opportunistic. Orca calves in captivity may develop calls with features similar to those produced by their companions (Kremers et al., 2012); this has also been reported for adults (Ford, 1991). Orcas may also copy features of man-made sounds present in their environment (van Heel, Kamminga,

& van der Toorn, 1982). There are some indications that orcas in the wild may imitate the calls of other orcas (Ford, 1991), or sounds produced by other marine animals such as sea lions (Foote et al., 2006). Interestingly, apparent reproductions of sea lion barks by a wild orca matched not only the features of individual sounds, but also the rhythmic production of repeated sounds typical of sea lions (Foote et al., 2006). The orca that was observed producing sea lion–like sounds was separated from its family group at a young age, which may have affected this individual's auditory experiences during vocal development.

Call matching and call sharing are generally not viewed as clear instances of vocal imitation (Egnor & Hauser, 2004; P. J. O. Miller et al., 2004; Tyack, 2008). Instead, such repertoire sharing is usually described either in terms of dialect usage (Deecke, Ford, & Spong, 2000) or vocal contagion (Andrew, 1962). This interpretation leads to the somewhat odd situation that if an orca replicates a call that it just heard (call matching) this would not qualify as an imitative act, but if it were to bark like a sea lion in response to that same call, then this would qualify as sound imitation (albeit deferred), because orcas do not normally bark. Although call matching could potentially be explained as vocal contagion, or as a case of an orca selecting a known call from its repertoire, it is important to keep in mind that these possibilities do not compel the inference that an orca matching another's call is not imitating that sound. The orca matching a call could be doing so by copying features of the call it recently heard. The presumption that a call should only be classified as imitative when all other alternative possibilities have been excluded lacks parsimony. If there is evidence that orcas can imitate sounds, and no evidence that they reactively produce calls of a particular type whenever they hear them (as is typical of vocal contagion), then the "simplest" explanation is the one for which there is evidence.

Convergence in Humpback Whale Singing

Humpback whales produce sounds in ways that differ substantially from how dolphins, belugas, and orcas produce sounds, and that are more similar to vocal production by terrestrial mammals (Cazau, Adam, Laitman, & Reidenberg, 2013; Mercado et al., 2010; Reidenberg & Laitman, 2007). The sounds produced by humpback whales are also subjectively quite different from those used by belugas or orcas. Humpbacks do not produce short duration ultrasonic clicks, and their sounds are not commonly classified as being either whistles or pulsed calls. Recent acoustic (Mercado et al., 2010) and anatomical (Reidenberg & Laitman, 2007) analyses suggest, however, that many of the qualitative

---

[4] similar to vocal fry produced by human singers

aural differences in the sounds used by humpback whales reflect quantitative differences in the size and configuration of their vocal organs rather than mechanistic differences in how they produce sounds. Humpback whales do produce click trains (Mercado et al., 2010; Stimpert, Wiley, Au, Johnson, & Arsenault, 2007), but their clicks are much longer in duration and lower in frequency than those used by delphinids[4]. Many of the sounds produced by humpbacks are acoustically comparable to the pulsed calls used by orcas, but shifted to lower pulse rates. Humpback whales also produce higher-frequency tonal sounds, referred to as "chirps" or "cries," that are comparable to the "whistles" produced by toothed whales, but with fundamental frequencies an octave or two lower. As in toothed whales and human singers, the sounds of humpback whales fall along a graded continuum that corresponds to modulations of the pulse rate produced by vibrating membranes (Mercado et al., 2010).

Despite these similarities in the acoustic properties of the sounds produced by humpback whales and toothed whales, there are some key differences in the ways that humpbacks use sounds. Most notably, the repertoire of sounds that a particular humpback whale uses varies from one year to the next (Mercado, Herman, & Pack, 2005; K. Payne & Payne, 1985). More famously, humpback whales rhythmically produce sounds in stereotypical sequences for hours with no break, a behavior that has traditionally been described as singing (R. S. Payne & McVay, 1971). During singing bouts, an individual whale may gradually or rapidly expand or compress the spectrotemporal features of sounds, shift them into different frequency bands, or vary the rate and elemental structure of sequences of sounds (K. Payne, Tyack, & Payne, 1983). The repertoire of sounds produced within songs changes annually such that in each year some distinctive sounds are often no longer evident and others that have not previously been recorded may be prevalent (K. Payne & Payne, 1985). Singing by humpback whales is one of the most dramatic displays of vocal flexibility in any species.

There are no scientific reports of humpback whales reproducing the sounds of other species or man-made sounds. Nevertheless, humpback whale singing is often described as providing the clearest and most impressive evidence of vocal imitation among all cetaceans (Herman, 1980; Janik, 2009b). This is because singing humpback whales in a particular region produce similarly structured songs, despite annual changes in songs. It has been argued that humpback whales *must* be copying the songs they hear being produced by neighboring whales to maintain regional song similarity (Janik & Slater, 1997; Noad, Cato, Bryden, Jenner, & Jenner, 2000; Rendell & Whitehead, 2001; Tyack, 2000). Consistent with this idea, singers may change the features of their songs after being exposed to novel songs. For example,

over a period of a year, whales along the Eastern coast of Australia gradually adopted the songs of a separate population of whales from the west coast, essentially abandoning their original song features in favor of those present within the novel song (Noad et al., 2000). An obvious explanation for such rapid turnover is that whales on the east coast of Australia copied the songs of whales from the west coast.

A musician recently collected further evidence that humpback whale singers alter their songs based on the sounds they experience when he attempted an improvisational duet with a singing humpback whale (Rothenberg, 2008). Rothenberg used an underwater speaker and hydrophone to create a two-way sound channel with a nearby singing whale. By broadcasting clarinet sounds underwater in coordination with the singing whale's sound production, Rothenberg was apparently able to induce the singer to modulate features of its song in ways that matched aspects of the clarinet sounds. A more conventional, non-interactive playback study also found evidence that singers modify their songs based on the features of other songs they hear in their environment (Cholewiak, 2008). Although neither of these acoustic interventions provides clear evidence of sound imitation by humpback whales, they both suggest that singing humpback whales can flexibly adjust their sound production in real time based on sounds they have recently experienced.

If singing humpback whales are copying song features produced by other whales, then this is a rather sophisticated case of deferred sound imitation. First, the songs produced by humpbacks usually last 15 minutes or more, and typically contain 100+ individual sounds produced in five to seven different sequential patterns. A singer would need to encode, retain, and recall multiple properties of an experienced song in order to be able to incorporate these features into an existing song[5]. Second, songs produced by an individual whale on any given day can vary considerably in duration and content, and do not always include all of the regionally prevalent patterns. In other words, individual whales hear and produce multiple renditions of songs that vary in numerous ways (e.g., the number and variety of sounds, which patterns are included, the number of times patterns are consecutively repeated, etc.). Third, singers in many locales will often be exposed to songs from multiple singers simultaneously. To encode songs received in such complex auditory scenes, singers would need to selectively attend to the songs of individual singers while simultaneously hearing other similar songs at different stages within the sequence, possibly including their own song. Finally,

---

[5] Singers typically modify songs by gradually inserting, deleting, or modifying existing patterns within their current song, rather than replacing their songs entirely.

a singing humpback whale would need mechanisms for comparing its own current song with other songs to determine how the songs differ, the kinds of changes required to make the singer's song more similar to those it hears, and whether such changes are warranted. Baleen whales have generally been viewed as cognitively unsophisticated compared to their toothed relatives. However, the perceptual, memory, and attentional processes required to continuously update song features across decades suggests that humpback whales, at least, possess auditory and sound generating capacities that may match or exceed those of delphinids.

Multidimensional Sound Imitation by Bottlenose Dolphins

The sound producing capacities of bottlenose dolphins have been studied more extensively than those of all other cetaceans combined. Much of this work has focused on understanding how dolphins use ultrasonic signals to echolocate (Au, 1993), or on how they use whistles to communicate (Janik, 2000, 2009a; Tyack, 2000; Tyack & Clark, 2000). Like belugas and orcas, bottlenose dolphins produce a variety of sounds and can produce multiple sound types simultaneously. Unlike the fortuitous observations of belugas spontaneously producing speech-like sounds in captivity, and of orcas producing sea lion–like sounds in the wild, the first indications that dolphins could imitate sounds from outside their typical repertoire came from laboratory studies[6]. Lilly (1963) described hearing "queer noises" while conducting brain stimulation experiments designed to investigate basic mechanisms of motivation and reward (Lilly, 1958). Recordings used to dictate notes during the neuroscience experiment revealed that some of the sounds being produced were similar to other sounds on the recordings, including laughter and vocal dictations. These early reports that dolphins appeared to be imitating man-made sounds were initially viewed as implausible (Lilly, 1963). Lilly subsequently performed several behavioral experiments designed to explore whether dolphins could learn to reproduce arbitrary sounds (Lilly, 1961, 1965, 1967, 1968; Lilly, Miller, & Truby, 1968). He discovered that: (1) dolphins could repeat properties of acoustic sequences on command (e.g., matching the number, rate, and rhythm of sound bursts); (2) dolphins typically did not replicate the sounds they copied, but instead reproduced only a subset of features, for instance by speeding up frequency-modulation rates and transposing frequencies into a more natural range; (3) novel vocalizations learned by one dolphin sometimes

are copied by companion dolphins; (4) an adult dolphin was able to learn to copy features of arbitrary sound sequences produced by humans in as little as 2 hours and immediately transferred this ability to copying sounds from tape recordings; (5) dolphins were willing to reproduce sound sequences without any food reinforcement; (6) given repeated presentations of a word or sequence, dolphins naturally modulated

Figure 7. Spontaneous vocalizations produced by a bottlenose dolphin after broadcasts of computer-generated tonal sounds show that dolphins initially imitate subcomponents of the experienced sound (top three images) before producing a more complete copy (adapted from Reiss & McCowan, 1993; Figure 3). Gray lines show spectrographic contours and harmonics of the broadcast sound, and black lines show the contours and harmonics of the dolphin's sounds. Arrows point to components of the target sound that are similar to the sound produced by the dolphin.

---
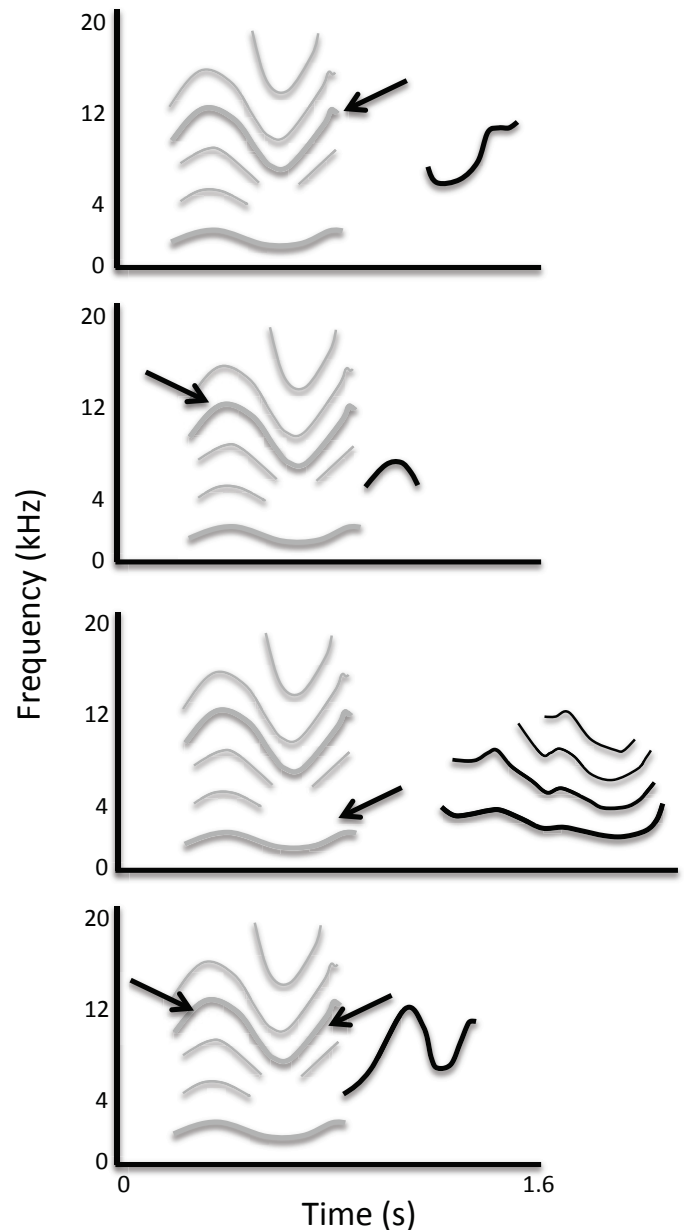
[6] Lilly (1967) noted that Aristotle reported that dolphins made sounds with "a voice like that of the human," so this discovery might be more accurately described as a rediscovery.

their production across repetitions, gradually improving the match of a subset of features; (7) dolphins persevered in reproducing sounds longer if there were natural variations in the targets than if the sound was reproduced exactly (e.g., by repeatedly playing back a recording of a stimulus); and (8) four of four dolphins were able to learn such tasks with varying fidelity. Although many have questioned the rigor and objectivity of Lilly's sound imitation experiments, particularly his reports that dolphins were imitating human speech, several of his observations regarding sound imitation by dolphins have since been independently confirmed.

Anecdotal reports of dolphins spontaneously producing "unnatural" sounds similar to ones they were exposed to in their surroundings provided additional evidence that dolphins could modify their vocalizations to match environmental features (Caldwell & Caldwell, 1972; Tayler & Saayman, 1973). More formal studies of spontaneous imitation in dolphins later confirmed that they reproduced components of computer-generated whistles after as few as 2–20 exposures (Hooper, Reiss, Carter, & McCowan, 2006; Reiss & McCowan, 1993), and that dolphins replicated not only individual sounds, but also rhythmic patterns of sounds (Crowell, Harley, Fellner, & Larsen-Plott, 2005). In the spontaneous imitation studies conducted by Reiss and colleagues, some electronic whistles were associated with the introduction of toys into the tank and others were presented alone. Dolphins reproduced the sounds in both cases, but were more likely to do so (and with higher fidelity) when the sound had been paired with a toy (Hooper et al., 2006). As noted by Lilly (1963), the dolphins often transposed novel sounds and compressed them in time when reproducing them (Hooper et al., 2006; Reiss & McCowan, 1993). Additionally, the dolphins' initial copies of electronic sounds contained only subcomponents of those sounds, which were later combined (Figure 7). In some cases, components of separate sounds were recombined to create novel sounds that the dolphins had never used or experienced previously (Reiss & McCowan, 1993). Reiss and colleagues found that dolphins reproduced sounds immediately after a sound was broadcast and also at later times. Kremers, Jaramillo, Boye, Lemasson, and Hausberger (2011) recently reported that captive dolphins could be heard producing sounds at night that were reminiscent of humpback whale sounds that were broadcast as part of public shows during the day. In this case, the sounds were transposed from the low frequency range produced by humpback whales into a range more typical of dolphin sound production. Dolphins appeared to match both the harmonic structure of the humpback whale calls, as well as their duration and direction of frequency modulation.

In one of the most controlled experimental studies of sound imitation to date, Richards, Wolz, and Herman (1984)
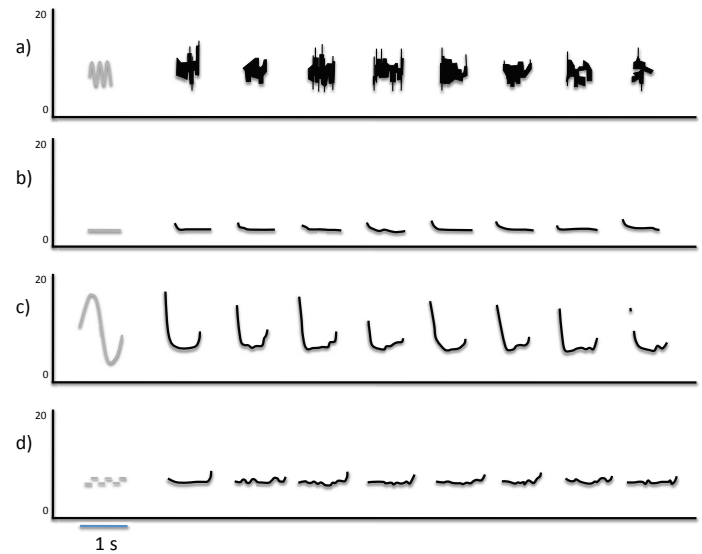

Figure 8. (a–d) Sound reproductions produced in experimental tests of a dolphin's imitation abilities across trials show that sound production is reliable across multiple repetitions and that the dolphin is more likely to replicate some features than others (adapted from Richards et al., 1984; Figure 4). Gray lines show spectrographic contours of four different broadcast sounds, and black lines show the contours of the dolphin's sounds on multiple trials for each of the sounds.

found that dolphins were able to learn to imitate computer-generated sounds on command. As reported by Lilly (1967) and Reiss and McCowan (1993), Richards and colleagues found that dolphins spontaneously imitated sounds before being trained to do so, rapidly learned to generalize the sound imitation task to novel sounds, transposed reproductions into a preferred vocal range, and gradually improved their copies of sounds across trials (Richards, 1986; Richards et al., 1984). Sigurdson (1993) also succeeded in training dolphins to reproduce specific frequency-modulated sounds, but only after extensive training. He concluded that the dolphins initially copied more general features of sounds, and then afterward learned to control details of sound structure through a process of vocal shaping. Richards and colleagues found that once a dolphin settled on a particular mode of imitating a sound, that imitations on subsequent trials were quite stable (Figure 8). Such stable renditions of specific targets can provide important information about the acoustic features that the dolphin attended to, as well as the precision with which dolphins can replicate these features. For instance, Figure 8 shows that a dolphin matched closely the duration of targets as well as the range of frequencies produced. The dolphin also more closely matched the final spectral properties of target sounds than earlier components. Although dolphins sometimes transpose sounds when imitating them, Figure 8 shows that they can also precisely

match absolute pitch. Similarly, although they may expand or compress spectrotemporal properties of a heard sound, they can also closely approximate rates of frequency modulation within sounds. In fact, Richards and colleagues noted that the dolphin even imitated transient distortions produced by the underwater speaker at the onsets of certain sounds.

Research on dolphins is unique in that, although relatively few experiments have been conducted, dolphins have consistently shown sound generating capacities that have yet to be observed in any other non-human species. Various songbirds are able to reproduce environmental sounds with astonishing fidelity (Dalziell & Magrath, 2012), but none have shown the ability to replicate electronic sounds on command, transpose copied sounds into a more appropriate range, or flexibly match the number, rhythm, and rate of sounds across trials. Dolphins are also the only non-human mammal that is known to spontaneously reenact observed episodes integrating both actions and sounds. For instance, in an early anecdotal report, dolphins were seen to imitate a scuba diver cleaning algae from the window of their tank. Not only did the dolphins use an object to scrub algae off the window, but they also released bubbles in bouts while doing so and made sounds described as being almost identical to those of the diver's air-demand valve (Tayler & Saayman, 1973). Such performances strongly suggest that dolphins can flexibly reproduce sounds other than those in their natural repertoire and will occasionally do so in contexts where sound imitation serves no obvious functional purpose.

Richards (1986) argued that the flexibility with which dolphins could imitate novel sounds in controlled experiments indicated that they possessed a generalized concept of imitation that extended to absolute frequency, relative frequency, amplitude modulation, and inadvertent click transients. Lilly (1963) had earlier noted similar generalization of a copying task across rhythm, rate, number, Bronx cheers, and possibly speech. The range and specificity with which dolphins can imitate sounds has yet to be determined. Given the wide range of sounds that dolphins are known to be able to imitate, and the fact that they can match the timing, number, and durations of sound sequences, they are likely able to reproduce at least some sequences of sounds (Crowell et al., 2005). Dolphins might also automatically imitate idiosyncratic features of sounds produced by conspecifics (as has been observed in studies of human speech), although this has yet to be reported. Dolphins have been trained to produce a wide range of sounds on command and to produce matching sounds when they hear another dolphin produce them (Jaakkola, Guarino, & Rodriguez, 2010). Such performances traditionally have been viewed as instances of contextual learning, because the sounds the dolphins reproduce are not novel.

However, no quantitative measures have been made to assess whether dolphins in these situations naturally adjust their vocalizations to match those of recently heard sounds.

Herman (1980) described primates and cetaceans as "cognitive cousins" because despite millions of years of evolutionary divergence within radically different environments, both groups appear to have converged on similar cognitive mechanisms for classifying, remembering, and discovering relationships between events (reviewed by Herman, 1980; Mercado & DeLong, 2010). In the case of sound imitation, this convergence is particularly noteworthy because humans and dolphins are the only mammals that have shown the ability to voluntarily imitate novel sounds. Above, we proposed that the ability of adult humans to imitate sounds is an acquired cognitive skill. Non-human primates do not appear to naturally acquire such skills (at least not vocally), raising the question of why dolphins would acquire a skill that other mammals typically do not. In section five, we suggest that the answer to this question may relate to cetaceans' advanced perceptual use of sound underwater.

Synthesis

A century ago, researchers were optimistic that with the right training, encultured chimpanzees would eventually be able to learn to reproduce speech sounds. At that time, the idea that a dolphin might be better at imitating sounds than a chimpanzee would have been considered absurd. Experimental studies have since shown that dolphins' capacities for imitating sounds exceed those of all non-human primates, and opportunistic observations suggest that other cetaceans may share this capacity. The evidence for sound imitation abilities in other cetaceans is anecdotal, but remains stronger than for most mammals, including non-human primates. The extent to which adult cetaceans use their imitative abilities in their daily lives remains unclear. There have been no studies of automatic imitation in any cetacean. It is also not known how the fidelity with which cetaceans can copy different acoustic features varies either within or across species and individuals. The few laboratory studies of sound imitation by cetaceans to date have focused on showing that they can imitate sounds rather than on revealing how they are able to do this. The extent to which the sound imitation abilities of adult cetaceans depend on practice is uncertain, but clearly dolphins can learn to refine their ability to reproduce man-made sounds, and field observations suggest that they may also regularly reproduce conspecific sounds in natural interactions.

The willingness of dolphins to interact with humans in experimental contexts provides numerous opportunities for sound imitation studies that would be impossible to conduct with humans. Individuals can be trained across

multiple years to perform a wide range of tasks. In principle, one might control a dolphin's exposure to many complex acoustic events, including musical patterns and sequences of speech sounds. Most importantly, such studies potentially allow for cross-species comparisons that are not feasible with non-human primates. If the ability to imitate sounds depends on evolutionarily specialized processing, then one would expect that cetaceans' abilities to imitate sounds should differ systematically from those of humans in ways that directly reflect the many differences in their ecological circumstances. If, however, these abilities are highly dependent on training and practice, then it might be possible to endow individuals from different species with similar imitative capacities by training them on tasks with similar demands. Given that the functions of sound imitation in adult mammals are poorly understood, it remains possible that cetaceans and humans evolved similar capacities to learn imitative skills because they faced similar perceptual or cognitive challenges. In the following section, we consider this possibility more closely.

V. Proposed Origins and Functions of Sound Imitation Abilities

When biologists and psychologists discuss vocal imitation as a learning mechanism, it invariably is in the context of explaining how and why individuals acquire communicative skills during development. Consequently, when it comes to explaining why some species have the ability to imitate sounds, many researchers focus on describing the benefits associated with effective communicative systems, such as enhanced mating opportunities, greater possibilities for complex social interactions, ability to identify familiar individuals, and so on. Although such explanations provide plausible reasons for why adaptations for sound imitation abilities might persist once they appear in a species, they are less able to account for why these abilities are so rare among terrestrial mammals.

The sophisticated sound imitation abilities of adult humans suggest that these abilities may be advantageous for reasons other than (or in addition to) learning to talk, such as predicting and perceiving the actions of others (Wilson & Knoblich, 2005). Below, we consider whether this might also be true for cetaceans. We conclude that current evolutionary and functional explanations for the prevalence of sound imitation abilities in cetaceans, which focus on the role of vocal learning in social communication, are inadequate, and propose an alternative explanation in which increasing perceptual-motor and cognitive demands related to non-visually guided movement coordination led to advanced sound localization abilities that are enhanced by sound imitation

capacities. First, however, we review past attempts to explain why cetaceans evolved the ability to imitate sounds.

Do Mammals Imitate Sounds to Enhance Social Communication?

The main hypotheses typically proposed for why different cetaceans imitate sounds are that this ability: (1) enables group recognition and maintenance of group cohesion (e.g., in orcas); (2) aids in the learning of a vocal badge that can be used as a password for access to local resources (e.g., in bottlenose dolphins); (3) provides a way for males to increase the complexity of sound production, thereby increasing their attractiveness to females (e.g., humpback whales); (4) enables individuals to display their prowess and better fend off competing males; and (5) helps individuals to recognize each other in noisy environments (Janik & Slater, 1997; Tyack, 2000). Janik (1999) collapsed these possibilities into two global hypotheses: the sexual selection hypothesis and the individual recognition hypothesis. In both cases, the proposed driving force for the evolution of vocal learning and imitation abilities in cetaceans is a need to facilitate communication of either fitness or identity. Janik (2009a) further suggested that sound imitation abilities in cetaceans subserve complex communication mechanisms that are necessitated by complex social systems.

The idea that bottlenose dolphins evolved the ability to imitate sounds to enable them to develop individual-specific whistles (referred to as signature whistles) that serve as a vocal badge or naming signal (Fripp et al., 2005; Janik, 2000; Janik & Sayigh, 2013; Janik, Sayigh, & Wells, 2006; King, Sayigh, Wells, Fellner, & Janik, 2013; Quick & Janik, 2012), arose from early observations that captive dolphins in isolation often repeatedly produced a stereotyped whistle with distinctive features that were specific to the vocalizing individual (Caldwell & Caldwell, 1965; see Harley, 2008, for a review). It was later noted that in some situations, dolphins would produce a whistle that was highly similar to the signature whistle of a tank-mate; these whistles have been described as being signature whistle imitations (Agafonov & Panova, 2012; Tyack, 1986). Researchers have hypothesized that a dolphin might imitate a signature whistle to communicate with or about specific individuals (Janik, 1999; King et al., 2013; Richards et al., 1984; Tyack, 1991).

It has also been suggested that sound imitation serves an important role during dolphin vocal development, enabling young bottlenose dolphins to acquire signature whistles that reflect their lineage (Sayigh, Tyack, Wells, & Scott, 1990; Sayigh, Tyack, Wells, Scott, & Irvine, 1995; Sayigh et al., 1999), or social affiliations (Fripp et al., 2005; Watwood, Tyack, & Wells, 2004). The role of vocalizing "tutors" in the vocal development of bottlenose dolphins is

generally thought to be similar to what is seen in human children and songbirds (Reiss & McCowan, 1993). Observations of dolphins born in captivity support the idea that vocal development is shaped by the sounds dolphins experience in their surroundings (Caldwell & Caldwell, 1979; McCowan & Reiss, 1995; Miksis, Tyack, & Buck, 2002; Reiss & McCowan, 1993; Tyack & Sayigh, 1997). However, such observations provide little evidence that experience-dependent repertoire acquisition serves primarily to distinctively signify a vocalizing dolphin's identity, or that sound imitation plays any role in such a process.

Whereas sound imitation abilities in toothed whales have been postulated to be important for the learning and development of acoustic identifiers, the apparent sound imitation abilities of humpback whales have been described as serving a role in sexual advertisement (Janik, 2009b; R. S. Payne & McVay, 1971; Smith, Goldizen, Dunlop, & Noad, 2008). For instance, Tyack and Sayigh (1997, p. 229) suggest that, in humpback whales, "vocal learning appears to function to produce more complex displays through sexual selection." Janik (1999) suggested that the ancestors of humpback whales may have initially evolved sound imitation abilities for individual recognition functions, but that over time this ability came to serve a reproductive function.

Limitations of Current Evolutionary and Functional Hypotheses

A prevalent assumption regarding vocal learning and imitation in cetaceans is that because different species have divergent social systems, the origins and functions of sound imitation must be similarly diverse (Janik & Slater, 1997; Tyack & Sayigh, 1997). For instance, Tyack (2000, p. 307) speculated that, "it is possible that vocal learning[7] evolved *de novo* in these different taxa as independent solutions to different problems posed by their different social organizations." While it is certainly possible that different cetacean species developed sound imitation abilities independently in response to their particular social and reproductive pressures, it is also possible that the origins and functions of sound imitation in cetaceans are not as disparate as they might at first appear. For instance, Deacon (1997) hypothesized that cetacean sound imitation abilities are an exaptation of adaptations for skeletal motor control of airflow. In this scenario, new demands on motor control related to voluntary breathing gave rise to increased vocal flexibility, as well as a dissociation between mechanisms involved in producing reactive/emotive vocalizations and other more voluntarily produced sounds (see also Mithen, 2009).

Deacon's hypothesis makes no assumptions about the functions of either sound imitation or the sounds being imitated, and can potentially account for the emergence of vocal imitation abilities in all cetacean species as well as in humans. A limitation of his hypothesis is that it does not explain any benefits cetaceans might gain from imitating sounds. In fact, Deacon suggests that some mammals famous for imitating speech may have been showing signs of neural dysfunction.

Past proposals that vocal imitation is an evolutionary outcome of either sexual selection or adaptations for enhanced individual recognition suffer from several limitations. First, these hypotheses attempt to account for the emergence of vocal learning and imitation abilities in cetaceans in terms of hypothetical functions of the sounds cetaceans produce. However, the specific functions of most cetacean sounds have yet to be established experimentally. The hypothesis that humpback whale songs function as reproductive displays to attract females and repel males is based on circumstantial evidence (Frazer & Mercado, 2000; Mercado & Frazer, 2001), and does not account for many known behaviors of singing whales (Darling, Jones, & Nicklin, 2012; Darling, Meagan, & Nicklin, 2006; Stimpert, Peavey, Friedlaender, & Nowacek, 2012). Although there is substantial evidence that bottlenose dolphins produce whistles that humans can use to identify them (Harley, 2008; Janik & Sayigh, 2013), there is no evidence that this is a primary function of whistles or that dolphins have difficulty identifying other dolphins that are not producing signature whistles (McCowan & Reiss, 2001). Second, neither the sexual selection nor the individual recognition hypothesis leads to predictions other than those related to the speculated functions of a small subset of cetacean sounds. Consequently, these evolutionary hypotheses are little more than a restatement of pre-existing functional hypotheses. Third, the sexual selection and individual recognition hypotheses require one either to assume that all cetaceans have a common ancestor that developed sound imitation abilities for sexual or identification purposes, after which the functions of these abilities later diverged dramatically across different species depending on their social systems (Janik, 1999), or that each species of cetacean independently evolved sound imitation abilities to meet their particular social needs (Tyack, 2000). Why either of these scenarios might have occurred in cetaceans, but not other mammals, is unclear given that many mammals (e.g., primates and canids) often engage in complex social interactions in situations where visual information is limited.

---

[7] Here, the term vocal learning is meant to include vocal imitation.

A New Hypothesis:
Imitatible Sounds Are More Localizable

Most current hypotheses regarding the origins and functions of sound imitation in cetaceans were originally developed as explanations for the evolution of song learning by birds (Thorpe, 1969; Thorpe & North, 1965). Here, we consider whether sound imitation abilities may provide adult cetaceans with other previously unsuspected benefits. Specifically, we assess the possibility that the capacity to imitate sounds might enable cetaceans to localize sound sources more accurately. In this scenario, sound imitation abilities may have appeared early in the evolution of cetaceans and then been preserved throughout the differentiation of species because the advantages of such capacities persisted despite differences in social organization and behavior.

As with the communication-focused hypotheses described above, the idea that vocal learning or imitation might enhance spatial perception was originally proposed to account for the evolution of song learning in birds (Morton, 1982, 1986, 1996, 2012). This hypothesis, referred to as the "ranging hypothesis," states that a listening bird will be better able to estimate its distance from a singing bird if the listener can compare received songs with an internal representation of the song as it would appear at the source. *Ranging* is a perceptual process in which an individual uses a received sound (or sounds) to estimate the distance to the source of that sound. Sound transmission can degrade the acoustic features of a song. By comparing an undistorted representation of the song with the received song, the listener may identify how transmission has changed song features. Changes in songs caused by propagation are thought to be the primary cues that enable birds and mammals to estimate auditory distance (Naguib & Wiley, 2001). The ranging hypothesis thus suggests that the accuracy with which a listener can judge auditory distance is constrained by its ability to compare received songs with internal representations of "pristine" songs. The ability to imitate a received song, either overtly or covertly, gives the listener direct access to features of the song as they would appear at the source. Thus, the ability to imitate sounds could improve a bird's ability to estimate auditory distance, which could give the bird a selective advantage in spatial interactions with competitors. Morton (1996) also suggested that male songbirds might selectively sing songs with acoustic features that make ranging difficult so that other males have problems locating them during territorial disputes.

Playback studies in songbirds have tested the ranging hypothesis by comparing territorial birds' responses to familiar and unfamiliar songs broadcast within and outside of a listener's territory (Falls & Brooks, 1975; Morton, Howlett, Kopysh, & Chiver, 2006; Shy & Morton, 1986). Listening birds responded more aggressively to familiar songs produced inside their territory than to those outside of their territory (Shy & Morton, 1986). Listening birds also expended more energy searching when unfamiliar songs were produced outside of their territory, suggesting that they may have been less certain of the singer's location. Finally, birds approached a playback speaker more closely when the song was familiar (Morton et al., 2006), indicating that they were better able to localize the speaker when it was broadcasting familiar songs. Although one cannot assume that all familiar songs are more imitatible than unfamiliar songs, if a song is familiar because it is within the listening bird's repertoire, then it is likely to be highly imitatible.

Cetaceans are not generally territorial, but they often encounter situations in which precise spatial hearing is important, as evidenced by their use of echolocation. Echolocation differs from ranging in that an echolocating animal controls the sounds it uses to localize environmental features, whereas a ranging animal uses sounds produced by other animals to localize them. Possible links between the evolution of echolocation and the emergence of sound imitation abilities have been previously noted (Tyack & Clark, 2000), but have received little scientific attention. Applied to cetaceans, the ranging hypothesis suggests that sound imitation capacities may have developed in cetaceans for the same reason as echolocation—to enhance auditory spatial perception in a visually limited environment.

Determining the distance to a sound source might seem like a rather trivial ability, one that an organism could easily achieve through mechanisms less complex than sound imitation. Intuitively, one might suspect that simply looking at the source would usually solve the problem. When a source is not visible, as may often be the case for cetaceans, then variations in amplitude might appear to suffice (e.g., the quieter the sound, the farther the source). Amplitude cues are only grossly correlated with source distance, however, and for sounds propagating in the ocean such cues would provide little if any information about the trajectory of a vocalizing conspecific. The ambiguity of amplitude cues arises, in part, because individuals may vary how loudly they produce sounds and because sounds repeatedly reflect from the ocean surface and bottom, creating complex patterns of constructive and destructive interference. As a result, amplitude can fluctuate dramatically for reasons unrelated to variations in distance (e.g., Mercado & Frazer, 1999). It might also seem that if a species can echolocate, then additional mechanisms for locating other individuals would be redundant. Undoubtedly, cetaceans do sometimes use echolocation to range other animals. This is a much less efficient means of coordinating the movements

of multiple individuals than passive localization, however, because it requires that every individual continuously echolocate in multiple directions to keep track of all the other individuals in a group. Furthermore, such active sound production would reveal the locations of all members of the group to prey or competitors, which is likely to be disadvantageous in many situations. Humpback whales, belugas, orcas, and bottlenose dolphins are all known to engage in sophisticated foraging strategies in which multiple animals must coordinate their underwater movements in three-dimensions to corral prey (Connor, 2000; Wiley et al., 2011), and they often synchronize their movements within groups (Fellner, Bauer, & Harley, 2006; Perelberg & Schuster, 2008). Coordinating invisible movements in the ocean can be a highly challenging task. A listening whale or dolphin may need to track multiple sources simultaneously and to move or produce sounds contingently based on the sounds it hears. If sound imitation abilities enhance a cetacean's capacity to monitor and predict the movements of conspecifics, then sound imitation may be more prevalent in cetaceans than in terrestrial mammals because reduced availability of visual cues for coordinating actions underwater increased reliance on alternative perceptual strategies.

Do Mammals Imitate Sounds to Enhance Their Perception of Actions?

A specific prediction of the ranging hypothesis is that a listener will be better able to localize the source of a sound if the listener can reproduce that sound. Unfortunately, it is not known whether cetaceans' auditory distance estimates vary with sound type. In fact, there are no measures of the accuracy with which cetaceans can judge the auditory distance of any sound source other than targets that they have echolocated (Au, 1993). To test whether imitatible sounds are easier for cetaceans to localize, one would need to broadcast various sounds at known distances, and then assess how accurately individuals can estimate the distance of the source[8]. Given the logistical difficulties associated with conducting such experiments with cetaceans, an alternative approach is to first investigate whether other species (e.g., humans) show improved spatial processing of imitatible sounds.

Predictions of the ranging hypothesis have never been explicitly tested in mammals, but there have been numerous studies of auditory distance estimation in humans. Human sound localization abilities are quite good relative to other mammals (Blauert, 1997). Nevertheless, the accuracy with which humans can estimate the distance to a sound source

varies considerably (Zahorik, Brungart, & Bronkhorst, 2005). Familiarity with sound features can dramatically improve an individual's ability to range the source of that sound (Coleman, 1962; Little, Mershon, & Cox, 1992). Humans are also known to be better at ranging speech than artificial sounds (Gardner, 1969), and to be better at ranging forward speech than speech played backward (McGregor, Horn, & Todd, 1985; Wisniewski, Mercado, Gramann, & Makeig, 2012). Because backward speech contains all of the acoustic information present in forward speech, any environmental degradation of sound features associated with propagation will be the same for both forward and backward speech. Consequently, any differences in an individual's ability to estimate the distance of these sounds lies within the listener, not within the received signals.

A recent study of auditory distance estimation by humans hearing familiar and foreign speech sounds found that the advantage for forward speech still holds for an unfamiliar foreign language (Wisniewski et al., 2012). Thus, familiarity per se does not seem to be the key factor that makes speech more localizable. According to the ranging hypothesis, the greater accuracy at ranging forward speech comes from the fact that speech is a highly imitatible acoustic event, and thus is encoded in ways that make replication of the heard sounds possible (Skoyles, 1998). Backward speech, in contrast, contains acoustic trajectories that would be difficult or impossible to reproduce with vocal acts (Cowan, Braine, & Leavitt, 1985), and so cannot be reconstructed with the same fidelity. The ranging hypothesis thus provides a possible explanation for differences in the accuracy with which humans can judge auditory distance for particular sound types.

The ranging hypothesis is similar in many respects to Wilson's (2001b) and Wilson and Knoblich's (2005) hypotheses that imitation may enhance an individual's ability to perceive and predict the actions of conspecifics. Wilson (2001b) suggested that mental representations formed during covert imitation facilitate the flow of information processing between perception and action, especially when the stimuli and actions are familiar. More specifically, Wilson and Knoblich proposed that visual perception of other persons' behaviors activates covert imitative motor representations that feed back into the perceptual processing of observed actions, leading to expectations and predictions of ongoing action trajectories. Consistent with this proposal, people are better able to recognize actions via point-light displays if the actions are ones that they themselves can perform (Blake & Shiffrar, 2007; Casile & Giese, 2006). Auditory processing of conspecifics' vocal acts might similarly activate covert imitative motor representations that facilitate the mental representation of non-visible

---

[8] One complication of this approach is that it is difficult to establish how well a listener can imitate a sound unless the listener is known to make that sound or actually imitates the broadcast sound.

movements of a sound's source through space. Although the ranging hypothesis, as proposed by Morton (2012), does not specifically address the possibility that such acoustically triggered representations might facilitate a listener's ability to track or predict a singer's future actions, it is likely that more accurate ranging of vocalizing conspecifics would facilitate monitoring of their movements.

Past laboratory studies of sound imitation by dolphins suggest that they gradually improve the fidelity of their copies through repeated practice (Lilly et al., 1968; Reiss & McCowan, 1993; Richards et al., 1984; Sigurdson, 1993), and that this gradual improvement reflects incremental refinement of vocal control. Perceptual-motor skill learning related to vocal production thus likely plays a role in the development of capacities for imitating specific sounds. The ranging hypothesis predicts that as an individual's facility at producing a particular sound improves, his or her ability to represent and imitate that sound should also gradually improve, which could indirectly lead to improvements in spatial localization abilities. Thus, vocal learning may play an important role in the functionality of sound imitation for cetaceans, but in the opposite direction from what is typically assumed. Most researchers assume that the purpose of vocal imitation is to enable individuals to rapidly learn new ways of producing sounds from others (e.g., Whiten & Ham, 1992). The ranging hypothesis suggests instead that individuals may learn new sound production skills to enhance existing perceptual capacities (for a review of how motor skills can enhance perception, see Wilson & Knoblich, 2005). Specifically, rather than imitating novel sounds to increase or specialize their vocal repertoire, cetaceans may practice producing different sounds to increase their vocal flexibility, thereby increasing the variety of sounds that they can imitate, which in turn might increase their ability to localize sources of similar sounds.

Synthesis

Current explanations for why cetaceans evolved the ability to imitate sounds focus heavily on the role of imitation in vocal repertoire formation and modification. Such explanations meld well with proposed functions of vocal imitation in speech and language learning by young children. When cetaceans' abilities are viewed through the lens of vocal imitation research in adult humans, however, an alternative possibility emerges. Namely, that the benefits of sound imitation abilities for adult cetaceans may relate more to enhancing the perception and dynamic coordination of movements than to cementing social bonds, selecting a moniker, or attracting a mate. Of course, enhanced perceptual and coordination abilities may facilitate a wide array of functions, including mating, communicating, and other social functions. Nevertheless, sexual selection for fitness revealing traits and adaptive specializations for species-specific social needs are likely to involve different adaptations and mechanisms from those associated with natural selection for basic perceptual abilities. The hypothesis that vocal imitation in cetaceans is a perceptual adaptation predicts that the most proficient imitators will be adults rather than immature individuals, and that through extensive practice, cetaceans may be able to increase not only their sound imitation skills, but also their capacity to localize sound sources, and their ability to represent and predict dynamic events. In the following section, we consider more closely the role that learning plays in the refinement of sound imitation abilities and explore whether a unified framework can potentially describe and explain these abilities in both cetaceans and primates.

VI. Proposed Mechanisms for Imitating Sounds

A successful model of vocal imitation, and of sound imitation more generally, must be able to account for known flexibilities in imitative abilities and for documented sensitivities to stimulus complexity. Ideally, the model should also be able to account for the role of sound imitation in perception and production across mammalian species. Having discussed empirical findings from both primates and cetaceans, we now review some of the leading theoretical models of vocal imitation and consider how well they account for the available data. In so doing, we revisit general themes discussed at the beginning of the paper, but with respect to specific mechanisms proposed by different theories. By our reading, the literature to date supports the notion that vocal imitation abilities emerge in mammals as a learned skill that is suited to the particular constraints faced by the species in question, and that involves the construction of multimodal representations of acoustic events.
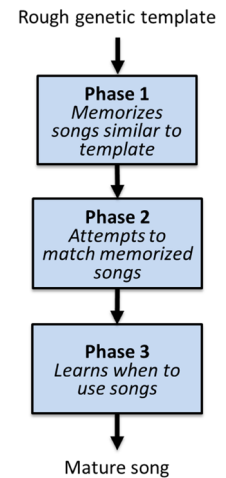

Rough genetic template

Phase 1
Memorizes songs similar to template

Phase 2
Attempts to match memorized songs

Phase 3
Learns when to use songs

Mature song

Figure 9. Template model of vocal learning and imitation originally developed to explain how birds learn songs and subsequently used as a model of vocal imitation.

Vocal Imitation as Template Matching

Researchers studying animals other than humans have often described the processes underlying vocal imitation as simple, unimodal, and transparent to the vocalizing individual. For instance, Whiten and Ham (1992) suggested that to reproduce a sound, a bird only needed to adjust its output until the produced sound matched what the bird had originally heard. They contrasted this process with visually based motor imitation, which they described as requiring additional levels of representation and greater computational capacities. This auditory-feedback based explanation of the processes involved in vocal imitation is derived from a model that was originally developed to account for song learning by birds—the auditory template model (Konishi, 1965; Marler, 1976b). In the template model, birds start out with an internal auditory representation of what a song should sound like (acquired either genetically or through memorization, Marler, 1997), and then gradually learn to produce sounds that match this auditory template through a process of sensorimotor learning (Margoliash, 2002; Marler, 1976b, 1997). Computational instantiations of the template model show that such error-correction mechanisms are sufficient to generate sound patterns that match a prescribed target (Troyer & Doupe, 2000a, 2000b). When vocal imitation is construed as an instance of vocal learning, this model provides a relatively simple account of the necessary underlying mechanisms (Figure 9).

The auditory template model rests on several assumptions that make it problematic as a simple account of vocal imitation abilities, however, including: (1) heard sounds are selectively filtered such that particular sequences produced by conspecifics trigger unique auditory memory mechanisms; (2) experiences of these favored sequences are internally stored via something like the auditory equivalent of eidetic memory after a single or very few exposures; (3) once formed, these memories last indefinitely and are immediately reactivated whenever an individual vocalizes; (4) any mismatch between the permanent auditory template and a produced sound will lead to changes in sound production to minimize those differences; and (5) the fundamental process enabling vocal imitation is auditory feedback (see Petrinovich, 1988, for a more detailed critique of these assumptions in relation to theories of bird song learning).

Vocal imitation studies in adult humans suggest that neither detailed long-term auditory memories nor auditory feedback are necessary to reproduce sounds. A human can readily imitate a novel melody even if masking noise is presented over headphones such that it is very difficult for the person to hear their own vocalizations (Pfordresher & Brown, 2007), although intonation may deteriorate slightly

(Mürbe, Friedmann, Hofmann, & Sundberg, 2002; Ward & Burns, 1978). According to the template model, the mismatch between what is produced (voiced pitches) and what is heard (noise) should lead to large changes in sound production; however, no such changes have been reported. In fact, much larger changes in vocal production are observed when auditory feedback exactly matches the produced sound, but shifted slightly in time (Pfordresher & Mantell, 2012; Smotherman, 2007). More generally, comparisons between produced sounds and previously heard sounds are *not* necessary for an adult human to reproduce novel sounds. In particular, when a person imitates a novel sound for the first time, feedback cannot guide the vocal act because the motor acts that constitute the imitative act are selected and executed prior to any feedback being available. Thus, organisms that can accurately imitate novel sounds upon first presentation are controlling their sound producing actions such that they will generate perceived similarities, rather than using those similarities to discover what actions to perform. Theories of vocal imitation that assume auditory feedback renders vocal imitation fundamentally different from other forms of imitation have conflated the act of vocal reproduction with an individual's post-hoc assessment of similarities between produced sounds and remembered sounds. The availability of auditory feedback can be an important component of vocal learning, but it is neither necessary nor sufficient for flexible vocal imitation, and in some cases may even degrade an individual's imitation abilities.

Despite the limitations of the auditory template model as a model of vocal imitation, it can potentially provide insights into how and why mammals change the way they imitate sounds over time. For instance, past studies of spontaneous and instructed vocal imitation by bottlenose dolphins consistently show that dolphins gradually refine their reproductions of experienced sounds (Lilly, 1967; Reiss & McCowan, 1993; Richards et al., 1984; Sigurdson, 1993), with later renditions showing more similarities to targets than earlier versions. The template model provides a reasonable account of such gradual adjustments in performance.

Vocal Imitation as the Operation of Adaptively Specialized Modules

Marler (1997), recognizing that the auditory template model was insufficient to account for vocal imitation by birds (especially when the sounds being imitated were from other species), proposed two distinct modes of vocal learning: one involving the template-based system that is specialized for learning songs produced by conspecifics, and a second system, described as "general auditory mechanisms," that enabled birds to imitate other sounds by bypassing or overriding the template-based system. Marler's proposal that

separate auditory mechanisms might be used for imitating different kinds of sounds converges with a second way of conceptualizing vocal imitation—as operations performed by one or more specialized cognitive modules.

A modular architecture of cognition, as originally proposed by Fodor (1983), assumes that certain cognitive functions are driven by specialized processors that operate independently from each other. Fodor and others who have grappled with the notion of modularity have proposed many features of cognitive modules, but the two features that dominate the literature include informational encapsulation (a module's functioning is not influenced by processing of information in other modules) and domain specificity (a module is selective with respect to the type of input it will process). Modular approaches that are relevant to vocal imitation have proposed distinct modules for imitation, thus leading to the possibility that stimulus-specific modules may mediate certain kinds of imitation.

One modular approach to imitation in general, and not just vocal imitation, was proposed by Subiaul and colleagues (Subiaul, 2010; Subiaul et al., 2012). In his "multiple imitation mechanisms" approach, imitative modules are divided into vocal imitation, motor imitation (imitation of visually presented information through manual gestures), and cognitive imitation (copying an inferred pattern of thought). Superordinate to this division, he further divides imitation into separate processes for the imitation of novel versus familiar stimuli. This approach shares with the perspective we have advocated the idea that vocal imitation is genuinely a form of imitation, albeit one that may be guided by distinct mechanisms from other forms of imitation. However, in proposing that vocal imitation abilities depend on six specialized cognitive modules, Subiaul's model diverges from the present account in two important respects. First, the conceptualization of vocal imitation as being the dedicated function of two adaptively specialized systems runs against the present argument that imitative skills are learned. Second, because the domain specificity in Subiaul's model is limited to auditory inputs, the model does not explain differences in imitation across domains such as music and language, which we turn to next.

A highly influential modular architecture of auditory processing was proposed by Peretz and Coltheart (2003). Although this model was not intended to be a model of imitation per se, its scope is broad enough to make systematic predictions about imitation within each domain. According to the Peretz and Coltheart model, individuals are endowed with processing modules specialized for

analyzing particular features of sounds that are then used as a basis for guiding vocal actions. These features are processed differently for inputs that represent linguistic versus musical domains. One might use a module specialized for extracting pitch when imitating melodies, another focused on phonology when imitating speech, and possibly a third when vocally reproducing percussive rhythms. This framework suggests that different processing mechanisms are required to form particular kinds of auditory templates and that which template formation process is used depends on categorical features of auditory inputs. This approach is consistent with some of the vocal imitation data from adult humans, in particular the general advantage for imitating absolute pitch content within the domain of music as opposed to speech[9] (Mantell & Pfordresher, 2013). The assumption that these auditory modules are informationally encapsulated is, however, inconsistent with the observed effects of phonetic information on the imitatibility of both musical and spoken sentences.

A variant of this multiple module approach was recently proposed by Patel (2003), in which musical and linguistic representations are separately constructed by independent, specialized processing systems, but then manipulated or used by a third shared system that constrains how both types of representations are used. For example, an individual's ability to parse syntactical structures or to recognize chord progressions might both depend on integrating multiple elements within a sound sequence. A shared system for sequence integration might thus lead to correlations in an individual's fidelity at imitating different sound sequences, even if the auditory templates formed by different categories of sounds are independent of one another. In this view, there are specialized mechanisms for representing different categories of sound sequences (and forming associated templates), as well as general cognitive mechanisms that may constrain an individual's ability to reproduce all kinds of sequences.

Vocal Imitation as Auditory-Motor Recoding

All of the above models focus on comparisons of auditory representations of sounds as being the key mechanism of vocal imitation, while minimizing the role of other contributing mechanisms, such as characteristics of the vocal motor system. These models beg the question of why so few mammals show vocal imitation abilities, given that many mammals (including all primates) have sophisticated auditory systems. As noted earlier, some researchers have suggested that a more crucial mechanism underlying vocal imitation relates to neural control of skeletal muscles involved in vocalizing (Arriaga & Jarvis, 2013; Deacon, 1997; Fitch, 2010). Humans have greater control of tongue and laryngeal movements than most

---

[9] However, an advantage was not found for the imitation of relative pitch content.

other primates and may possess specialized neural regions for directly controlling these movements. Other species known to imitate sounds, such as some songbirds, also have more fine control over vocal membranes than is typical for mammals. The basic idea proposed by Deacon, Fitch, and Jarvis is that these specialized motor control circuits provide humans and a few other mammals with uniquely flexible vocal control processes, and that it is this heightened vocal dexterity that makes vocal imitation possible.

The role of the motor system in vocal imitation, and more broadly in perception, has been assessed in studies of human speech production and imitation. Speech researchers have posited additional mechanisms that may shed some light on processes that facilitate the imitation of sounds. Foremost among these is the proposal that received speech sounds are encoded not only via auditory representations, but also in terms of the motor gestures required to generate particular speech sounds (Corballis, 2010; Galantucci, Fowler, & Turvey, 2006; Liberman & Mattingly, 1985; Lindbolm, 1996; Vallabha & Tuller, 2004; Yuen, Davis, Brysbaert, & Rastle, 2010), and possibly in terms of the somatosensory signals that occur during the production of speech (Guenther, 1995; Studdert-Kennedy, 2000). Wilson (2001b) similarly suggested that imitatible stimuli are not represented solely in terms of their unimodal perceptual properties, but also in terms of articulatory gestures. Such mechanisms provide a ready explanation for how an individual might reproduce novel sounds without auditory feedback on a first attempt. Specifically, if the representation that guides one's vocal acts during vocal imitation is the motor representation required to produce a heard sound, then mismatching auditory feedback (or the lack of repeated instances of mismatching feedback) would have relatively little impact on vocal performance. Numerous theories have been proposed for how one might transform acoustic inputs into "matching" vocal gestures (reviewed by Galantucci et al., 2006), as this is often suggested as a fundamental mechanism of theories of speech imitation. When applied to vocal imitation, this perspective can be viewed as a multimodal representational model in which the key mechanisms correspond to cross-modal transformations rather than error correction based on unimodal auditory comparisons.

A related model of rapid speech imitation (shadowing) developed by Fowler and colleagues similarly suggests that speech sounds may be encoded in terms of motor representations (Galantucci et al., 2006; Honorof et al., 2011; Shockley et al., 2004). Specifically, they suggest that speech may be encoded in terms of the actual motor commands used to control vocal acts rather than (or in addition to) representations of gestures and associated kinesthetic stimuli. Such abstract control parameters might relate to constraints on

trajectories of movement patterns (e.g., the order of speech primitives) rather than the specific motor gestures required to implement those patterns. The mechanisms emphasized by this model relate to controlling a nonlinear dynamical system rather than to creating analog representations of perceived events (Shockley et al., 2004). This approach provides a plausible account of why humans automatically imitate certain features of speech and may also be able to explain vocal convergence within social groups of non-humans.

A problem this sort of model confronts in accounting for the present data has to do with the flexibility of imitation as well as the etiology of imitative deficits. On the one hand, in proposing a specific auditory-vocal equivalence, such motorically constrained models seem ill equipped to account for the fact that imitation of sounds can be performed non-vocally, and that non-vocal sounds can be imitated vocally, often with high accuracy. On the other hand, in proposing a simple perceptual/motor equivalence, which is associated with fluency in speech, such models have difficulty accounting for the fact that imitative deficits can occur in individuals who are apparently able to fluently control phonation and articulation. Moreover, suggestions of perceptual/motor equivalence assume that the transformation from sensory to motor representations is effectively a non-issue. This stands in contrast to the apparent basis of poor-pitch singing, which appears to reflect a deficit of sensorimotor translation (Pfordresher & Brown, 2007).
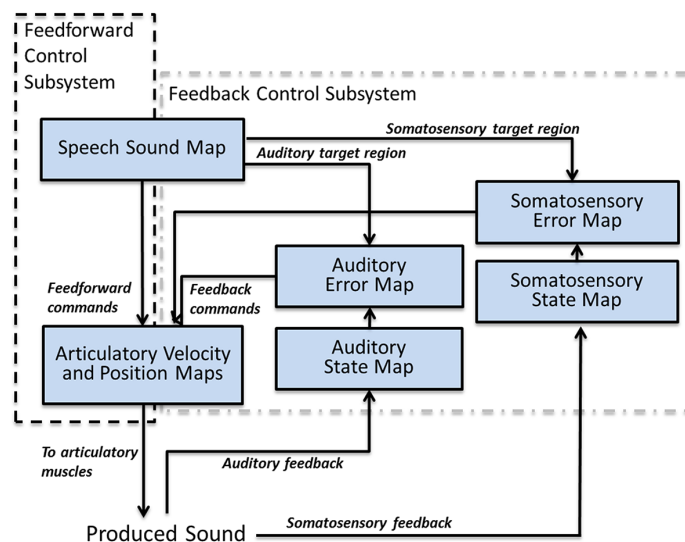
Vocal Imitation as Multimodal Mapping

Another approach to modeling vocal imitation is also based on sensorimotor interactions, but adopts a broader, more flexible framework than the theories discussed above. This approach suggests that sensorimotor translation effects can span multiple perceptual and motor modalities. Such ideas stem from music cognition researchers who have suggested that the capacity of musicians to imitate sounds depends on coordinated auditory, kinesthetic, visual, and spatiomotor processes (described by Baily, 1985, as "auro-motor coordination"), which are developed through experience and which enable some individuals to immediately reproduce musical patterns either vocally or instrumentally. At the core of such musical reproduction abilities lies hypothetical mechanisms of auditory imagery, which enable one to plan and control the production of complex, extended sound sequences (Baily, 1985; Pfordresher & Halpern, 2013). Auditory imaging can be viewed as analogous to visualization processes, enabling a musician not only to reproduce songs, but also to creatively modify those songs (e.g., transforming them into the styles of various musical genres, transposing them into different keys, etc.). Expert musical reproduction is also thought to require sophisticated

conceptual processes acquired through extensive training, allowing heard (or imagined) sound sequences to be reproduced in the form of symbolic visual notations (Gordon, 2007). The supplemental mechanisms required for such flexible reproduction of sounds are not well specified, but clearly involve more than simple unimodal comparisons. In particular, they seem to require some means of voluntarily controlling vocal imitation. Consideration of the possible mechanisms that give rise to voluntary acts is beyond the scope of the current review. Recent work points to perception-action links and cognitive control as critical components (Jeannerod, 2006; Nattkemper, Ziessler, & Frensch, 2010; Zhang, Hughes, & Rowe, 2012).

Current computational models of speech acquisition provide quantitative hypotheses regarding the roles multi-modal learning and representations play in vocal control and production (Kroger, Kannampuzha, & Neuschaefer-Rube, 2009; Tourville & Guenther, 2011; Westermann & Reck Miranda, 2004). These models can also provide a useful framework for thinking about how learning contributes to vocal imitation and about the form of the representations that make sound imitation possible. For instance, the DIVA model is an adaptive neural network model that can be used to simulate the acquisition of speech by humans (Guenther, 1994, 1995, 2006; Tourville & Guenther, 2011). The key components of this model closely match several mechanisms hypothesized to underlie vocal learning and imitation (Figure 10). In this model, vocal acts generate auditory and tactile feedback that is compared with auditory and somatosensory templates. The outcomes of these comparisons in turn modulate how sounds are produced. The model is adaptive in terms of how heard phonemes become mapped to somatosensory patterns, how somatosensory patterns are mapped to articulatory control, and how sounds are mapped to phonemes.

The DIVA model initially learns to generate pre-specified phonemes based on the results of essentially random babbling followed by specific practice (i.e., no vocal imitation is involved). This learning process can be viewed as a multimodal instantiation of the template model of vocal learning. In the model, babbling corresponds to induced random motions of speech articulators. Acquisition of phoneme production involves finding appropriate parameters to establish desired mappings. Initially, the model learns to map sensed mouth movements to particular articulator movements. Babbled movements produce tactile feedback. This stage basically leads to specific coordinated groupings of articulator movements that generate target tactile patterns. The mapping that the model learns transforms current states into desired states. Mappings from auditory representations to tactile representations are similarly learned so that certain sounds become associated with certain tactile configurations. Essentially, the model learns the different effector positions that are associated with different sounds. Importantly, multiple effector positions can lead to similar sound outputs and the model learns to approximate these many-to-one mappings. The targets for production are thus not a single auditory template for each sound, but a multidimensional space of possible effector configurations (learned from prior production experiences) that lead to that sound. The DIVA model assumes that the speaker (typically construed as a developing child) has a good representation of the sounds that need to be produced prior to vocal learning. However, it is well known that perception of speech sounds is experience dependent and that perceptual learning and speech production learning often occur in parallel. It is likely that perceptual learning gradually refines the target(s) with accumulated experience. The assumption that perception of speech stabilizes before productive learning begins is thus an oversimplification.

Within the DIVA framework, vocal imitation can be described as a process whereby new auditory-tactile targets can be incorporated into a pre-existing vocal control system. The factors that constrain how well the model can imitate a novel target sound include: (1) how the novel sound is represented; (2) the current set of learned tactile configurations; and (3) how closely components of the target sound map onto existing production templates. Ultimately, vocal imitation requires the model to generalize from past learning. However, whenever a new target

Figure 10. Guenther's computational model of speech acquisition (adapted from Tourville and Guenther, 2011; Figure 1). In this model, multimodal maps acquired through experience make it possible for an individual to rapidly learn to reproduce novel sounds.

is added to the repertoire of the model, this will initiate a new wave of adaptive changes to connections in the model that over time can change the model's ability to accurately reproduce both familiar and novel sounds.

A trained DIVA model can rapidly learn to produce new speech sounds based on audio samples provided to it (Tourville & Guenther, 2011. This is possible because the learned maps represent both subcomponents of sounds and combinations of sounds (corresponding to phonemes, syllables, and words). Consequently, novel sounds are essentially indexing combinations of speech motor programs as well as expected somatosensory targets. Feedback is critical in this model for adjusting movements to reduce errors. Feedback is not what makes vocal imitation possible, however. It is instead the incrementally learned mappings based on past auditory, somatosensory, and vocal experiences. Note that in this model there are no specialized "imitation" modules or processors that transform sounds into vocal acts. Rather, it is the adaptive connections between different modalities, as well as the resolution of representations within each of these modality-specific processors that enable the model to imitate sounds. The DIVA model incorporates several features of earlier unimodal models, including error-correction learning mechanisms similar to those of the template model, auditory-to-motor recoding, and specializations for processing speech sounds. Because it is a computational model, it can be used to explore the effects of different experiences on imitative abilities and to generate specific predictions about how different auditory-motor coding schemes might impact an organism's ability to imitate speech sounds.

A limitation of the DIVA model is that it assumes a single target for a given speech sound. Consequently, it would not be able to reproduce the melodic structure of sung speech, nor the individual-specific qualities of a person's voice. It would also not be able to account for the transposition or temporal compression of heard sounds during imitation. Nevertheless, the DIVA model illustrates how vocal imitation abilities can potentially be achieved without any specialized learning mechanisms, and how incremental multimodal learning may be critical to the development of vocal imitation capabilities. The model also highlights the idea that the ability to vocally imitate depends on the flexibility with which sounds are encoded, as well as the capacity to cross-modally associate and monitor dynamic sensorimotor patterns related to vocal control. Because the DIVA model does not include any mechanisms that are unique to humans (other than predefined speech targets), it may be applicable to other species, including non-human primates and cetaceans. The model does not directly account for why vocal imitation abilities are rare among mammals.
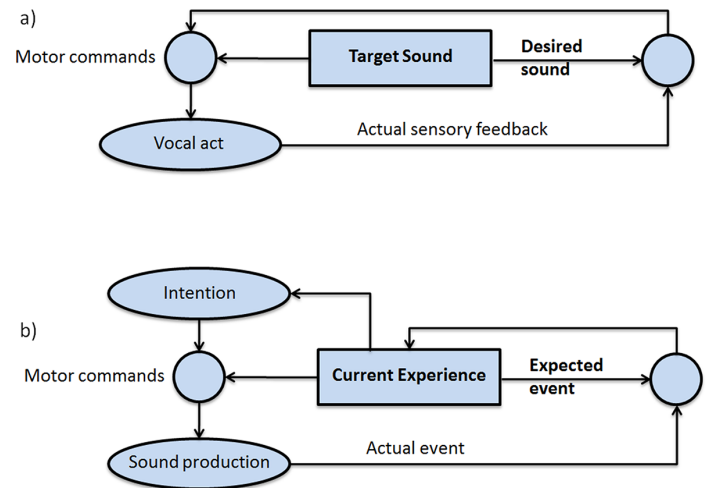

Figure 11. (a) In the standard portrayal of vocal imitation as a learning mechanism, memories of sounds enable an individual to produce somewhat similar sounds that can be compared with the remembered sounds. Differences between the produced and remembered sounds serve as an error signal that is used to adjust future sound production. (b) In a more cognitive characterization of vocal imitation, multimodal representations of ongoing acoustic events (Current Experience) and memories of past events are used to predict future events and to generate and modulate plans for vocal actions, including intentional sound reproduction. In this framework, differences between expected events and perceived events adjust how events are represented.

Synthesis

The key mechanisms postulated in most current models of vocal imitation are auditory representations, motor control systems, a means of comparing past and present representations of sensorimotor events, and error-correction learning (Figure 11a). None of the models explicitly portrays vocal imitation as potentially involving maintenance and recall of past episodes, selective attention, or goal planning, and none distinguishes voluntary imitation from involuntary imitation (Figure 11b illustrates how such processes might be incorporated into a more cognitive model of sound imitation). Theories that describe vocal imitation in the context of acquired multimodal coordination of actions come closest to capturing the complexity of processing typically associated with the voluntary performance of a cognitive skill. These theories suggest that an organism's capacity to imitate sounds is gained through extensive practice producing, feeling, hearing, and recalling different sounds. Although originally developed as models of speech learning and imitation, such multimodal mapping models might be applicable to sound imitation more generally. This could entail introducing multiple, specialized sensorimotor modules that vary across species and/or sound types to account for

differences in the imitatibility of different sounds—such specialized modules might reflect either adaptive specializations or domain-specific customization from prior experiences. It remains to be seen whether any of these models can be modified such that their representations facilitate the perception, prediction, tracking, or coordination of actions.

VII. Conclusions

In the end, the success of any framework for explaining vocal imitation rests less on the terminology it prescribes than on the novel findings that it provides. Our general assessment is that current frameworks that describe vocal imitation as either a specialized communicative learning mechanism or alternatively as an instrumentally conditioned copying response are inadequate for explaining either what vocal imitation entails or its apparent rarity among mammals. For those who may have missed the gist of our argument within the jungle of details provided above, we briefly summarize the main points that led us to this conclusion.

Adult humans vary considerably in their abilities to imitate sounds, both vocally and non-vocally. They imitate speech automatically and unconsciously in contexts that are unlikely to lead to significant learning. They voluntarily imitate singing styles, accents they find amusing, and commercial slogans. Some imitate sounds for a living. Others imitate sounds covertly, including their own vocalizations (e.g., when mentally practicing lines for a play). Human toddlers readily imitate many sounds they hear, both vocally and non-vocally, not all of which are speech sounds. But, despite the frequency with which toddlers copy sounds, their fidelity is poor compared to that of a professional impersonator. This is because the professional has honed his or her imitative skills through extensive practice. Proficient imitation of sounds is a multifaceted skill that arises through learning and that takes much longer to master than the ability to speak. In humans, at least, it is learning that gives rise to vocal imitation abilities. Certainly, the ability to imitate sounds can catalyze communicative learning, but this is just one of many benefits that imitative abilities afford and perhaps not the most important.

Imitation of sounds by cetaceans is not nearly as evident as human vocal imitation. When put to the test, however, dolphins show astounding fidelity in reproducing artificial sounds. While it is true that dolphins do not reproduce speech with the accuracy shown by some birds, it is arbitrary to treat speech as the gold standard of sound imitation. In a competition where dolphins and humans are challenged to reproduce artificial sounds, dolphins would likely outperform many humans. The fact that adult dolphins (and humans) can imitate arbitrary novel sounds when instructed to do so strongly implies they that have

flexible voluntary control of this ability. Is such control necessary for adding new sounds to a vocal repertoire? Would automatic imitation abilities not suffice? These are questions that most current explanatory frameworks cannot readily address, because they make no distinction between voluntary and involuntary sound imitation.

Current ideas about the nature of vocal imitation are largely derivative of hypotheses proposed by Thorndike (1911) over a century ago. Assumptions about the mechanisms underlying sound imitation in mammals have led researchers to underestimate the range of cognitive processes involved. The proposal that vocal imitation only requires comparing percepts of self-produced sounds with memories of previously experienced sounds is inadequate. All mammals that produce sounds can perceive their sound-producing actions through multiple modalities, but most show no capacity to imitate sounds. Auditory feedback can be an important guide to learning, but it is neither necessary nor sufficient for sound imitation. Mismatches between produced sounds and remembered sounds do not automatically lead to changes in sound production. Studies from adult humans suggest that an individual's ability to map perceived sounds onto performable actions, to retain representations of sufficient detail for later reenactments, and to acquire the motor control necessary to flexibly reenact perceived events are key elements of successful sound imitation.

A New View on Vocal Imitation: Imitating Sounds Is a Complex Cognitive Skill

Thorndike had many useful insights about how animals learn and about how to identify limitations in their mental abilities. His ideas about vocal imitation, however, reflect the limited data available at the time. Now, we know much more about the imitative abilities possessed by humans and other animals. The time has come to move beyond Thorndike's (1911, p. 77) idea that vocal imitation abilities are "a specialization removed from the general course of mental development."

We suggest that sound imitation abilities should instead be viewed as a sophisticated skill (or set of skills) that relatively few organisms have the representational or vocal flexibility to master. The limited evidence currently available is consistent with the idea that sound imitation by primates and cetaceans may be mediated by learning, memory, attention, and vocal control mechanisms that involve experience-dependent multimodal representations of events. Multimodal representations appear to play a particularly important role in the sound imitation abilities of adult humans, and may also contribute to sound imitation by non-humans. The availability of such multimodal representations of events can enhance an individual's ability to represent, predict, and

reconstruct perceived events (Wilson & Knoblich, 2005), predict the future actions of conspecifics (Knoblich & Jordan, 2003; Loehr, Kourtis, Vesper, Sebanz, & Knoblich, 2013; Vesper, van der Wel, Knoblich, & Sebanz, 2013), socially communicate with others (Chartrand & Lakin, 2013; Lakin & Chartrand, 2003; Lakin, Chartrand, & Arkin, 2008), and monitor self-produced actions (Wilson, 2001a). An integrative framework in which sound imitation by any mammal is viewed as a skilled performance may provide new insights into the mechanisms underlying this ability.

Viewing voluntary sound imitation as a cognitive skill shifts emphasis away from describing its role in repertoire acquisition and more toward understanding how it fits within the broader domain of cognitive skill learning. Techniques that have been developed to study cognitive skill acquisition in other domains (e.g., comparisons between the performances of experts and amateurs) can potentially be brought to bear in studies of sound imitation. Such approaches, which often focus on individual differences, have rarely been applied in comparative cognition research. Understanding an individual's ability to imitate sounds may require one to relate the individual's perceptual, motor control, memory, selective attention, and conceptual capacities to his or her ability to produce sounds. Vocal imitation abilities may vary within and across species in ways that closely match variations in more basic cognitive capacities, and may reflect global constraints on how learning experiences impact perceptual-motor and cognitive abilities (Mercado, 2008).

Models of the representational processes involved in the control, imagery, and perception of movements (Grossberg & Paine, 2000; Grush, 2004; Hurley, 2008; Jeannerod, 2006) may provide a useful starting point for developing new ways of understanding sound imitation abilities. Most of these models (including the DIVA model described above) were developed to account for human learning and behavior, but they typically do not invoke mechanisms that are specific to humans. Exploring how well existing models can account for sound imitation abilities in different species will be an important step toward identifying the minimal sensorimotor and cognitive requirements for both the automatic and voluntary reproduction of sounds, as well as toward identifying contexts in which either covert imitation or self-imitation are likely to occur.

Vocal imitation is best understood as a subtype of sound imitation, which in turn can be viewed as a representational process that enables an individual to better perceive and predict ongoing actions (of which vocalizations are only a small subset). Imitation of sounds is as true a form of imitation as any based on visual inputs, and no less mysterious. Treating any imitative process as primarily a learning mechanism requires one to ignore any distinction between learning and performance. It would be more accurate to say that imitation depends on an individual's capacity to flexibly represent observed events and to voluntarily control actions, independently of whether the observer attempts to reproduce components of what was observed (as suggested by Bandura, 1986). Unless one considers perception to be synonymous with learning, the framework we propose suggests that sound imitation is a learned process of interpreting the world rather than a specialized mechanism for learning how to produce vocalizations. Studies of adult humans and cetaceans are particularly useful for investigating such processes because these are the only mammals that have consistently shown the ability or motivation to learn to imitate arbitrary sounds in experimental settings.

This account of vocal imitation will remind many readers of the purported role of the "mirror neuron system", in facilitating action understanding (Aziz-Zadeh & Ivry, 2009; Corballis, 2010; Molenberghs, Cunnington, & Mattingley, 2009; Ocampo & Kritikos, 2011). Although we clearly share with this view the sense that the intersection of perception and action plays a critical role in cognition, our perspective differs in several ways from these accounts. First, and most important, we propose that vocal imitation is an ability that is acquired through learning and reflects generalization of learned associations within a complex cognitive architecture. By contrast, the mirror neuron hypothesis, in our view, implies a more hard-wired, modular system. Second, we propose that vocal imitation is a manifestation of the organism's tendency to generate multimodal representations based on associations. Thus we are not suggesting that there is something special about perception/action intersections but rather that intersections may cross multiple perceptual and motor modalities. Finally, a shortcoming we see in the mirror neuron hypothesis is that it is computationally underspecified. This comment is meant more as an observation than a criticism. Mirror neurons emerged from the neuroscience literature and are thus a biological reality. As cognitivists, however, we think it is important to focus on the underlying functional architecture. The mirror neuron hypothesis specifies no such architecture, because the system itself exists at a different level of analysis. Ultimately we find ourselves aligned with the mirror neuron hypothesis in broad philosophical terms, while differing significantly in several important details.

Some Open Questions

When one considers sound imitation freed from the restraints of past assumptions regarding its nature and functions, this can change not only how one describes vocal imitation, but also how it is studied scientifically (Galef,

2013). We end by considering just a few of the many challenging questions raised by this new view of sound imitation.

When does imitative processing contribute to sound perception and production?

Because past criteria for designating phenomena as vocal imitation have been conservative, it is possible that the prevalence of sound imitation abilities in mammals has been underestimated. For instance, a vocalizing cetacean might produce several similar calls in a row. Such repetitive vocal behavior has not been previously viewed as being imitative, but it could involve either deferred imitation of others or self-imitation (Mercado, Murray, Uyeyama, Pack, & Herman, 1998). Piaget (1962) suggested that when infants produce series of similar vocalizations, that these vocal acts should be viewed as self-imitative. Currently, there are no validated, objective metrics for distinguishing an imitative vocal act from a non-imitative one, for distinguishing self-imitation from repetition, or for determining when sounds might be covertly imitated. Consequently, it is difficult to observationally distinguish a series of self-imitated sounds from a series of independently generated sounds with similar features. The extent to which self-imitation contributes to vocal production is not known for any species, including humans, so this possibility cannot yet be excluded. Studies of self-imitation in adults may reveal new ways of distinguishing self-imitative sound production from other non-imitative sound-producing acts, thereby clarifying how often imitative mechanisms are engaged during sound production and perception. When one considers that an imitative vocal act might be deferred for extended periods, the possibility arises that most of the sounds an individual produces might be imitative.

Modern models of sound production posit that many of the mechanisms underlying sound imitation may be automatically engaged every time an organism with the capacity to imitate hears a sound. In that case, observations of overt sound copying would not accurately reflect the frequency of imitative processing. Cases in which individuals use their voice to voluntarily or automatically imitate novel sounds may reflect only a small proportion of instances in which sound imitation skills are engaged during the processing of acoustic events. By analogy, an adult human may think many thoughts each day and yet only occasionally state, "I was just thinking that . . . ." The rarity of such statements does not accurately reflect the frequency of thinking. Representing certain sounds in terms of the vocal or motor acts required to reproduce those sounds may be the default perceptual process rather than a selectively applied approach (Möttönen, Dutton, & Watkins, 2013; Wilson

& Knoblich, 2005; Yuen et al., 2010). Techniques are needed for identifying when auditory perception by both humans and non-humans engages imitative mechanisms.

What determines which acoustic events are imitated?

Mowrer (1960) suggested that the motivation for parrots to imitate human speech was driven by an emotional attachment between a bird and its caretaker and by the frequency with which the bird encountered situations where its needs were not met. This interpretation predicts that differences in "personality" within or across species might correlate with the likelihood that an individual imitates particular sounds, and further suggests that an individual's imitativeness might provide an indirect indicator of his or her emotional state or relationship with a particular individual (Gewirtz & Stingle, 1968). Whether certain emotional states are a prerequisite for overt sound imitation remains unknown. It is interesting to note, however, that in many of the cases in which cetaceans have been observed to imitate environmental sounds (including speech), the imitator has been socially deprived relative to natural situations (Caldwell & Caldwell, 1972; Eaton, 1979; Foote et al., 2006; Ridgway et al., 2012). Miklosi (1999) suggested that many instances of imitation could be related to play behavior (see also Richards, 1986, and Pepperberg, 2005, 2010, for discussions of vocal play in dolphins and parrots). Consistent with this idea, many professional imitators are entertainers. The act of sound imitation (or observations of such acts) might thus sometimes serve to modulate an organism's emotional state, acting as a homeostatic mechanism rather than as a means of learning, localizing, or communicating. It remains unclear how any emotional functions of sound imitation might relate to other potential functions.

We know little about the sound qualities that mammals are most likely to reproduce or about the fidelity with which the most proficient imitators can reproduce these qualities. To reveal the full imitative capacities of non-human mammals may require extensive cognitive and vocal training over several years focusing on sounds that apes or cetaceans find naturally imitatible. If sound imitation in cetaceans or apes depends on comparable mechanisms to those used by humans, then it should be possible to train individuals to flexibly imitate a wide range of sounds or to specialize in imitating certain classes of sounds. Identifying the upper limits of imitative capacity in cetaceans may yield new insights into the constraints that prevent other mammals from learning to imitate sounds. If higher fidelity imitative representations enhance remote action monitoring, as suggested above, then this predicts that audiospatial sequences that correspond to natural events should be

more imitatible than randomly ordered sequences. To date, researchers have only explored the ability of non-human mammals to imitate individual sounds or repetitions of the same sound; essentially nothing is known about how the imitatibility of sound sequences varies across species.

What determines an individual's imitative capacity?

There is considerable debate about which if any imitative abilities of humans are genetically determined (Jones, 2007, 2009; Parton, 1976). Recent work tends to side with Piaget's (1962) conclusion that there is not a single specialized mechanism that gives rise to vocal imitation (or any other kind of imitation), but that a variety of perceptual, motor, and cognitive abilities that emerge sequentially during development contribute to the acquisition of vocal imitation abilities (Jones, 2007). Interestingly, the earliest imitative acts noted by Piaget (1962) and other developmental psychologists often have involved sound production (Jones, 2006, 2007), suggesting that sound plays a particularly important role in the development of imitative abilities. Furthermore, the first sounds imitated by infants are in some cases not sounds produced vocally by other humans, but are instead novel environmental sounds or percussive sounds (Piaget, 1962; E. Mercado, personal observation), contrary to what one might expect if sound imitation serves primarily as an adaptation for speech acquisition. Detailed investigations of the kinds of sounds naturally reproduced by humans and other animals at an early age, as well as systematic analyses of how imitative capacities vary across individuals, may help researchers to identify the genetic, neural, or experiential variables that impact an organism's capacity and tendency to imitate particular sound features. For instance, a recent study of expert phoneticians (who specialize in transcribing speech) identified both anatomical predispositions and acquired brain morphology that were correlated with an individual's ability to transcribe novel phonetic contrasts (Golestani, Price, & Scott, 2011).

As noted earlier, several researchers have suggested that the rarity of vocal imitation abilities in mammals reflects limitations in vocal control (Deacon, 1997; Fitch, 2010; Mowrer, 1960). Jarvis has made this argument most strongly and convincingly (Arriaga & Jarvis, 2013; Jarvis, 2004, 2013). Undoubtedly, variations in imitative capacity across species and individuals reflect differences in neural architecture, including differences in circuits involved in the control of vocal production. However, the brains of different individuals vary in many ways, and scientific history is replete with premature identifications of "brain differences that make the difference." To date, neuroscientific studies of vocal imitation have been designed primarily to identify neural substrates that instantiate the classic template matching model of vocal learning. To the extent that this model is inadequate for explaining sound imitation by adult mammals, research aimed at revealing how neural circuits implement template-based vocal learning will be similarly inadequate for understanding how and why mammals imitate sounds.

The flexibility with which dolphins and humans can reproduce sounds may reflect their general cognitive abilities rather than any specialized mechanisms of vocal control. Music cognition researchers often suggest that musicians can flexibly manipulate sophisticated representations of sound streams and associated visual or motor events, and that acquired musical concepts constrain a musician's performances. Dolphins have shown the ability to explicitly access representations of past events (Mercado, Uyeyama, Pack, & Herman, 1999), to form and use abstract concepts about such events, and to actively use sounds (and memories of sounds) to guide their actions (reviewed by Mercado & DeLong, 2010). The extent to which available concepts and memory mechanisms constrain sound imitation abilities has seldom been considered in past comparative studies. Future efforts to characterize and understand the nature and functions of sound imitation in mammals can benefit from integrated approaches that more fully consider the representational processes cetaceans and primates may bring to bear when imitating both familiar and novel sounds.

In this review, a cognitive approach to understanding sound imitation was presented as an alternative to the possibility that vocal imitation serves primarily as a mechanism for learning to produce novel sounds. The shift from a social communicative learning model of vocal imitation to a cognitive skill oriented model leads to novel hypotheses about cross-species commonalities in representational and perceptual processes and to new avenues for theoretical integration of comparative bioacoustic studies with studies of human auditory cognition. Experimental investigations of automatic imitation, individual differences in imitative capacity, and correlations between imitative fidelity and spatial acuity or auditory working memory capacity in non-human animals may reveal unsuspected similarities (or differences) in the imitative skills of primates and cetaceans.

Term	Description
vocal motor system	A collection of neural control circuits and associated muscles whose coordinated actions serve to produce sounds. This system level subsumes the standard tripartite division of vocal production into respiration, phonation, and articulation.
vocal learning	Modifying vocal output based on auditory experience, which may be short- or long-term in its effect. In some theories, vocal imitation is described as a type of vocal learning. In our view, these are separate phenomena and may share a reciprocal relationship, or may be unrelated.
vocal imitation	Imitation of sound using the vocal effector system. From a cognitive perspective, this process can be viewed as the vocal reenactment of previously experienced auditory events.
vocal contagion	Production of sound based on assimilation to present context that does not involve matching a specific stimulus, but does involve reproducing sounds that are predominant at the present point.
vocal performance	The outcomes of vocal motor actions. Performance is neutral with respect to whether it reflects learning, imitation, or anything else. The term ultimately refers to events generated by the vocal motor system.
cognitive skill	An ability, such as reading, that can be improved through practice and that involves judgments or processing beyond what is involved in performing perceptual-motor responses.
sound imitation	A process by which an individual either vocally or non-vocally generates sounds with qualities that reproduce elements of previously experienced sounds.
imitatible sound	Sound for which movements can be produced that map onto the form of the sound, even if the mapping is imperfect and only applies to certain sound features.
ranging hypothesis	A proposed explanation for why animals copy vocalizations that suggests that internal representations of familiar sound features can enhance an individual's ability to estimate the distance a sound has traveled.
template matching	A proposed mechanism for vocal learning and imitation based on the idea that the imitator matches auditory feedback from a vocalization to a memory representation of the target sound (the "template").
auditory-motor recoding	A proposed mechanism for vocal imitation based on the assumption that the motor system plays an active role in perception. In so doing, the motor system automatically processes auditory events with respect to the motor gestures needed to produce that event. A closely related construct in the cognitive literature is the notion of embodied cognition.
multimodal mapping	A theoretical proposal, applicable but not limited to vocal imitation, that all perceptual inputs are processed across multiple modalities.
auditory imagery	Mental simulation of the experience of auditory events, in the absence of the appropriate distal stimulus. Unlike auditory hallucinations, auditory imagery is phenomenally experienced as a mental event, and is not confused with a veridical experience.

References

Abrego-Collier, C., Grove, J., & Sonderegger, M. (2011). *Effects of speaker evaluation on phonetic convergence.* Paper presented at the Proceedings of the 17th International Congress of Phonetic Science.

Adret, P. (1993). Vocal learning induced with operant techniques: An overview. *Netherlands Journal of Zoology, 43,* 125–142.

Agafonov, A. V., & Panova, E. M. (2012). Individual patterns of tonal (whistling) signals of bottlenose dolphins (*Tursiops truncates*) kept in relative isolation. *Biology Bulletin, 39*, 430–440. doi: 10.1134/S1062359012050020

Akcay, C., Tom, M. E., Campbell, S. E., & Beecher, M. D. (2013). Song type matching is an honest early threat signal in a hierarchical animal communication system. *Proceedings of the Royal Society B - Biological Sciences, 280*, 20122517. doi: 10.1098/rspb.2012.2517

Amin, T. B., Marziliano, P., & German, J. S. (2012). *Nine voices, one artist: Linguistic and acoustic analysis.* Paper presented at the 2012 IEEE International Conference on Multimedia and Expo, Melbourne, Australia.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369–406. doi: 10.1037//0033-295X.89.4.369

Andrew, R. J. (1962). Evolution of intelligence and vocal mimicking: Studies of large-brained mammals promise to elucidate some problems of human evolution. *Science, 137*, 585–589. doi: 10.1126/science.137.3530.585

Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language, 124*, 96–116. doi: 10.1016/j.bandl.2012.10.002

Au, W. W. L. (1993). *The sonar of dolphins.* New York: Springer-Verlag.

Aziz-Zadeh, L., & Ivry, R. B. (2009). The human mirror neuron system and embodied representations. *Advances in Experimental Medicine and Biology, 629,* 355–376. doi: 10.1007/978-0-387-77064-2_18

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*, 177–189. doi: 10.1016/j.wocn.2011.09.001

Baer, D. M., & Deguchi, H. (1985). Generalized imitation from a radical-behavioral viewpoint. In S. Reiss & R. R. Bootzin (Eds.), *Theoretical issues in behavior therapy* (pp. 179–217). Orlando: Academic Press.

Baer, D. M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior, 10*, 405–416. doi: 10.1901/jeab.1967.10-405

Baily, J. (1985). Musical structure and human movement. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 237–258). London: Academic Press.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Baylis, J. R. (1982). Avian vocal mimicry: Its function and evolution. In D. E. Kroodsma (Ed.), *Acoustic communication in birds, vol. 2: Song learning and its consequences* (pp. 51–83). New York: Academic Press.

Beaman, C. P., & Williams, T. I. (2010). Earworms (stuck song syndrome): Towards a natural history of intrusive thoughts. *British Journal of Psychology, 101*, 637–653. doi: 10.1348/000712609X479636

Beckers, G. J., Nelson, B. S., & Suthers, R. A. (2004). Vocal-tract filtering by lingual articulation in a parrot. *Current Biology, 14*, 1592–1597. doi: 10.1016/j.cub.2004.08.057

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*, 47–73. doi: 10.1146/annurev.psych.57.102904.190152

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization.* Cambridge, MA: MIT Press.

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience, 11*, 747–759. doi: 10.1038/nrn2931

Byrne, R. W. (2002). Imitation of novel complex actions: What does the evidence from animals mean? *Advances in the Study of Behavior, 31,* 77–105. doi: 10.1016/S0065-3454(02)80006-7

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences, 21*, 667–684. doi: 10.1017/S0140525X98001745

Caldwell, M. C., & Caldwell, D. K. (1965). Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*). *Nature, 207*, 434–435. doi: 10.1038/207434a0

Caldwell, M. C., & Caldwell, D. K. (1972). Vocal mimicry in the whistle mode by an Atlantic bottlenosed dolphin. *Cetology, 9*, 1–8.

Caldwell, M. C., & Caldwell, D. K. (1979). The whistle of the Atlantic bottlenosed dolphin (*Tursiops truncatus*) - Ontogeny. In H. E. Winn & B. L. Olla (Eds.), *Behavior of marine animals: Current perspective in research. Vol 3. Cetaceans* (pp. 369–401). New York: Plenum Press.

Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology, 16*, 69–74. doi: 10.1016/j.cub.2005.10.071

Cazau, D., Adam, O., Laitman, J. T., & Reidenberg, J. S. (2013). Understanding the intentional acoustic behavior of humpback whales: A production-based approach. *Journal of the Acoustical Society of America, 134*, 2268–2273. doi: 10.1121/1.4816403

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology, 64*, 285-308. doi: 10.1146/annurev-psych-113011-143754

Chmelnitsky, E. G., & Ferguson, S. H. (2012). Beluga whale, *Delphinapterus leucas*, vocalizations from the Churchill River, Manitoba, Canada. *Journal of the Acoustical Society of America, 131*, 4821–4835. doi: 10.1121/1.4707501

Cholewiak, D. M. (2008). *Evaluating the role of song in the humpback whale (Megaptera novaeangliae) breeding system with respect to intra-sexual interactions.* Doctoral dissertation, Cornell University.

Clark, C. W. (1990). Acoustic behavior of mysticete whales. In J. A. Thomas & R. Kastelein (Eds.), *Sensory abilities of cetaceans: Laboratory and field evidence* (pp. 571–583). New York: Plenum.

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception, 10*, 317–341.

Clarke, E. F., & Baker-Short, C. (1987). The imitation of perceived rubato: A preliminary study. *Psychology of Music, 15*, 58–75. doi: 10.1177/0305735687151005

Coleman, P. D. (1962). Failure to localize the source distance of an unfamiliar sound. *Journal of the Acoustical Society of America, 34,* 345–346. doi: 10.1121/1.1928121

Connor, R. C. (2000). Group living in whales and dolphins. In J. Mann, R. C. Connor, P. L. Tyack, & H. Whitehead (Eds.), *Cetacean societies: Field studies of dolphins and whales* (pp. 199–218). Chicago: University of Chicago Press.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain and Language, 112*, 23–35. doi: 10.1016/j.bandl.2009.02.002

Cowan, N., Braine, M. D. S., & Leavitt, L. A. (1985). The phonological and metaphonological representation of speech: Evidence from fluent backward talkers. *Journal of Memory and Language, 24*, 679–698. doi: 10.1016/0749-596X(85)90053-1

Cranford, T. W., Elsberry, W. R., Van Bonn, W. G., Jeffress, J. A., Chaplin, M. S., Blackwood, D. J., . . . Ridgway, S. H. (2011). Observation and analysis of sonar signal generation in the bottlenose dolphin (Tursiops truncatus): Evidence for two sonar sources. *Journal of Experimental Marine Biology and Ecology, 407*, 81–96. doi: 10.1016/j.jembe.2011.07.010

Crowell, S., Harley, H. E., Fellner, W., & Larsen-Plott, L. (2005). *Vocal productions of rhythms by the bottlenose dolphin.* Paper presented at the 16th Biennial Conference on the Biology of Marine Mammals, San Diego, CA.

d'Alessandro, C., Rilliard, A., & Le Beux, S. (2011). Chironomic stylization of intonation. *Journal of the Acoustical Society of America, 129*, 1594–1604. doi: 10.1121/1.3531802

Dalla Bella, S., Giguere, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America, 121*, 1182–1189. doi: 10.1121/1.2427111

Dalziell, A. H., & Magrath, R. D. (2012). Fooling the experts: Accurate vocal mimicry in the song of the superb lyrebird, *Menura novaehollandiae*. *Animal Behaviour, 83*, 1401–1410.

Darling, J. D., Jones, M. E., & Nicklin, C. P. (2012). Humpback whale (*Megaptera novaeangliae*) singers in Hawaii are attracted to playback of similar song (L). *Journal of the Acoustical Society of America, 132*, 2955–2958. doi: 10.1121/1.4757739

Darling, J. D., Meagan, E., & Nicklin, C. P. (2006). Humpback whale songs: Do they organize males during the breeding season? *Behaviour, 143*, 1051–1101. doi: 10.1163/156853906778607381

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: W. W. Norton.

Deecke, V. B. (1998). *Stability and change of killer whale (Orcinus orca) dialects*. Masters thesis, University of British Columbia.

Deecke, V. B., Ford, J. K., & Spong, P. (2000). Dialect change in resident killer whales: Implications for vocal learning and cultural transmission. *Animal Behaviour, 60*, 629–638. doi: 10.1006/anbe.2000.1454

DeRuiter, S. L., Boyd, I. L., Claridge, D. E., Clark, C. W., Gagnon, C., Southall, B. L., & Tyack, P. L. (2013). Delphinid whistle production and call matching during playback of simulated military sonar. *Marine Mammal Science, 29*, E46-E59. doi: 10.1111/j.1748-7692.2012.00587.x

Domjan, M. (2000). Learning: Overview. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 5, pp. 1–3). New York: Oxford University Press.

Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*, 567–631. doi: 10.1146/annurev.neuro.22.1.567

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics, 53*, 25–33. doi: 10.3758/BF03211712

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition, 74*, 1–32. doi: 10.1016/S0010-0277(99)00061-X

Eaton, R. L. (1979). A beluga imitates human speech. *Carnivore, 2*, 22–23.

Edds-Walton, P. L. (1997). Acoustic communication signals of mysticete whales. *Bioacoustics, 8*, 47–60. doi: 10.1080/09524622.1997.9753353

Egnor, S. E., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences, 27*, 649–654. doi: 10.1016/j.tins.2004.08.009

Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders, 5*, 681–691. doi: 10.1016/j.rasd.2010.09.001

Eriksson, A. (2010). The disguised voice: Imitating accents of speech styles and impersonating individuals. In C. Llamas (Ed.), *Language and identities* (pp. 86–98). Edinburgh: Edinburgh University Press.

Eriksson, A., & Wretling, P. (1997). *How flexible is the human voice? A case study of mimicry*. Paper presented at the Fifth European Conference on Speech Communication and Technology.

Falls, J. B., & Brooks, R. J. (1975). Individual recognition by song in white-throated sparrows. II. Effects of location. *Canadian Journal of Zoology, 53*, 1412–1420. doi: 10.1139/z75-170

Fay, W. H. (1969). On the basis of autistic echolalia. *Journal of Communication Disorders, 2*, 38–47. doi: 10.1016/0021-9924(69)90053-7

Fay, W. H., & Coleman, R. O. (1977). A human sound transducer/reproducer: Temporal capabilities of a profoundly echolalic child. *Brain and Language, 4*, 396–402. doi: 10.1016/0093-934X(77)90034-7

Fellner, W., Bauer, G. B., & Harley, H. E. (2006). Cognitive implications of synchrony in dolphins: A review. *Aquatic Mammals, 32*, 511–516.

Filatova, O. A., Burdin, A. M., & Hoyt, E. (2010). Horizontal transmission of vocal traditions in killer whale (*Orcinus orca*) dialects. *Biology Bulletin, 37*, 965–971. doi: 10.1134/S1062359010090104

Filatova, O. A., Deecke, V. B., Ford, J. K. B., Matkin, C. O., Barrett-Lennard, L. G., Guzeev, M. A., . . . Hoyt, E. (2012). Call diversity in the North Pacific killer whale populations: Implications for dialect evolution and population history. *Animal Behavior, 83*, 595–603. doi: 10.1016/j.anbehav.2011.12.013

Fish, M. P., & Mowbray, W. H. (1962). Production of underwater sound by the white whale or beluga *Delphinapterus leucas* (Pallas). *Journal of Marine Research, 20*, 149–161.

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, MA: MIT Press.

Foote, A. D., Griffin, R. M., Howitt, D., Larsson, L., Miller, P. J. O., & Hoelzel, A. R. (2006). Killer whales are capable of vocal learning. *Biology Letters, 2*, 509–512. doi: 10.1098/rsbl.2006.0525

Ford, J. K. B. (1989). Acoustic behavior of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia, Canada. *Canadian Journal of Zoology, 67*, 727–745.

Ford, J. K. B. (1991). Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia. *Canadian Journal of Zoology, 69*, 1454–1483. doi: 10.1139/z91-206

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*, 396–413. doi: 10.1016/S0749-596X(03)00072-X

Frazer, L. N., & Mercado, E., III. (2000). A sonar model for humpback whale song. *IEEE Journal of Oceanic Engineering, 25*, 160–182. doi: 10.1109/48.820748

Fripp, D., Owen, C., Quintana-Rizzo, E., Shapiro, A., Buckstaff, K., Jankowski, K., . . . Tyack, P. (2005). Bottlenose dolphin (*Tursiops truncatus*) calves appear to model their signature whistles on the signature whistles of community members. *Animal Cognition, 8*, 17–26. doi: 10.1007/s10071-004-0225-z

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*, 361–377. doi: 10.3758/BF03193990

Galef, B. G. (1988). Imitation in animals: History, definition, and interpretation of data from the psychological laboratory. In T. R. Zentall & B. G. Galef (Eds.), *Social learning: Psychological and biological perspectives* (pp. 3–28). Hillsdale, NJ: Lawrence Erlbaum Associates.

Galef, B. G. (2013). Imitation and local enhancement: Detrimental effects of consensus definitions on analyses of social learning in animals. *Behavioural Processes, 100*, 123–130. doi: 10.1207/s15327604jaws0204_2

Garamszegi, L. Z., Eens, M., Pavlova, D. Z., Aviles, J., & Moller, A. P. (2007). A comparative study of the function of heterospecific mimicry in European passerines. *Behavioral Ecology, 18*, 1001–1009. doi: 10.1093/beheco/arm069

Garcia, E., Baer, D. M., & Firestone, I. (1971). The development of generalized imitation within topographically determined boundaries. *Journal of Applied Behavioral Analysis, 4*, 101–112.

Gardner, M. B. (1969). Distance estimation of 0 degrees or apparent 0 degrees-oriented speech signals in anechoic space. *Journal of the Acoustical Society of America, 45*, 47–53. doi: 10.1121/1.1911372

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science, 1*, 292–304. doi: 10.1111/j.1756-8765.2009.01020.x

Gewirtz, J. L., & Stingle, K. G. (1968). Learning of generalized imitation as the basis for identification. *Psychological Review, 75*, 374–397. doi: 10.1037/h0026378

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279. doi: 10.1037//0033-295X.105.2.251

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review, 11*, 716–722. doi: 10.3758/BF03196625

Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *Journal of Neuroscience, 31*, 4213-4220. doi: 10.1523/JNEUROSCI.3891-10.2011

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language, 109*, 55–67. doi: 10.1016/j.bandl.2008.01.005

Gordon, E. E. (2007). *Learning sequences of music.* Chicago: GIA Publications.

Grebner, D. M., Parks, S. E., Bradley, D. L., Miksis-Olds, J. L., Capone, D. E., & Ford, J. K. (2011). Divergence of a stereotyped call in northern resident killer whales. *Journal of the Acoustical Society of America, 129*, 1067–1072. doi: 10.1121/1.3531842

Green, G. A. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *Journal of Research in Music Education, 38*, 225–231. doi: 10.2307/3345186

Gregory, S. W., Jr., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology, 70*, 1231–1240. doi: 10.1037/0022-3514.70.6.1231

Grossberg, S., & Paine, R. W. (2000). A neural model of corticocerebellar interactions during attentive imitation and predictive learning of sequential handwriting movements. *Neural Networks, 13*, 999-1046. doi: 10.1016/S0893-6080(00)00065-4

Grossi, D., Marcone, R., Cinquegrana, T., & Gallucci, M. (2012). On the differential nature of induced and incidental echolalia in autism. *Journal of Intellectual Disability Research*. doi: 10.1111/j.1365-2788.2012.01579.x

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences, 27*, 377–396.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics, 72*, 43–53. doi: 10.1007/BF00206237

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102*, 594–621. doi: 10.1037/0033-295X.102.3.594

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39*, 350–365. doi: 10.1016/j.jcomdis.2006.06.013

Guinee, L. N., Chu, K., & Dorsey, E. M. (1983). Changes over time in the songs of known individual humpback whales (*Megaptera novaeangliae*). In R. Payne (Ed.), *Communication and behavior of whales* (pp. 59–80). Boulder, CO: Westview Press.

Halpern, A. R., & Bartlett, J. C. (2011). The persistence of musical memories: A descriptive study of earworms. *Music Perception, 28*, 425–432. doi: 10.1525/MP.2011.28.1.425

Harley, H. E. (2008). Whistle discrimination and categorization by the Atlantic bottlenose dolphin (*Tursiops truncatus*): A review of the signature whistle framework and a perceptual test. *Behavioural Processes, 77*, 243–268. doi: 10.1016/j.beproc.2007.11.002

Hauser, M. D. (2009). The illusion of biological variation: A minimalist approach to the mind. In M. Piatelli-Pamarini, J. Uriagereka, & P. Salaburu (Eds.), *Of minds and language: A Dialogue with Noam Chomsky in the Basque Country* (pp. 299–328). Oxford, UK: Oxford University Press.

Herman, L. M. (1980). Cognitive characteristics of dolphins. In L. M. Herman (Ed.), *Cetacean behavior: Mechanisms and functions* (pp. 363–429). New York: Wiley Interscience.

Herman, L. M. (2002). Vocal, social, and self-imitation by bottlenosed dolphins. In K. Dautenhahn & C. Nehaniv (Eds.), *Imitation in animals and artifacts* (pp. 63–108). Cambridge, MA: MIT Press.

Herman, L. M., & Tavolga, W. N. (1980). The communication systems of cetaceans. In L. M. Herman (Ed.), *Cetacean behavior: Mechanisms and functions* (pp. 149–209). New York: Wiley Interscience.

Heyes, C. M. (1994). Social learning in animals: Categories and mechanisms. *Biological Reviews, 69*, 207–231. doi: 10.1111/j.1469-185X.1994.tb01506.x

Heyes, C. M. (1996). Genuine imitation. In C. Heyes & B. G. Galef (Eds.), *Social learning in animals: The roots of culture* (pp. 371–389). New York: Academic Press.

Heyes, C. M. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463–483. doi: 10.1037/a0022288 2011-01604-001

Hoelzel, A. R., & Osborne, R. (1986). Killer whale call characteristics: Implications for cooperative foraging strategies. In B. C. Kirkevold & J. S. Lockard (Eds.), *Behavioral biology of killer whales* (pp. 373–403). New York: A. R. Liss.

Honorof, D. N., Weihing, J., & Fowler, C. A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics, 39*, 18–38. doi: 10.1016/j.wocn.2010.10.007

Hooper, S., Reiss, D., Carter, M., & McCowan, B. (2006). Importance of contextual saliency on vocal imitation by bottlenose dolphins. *International Journal of Comparative Psychology, 19*, 116–128.

Hopkins, W. D., Taglialatela, J., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour, 73*, 281–286. doi: 10.1016/j.anbehav.2006.08.004

Hoppe, D., Sadakate, M., & Desain, P. (2006). Development of real-time visual feedback assistance in singing training: A review. *Journal of Computer Assisted Learning, 22*, 308–316. doi: 10.1111/j.1365-2729.2006.00178.x

Hurley, S. (2008). The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences, 31*, 1–22. doi: 10.1017/S0140525X07003123

Hutchins, S., & Peretz, I. (2012). Amusics can imitate what they cannot discriminate. *Brain and Language, 123*, 234–239. doi: 10.1016/j.bandl.2012.09.011

Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *Journal of the Acoustical Society of America, 127*, 504–512. doi: 10.1121/1.3270391

Immelmann, K., & Beer, C. (1989). *A dictionary of ethology.* Cambridge, MA: Harvard University Press.

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition, 15*, 434–435. doi: 10.1017/S1366728912000041

Jaakkola, K., Guarino, E., & Rodriguez, M. (2010). Blindfolded imitation in a bottlenose dolphin (*Tursiops truncatus*). *International Journal of Comparative Psychology, 23*, 671–688.

James, W. (1890). *The principles of psychology.* New York, NY: Dover.

Janik, V. M. (1999). Origins and implications of vocal learning in bottlenose dolphins. In H. O. Box & K. R. Gibson (Eds.), *Mammalian social learning: Comparative and ecological perspectives* (pp. 308–326). Cambridge: Cambridge University Press.

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science, 289*, 1355–1357. doi: 10.1126/science.289.5483.1355

Janik, V. M. (2009a). Acoustic communication in delphinids. *Advances in the Study of Behavior, 40*, 123–158. doi: 10.1016/S0065-3454(09)40004-4

Janik, V. M. (2009b). Whale song. *Current Biology, 19*, R109–111. doi: 10.1016/j.cub.2008.11.026

Janik, V. M., & Sayigh, L. S. (2013). Communication in bottlenose dolphins: 50 years of signature whistle research. *Journal of Comparative Physiology A, 199*, 479–489. doi: 10.1007/s00359-013-0817-7

Janik, V. M., Sayigh, L. S., & Wells, R. S. (2006). Signature whistle shape conveys identity information to bottlenose dolphins. *Proceedings of the National Academy of Sciences, USA, 103*, 8293–8297. doi: 10.1073/pnas.0509918103

Janik, V. M., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior, 26*, 59–99. doi: 10.1016/S0065-3454(08)60377-0

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour, 60*, 1–11. doi: 10.1006/anbe.2000.1410

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences, 1016,* 749–777. doi: 10.1196/annals.1298.038 1016/1/749

Jarvis, E. D. (2013). Evolution of brain pathways for vocal learning in birds and humans. In J. J. Bolhuis & M. Everaert (Eds.), *Birdsong, speech, and language: Exploring the evolution of mind and brain* (pp. 63–107). Cambridge, MA: MIT.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self.* Oxford: Oxford University Press.

Johnson, H. M. (1912). The talking dog. *Science, 35,* 749–751. doi: 10.1126/science.35.906.749

Jones, S. S. (2006). Infants learn to imitate by being imitated. In C. Yu, L. B. Smith & O. Sporns (Eds.), *Proceedings of the International Conference on Development and Learning*. Bloomington, IN: Indiana University.

Jones, S. S. (2007). Imitation in infancy: The development of mimicry. *Psychological Science, 18,* 593-599. doi: 10.1111/j.1467-9280.2007.01945.x

Kappes, J., Baumgaertner, A., Peschke, C., & Ziegler, W. (2009). Unintended imitation in nonword repetition. *Brain and Language, 111,* 140–151. doi: 10.1016/j.bandl.2009.08.008

Karlsen, J. D., Bisther, A., Lydersen, C., Haug, T., & Kovacs, K. M. (2002). Summer vocalisations of adult male white whales (*Delphinapterus leucas*) in Svalbard, Norway. *Polar Biology, 25,* 808–817. doi: 10.1007/s00300-002-0415-6

Kelley, L. A., & Healy, S. D. (2010). Vocal mimicry in male bowerbirds: Who learns from whom? *Biology Letters, 6,* 626–629. doi: 10.1098/rsbl.2010.0093

Kelley, L. A., & Healy, S. D. (2011). Vocal mimicry. *Current Biology, 21*(1), R9–10. doi: 10.1016/j.cub.2010.11.026

Killebrew, D. A., Mercado, E., III, Herman, L. M., & Pack, A. A. (2001). Sound production of a neonate bottlenose dolphin. *Aquatic Mammals, 27,* 34–44.

King, S. L., Sayigh, L. S., Wells, R. S., Fellner, W., & Janik, V. M. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society B - Biological Sciences, 280*, 20130053. doi: 10.1098/rspb.2013.0053

Knoblich, G., & Jordan, J. S. (2003). Action coordination in groups and individuals: Learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1006-1016. doi: 10.1037/0278-7393.29.5.1006

Koda, H., Oyakawa, C., Kato, A., & Masataka, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour, 144*, 681–692. doi: 10.1163/156853907781347817

Kojima, S. (2003). *A search for the origins of human speech: Auditory and vocal functions of the chimpanzee.* Victoria, Australia: Trans Pacific Press.

Konishi, M. (1965). The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Zeitschrift fur Tierpsychologie, 22*, 770–783.

Kremers, D., Jaramillo, M. B., Boye, M., Lemasson, A., & Hausberger, M. (2011). Do dolphins rehearse show-stimuli when at rest? Delayed matching of auditory memory. *Frontiers in Psychology, 2*, 386. doi: 10.3389/fpsyg.2011.00386

Kremers, D., Lemasson, A., Almunia, J., & Wanker, R. (2012). Vocal sharing and individual acoustic distinctiveness within a group of captive orcas (*Orcinus orca*). *Journal of Comparative Psychology, 126*, 433–445. doi: 10.1037/a0028858

Kroger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C. (2009). Towards a neurocomputational model of speech production and perception. *Speech Communication, 51*, 793–809. doi: 10.1016/j.specom.2008.08.002

Kuczaj, S. A., II, & Yeater, D. B. (2006). Dolphin imitation: Who, what, when, and why. *Aquatic Mammals, 32*, 413–422. doi: 10.1578/AM.32.4.2006.413

Kymissis, E., & Poulson, C. L. (1990). The history of imitation in learning theory: The language acquisition process. *Journal of the Experimental Analysis of Behavior, 54*, 113–127. doi: 10.1901/jeab.1990.54-113

Lachlan, R. F., & Slater, P. J. B. (1999). The maintenance of vocal learning by gene-culture interaction: The cultural trap hypothesis. *Proceedings of the Royal Society B - Biological Sciences, 266*, 701–706.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science, 14*, 334-339. doi: 10.1111/1467-9280.14481

Lakin, J. L., Chartrand, T. L., & Arkin, R. M. (2008). I am too just like you: Nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychological Science, 19*, 816-822. doi: 10.1111/j.1467-9280.2008.02162.x

Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H., . . . Wich, S. A. (2013). Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *Journal of the Acoustical Society of America, 134*, 2326–2335. doi: 10.1121/1.4817929

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior and Development, 13*, 343–354. doi: 10.1016/0163-6383(90)90039-B

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*, 78–106. doi: 10.1016/0010-0285(82)90005-6

Lévêque, Y., Giovanni, A., & Schön, D. (2012). Pitch-matching in poor singers: Human model advantage. *Journal of Voice, 26*, 293–298. doi: 10.1016/j.jvoice.2011.04.001

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1–36. doi: 10.1016/0010-0277(85)90021-6

Lieberman, P. (2012). Vocal tract anatomy and the neural bases of talking. *Journal of Phonetics, 40*, 608–622. doi: 10.1016/j.wocn.2012.04.001

Lilly, J. C. (1958). Some considerations regarding basic mechanisms of positive and negative types of motivation. *American Journal of Psychiatry, 115*, 498–504.

Lilly, J. C. (1961). *Man and dolphin.* New York: Doubleday.

Lilly, J. C. (1963). Productive and creative research with man and dolphin. *Archives of General Psychiatry, 8*, 111–116.

Lilly, J. C. (1965). Vocal mimicry in *Tursiops*: Ability to match numbers and durations of human vocal bursts. *Science, 147*, 300–301. doi: 10.1126/science.147.3655.300

Lilly, J. C. (1967). Dolphin's mimicry as a unique ability and a step towards understanding. In K. Salzinger & S. Salzinger (Eds.), *Research in verbal behavior and some neurophysiological implications* (pp. 21–27). New York: Academic Press.

Lilly, J. C. (1968). Sound production in *Tursiops truncatus* (bottlenose dolphin). *Annals of the New York Academy of Sciences, 155*, 321–341. doi: 10.1111/j.1749-6632.1968.tb56778.x

Lilly, J. C., Miller, A. M., & Truby, H. M. (1968). Reprogramming of the sonic output of the dolphin: Sonic burst count matching. *Journal of the Acoustical Society of America, 43*, 1412–1424. doi: 10.1121/1.1911001

Lim, S. J., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science, 35*, 1390–1405. doi: 10.1111/j.1551-6709.2011.01192.x

Lindbolm, B. (1996). Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America, 99*, 1683–1692. doi: 10.1121/1.414691

Lipkind, D., Marcus, G. F., Bemis, D. K., Sasahara, K., Jacoby, N., Takahasi, M., . . . Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature, 498*, 104–108. doi: 10.1038/nature12173

Little, A. D., Mershon, D. H., & Cox, P. H. (1992). Spectral content as a cue to perceived auditory distance. *Perception, 21*, 405–416. doi: 10.1068/p210405

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn_a_00388

Madsen, P. T., Jensen, F. H., Carder, D., & Ridgway, S. (2012). Dolphin whistles: A functional misnomer revealed by heliox breathing. *Biology Letters, 8*, 211–213. doi: 10.1098/rsbl.2011.0701

Majewski, W., & Staroniewicz, P. (2011). Imitation of target speakers by different types of impersonators. In A. Esposito, A. Vinviarelli, K. Vicsi, C. Pelachaud, & A. Nijholt (Eds.), *Analysis of verbal and nonverbal communication and enactment: The processing issues* (Vol. 6800, pp. 104–112). Berlin: Springer.

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition, 127*, 177–202. doi: 10.1016/j.cognition.2012.12.008

Margoliash, D. (2002). Evaluating theories of bird song learning: Implications for future directions. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 188*, 851–866. doi: 10.1007/s00359-002-0351-5

Marino, L., Connor, R. C., Fordyce, R. E., Herman, L. M., Hof, P. R., Lefebvre, L., . . . Whitehead, H. (2007). Cetaceans have complex brains for complex cognition. *PLoS Biology, 5*, 966–972. doi: 10.1371/journal.pbio.0050139

Marler, P. (1970). Birdsong and speech development: Could there be parallels? *American Scientist, 58*, 669–673.

Marler, P. (1976a). An ethological theory of the origin of vocal learning. *Annals of the New York Academy of Sciences, 280*, 386–395. doi: 10.1111/j.1749-6632.1976.tb25503.x

Marler, P. (1976b). Sensory templates in species-specific behavior. In J. C. Fentress (Ed.), *Simpler networks and behavior* (pp. 314–329). Sunderland, MA: Sinauer.

Marler, P. (1997). Three models of song learning: Evidence from behavior. *Journal of Neurobiology, 33*, 501–516. doi: 10.1002/(SICI)1097-4695(19971105)33:5<501::AID-NEU2>3.0.CO;2-8

Marshall, A. J., Wrangham, R. W., & Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour, 58*, 825–830. doi: 10.1006/anbe.1999.1219

Masataka, N. (2003). *The onset of language.* Cambridge: Cambridge University Press.

May-Collado, L. J. (2010). Changes in whistle structure of two dolphin species during interspecific associations. *Ethology, 116*, 1065–1074. doi: 10.1111/j.1439-0310.2010.01828.x

McCowan, B., & Reiss, D. (1995). Whistle contour development in captive-born infant bottlenose dolphins (*Tursiops truncatus*): Role of learning. *Journal of Comparative Psychology, 109*, 242–260. doi: 10.1037//0735-7036.109.3.242

McCowan, B., & Reiss, D. (1997). Vocal learning in captive bottlenose dolphins: A comparison with humans and nonhuman animals. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 178–207). Cambridge: Cambridge University Press.

McCowan, B., & Reiss, D. (2001). The fallacy of 'signature whistles' in bottlenose dolphins: A comparative perspective of 'signature information' in animal vocalizations. *Animal Behaviour, 62*, 1151–1162. doi: 10.1006/anbe.2001.1846

McGregor, P., Horn, A. G., & Todd, M. A. (1985). Are familiar sounds ranged more accurately? *Perceptual and Motor Skills, 61*, 1082.

McPherson, G. E., & Gabrielsson, A. (2002). From sound to sign. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance: Creative strategies for teaching and learning* (pp. 99–116). New York: Oxford University Press.

Mercado, E., III. (2008). Neural and cognitive plasticity: From maps to minds. *Psychological Bulletin, 134*, 109–137. doi: 10.1037/0033-2909.134.1.109

Mercado, E., III, & DeLong, C. M. (2010). Dolphin cognition: Representations and processes in memory and perception. *International Journal of Comparative Psychology, 33*, 344–378.

Mercado, E., III, & Frazer, L. N. (1999). Environmental constraints on sound transmission by humpback whales. *Journal of the Acoustical Society of America, 106*, 3004–3016. doi: 10.1121/1.428120

Mercado, E., III, & Frazer, L. N. (2001). Humpback whale song or humpback whale sonar? A reply to Au et al. *IEEE Journal of Oceanic Engineering, 26*, 406–415. doi: 10.1109/48.946514

Mercado, E., III, Herman, L. M., & Pack, A. A. (2005). Song copying by humpback whales: Themes and variations. *Animal Cognition, 8*, 93–102. doi: 10.1007/s10071-004-0238-7

Mercado, E., III, Murray, S. O., Uyeyama, R. K., Pack, A. A., & Herman, L. M. (1998). Memory for recent actions in the bottlenosed dolphin (*Tursiops truncatus*): Repetition of arbitrary behaviors using an abstract rule. *Animal Learning & Behavior, 26*, 210-218. doi: 10.3758/BF03199213

Mercado, E., III, Schneider, J. N., Pack, A. A., & Herman, L. M. (2010). Sound production by singing humpback whales. *Journal of the Acoustical Society of America, 127*, 2678–2691. doi: 10.1121/1.3309453

Mercado, E., III, Uyeyama, R. K., Pack, A. A., & Herman, L. M. (1999). Memory for action events in the bottlenosed dolphin. *Animal Cognition, 2*, 17-25. doi: 10.1007/s100710050021

Miklosi, A. (1999). The ethological analysis of imitation. *Biological Reviews, 74*, 347-374. doi: 10.1017/S000632319900537X

Miksis, J. L., Tyack, P. L., & Buck, J. R. (2002). Captive dolphins, *Tursiops truncatus*, develop signature whistles that match acoustic features of human-made model sounds. *Journal of the Acoustical Society of America, 112*, 728–739. doi: 10.1121/1.1496079

Miller, N. E., & Dollard, J. (1941). *Social learning and imitation.* New Haven: Yale University Press.

Miller, P. J. O., Shapiro, A. D., Tyack, P. L., & Solow, A. R. (2004). Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*. *Animal Behaviour, 67*, 1099–1107. doi: 10.1016/j.anbehav.2003.06.017

Miller, R., Sanchez, K., & Rosenblum, L. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics, 72*, 1614–1625. doi: 10.3758/APP.72.6.1614

Mithen, S. (2009). The music instinct: The evolutionary basis of musicality. *The Neurosciences and Music III - Disorders and Plasticity: Annals of the New York Academy of Science, 1169*, 3–12. doi: 10.1111/j.1749-6632.2009.04590.x

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition, 109*, 168–173. doi: 10.1016/j.cognition.2008.08.002

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neuroscience and Biobehavioral Reviews, 33*, 975–980. doi: 10.1016/j.neubiorev.2009.03.010

Molles, L. E., & Vehrencamp, S. L. (1999). Repertoire size, repertoire overlap, and singing modes in the banded wren (*Thryothorus pleurostictus*). *Auk, 116*, 677–689.

Molliver, M. E. (1963). Operant control of vocal behavior in the cat. *Journal of the Experimental Analysis of Behavior, 6*, 197–202. doi: 10.1901/jeab.1963.6-197

Moore, B. R. (1992). Avian movement imitation and a new form of mimicry: Tracing the evolution of a complex form of learning. *Behaviour, 122*, 231–263. doi: 10.1163/156853992X00525

Moore, B. R. (2004). The evolution of learning. *Biological Reviews, 79*, 301–335. doi: 10.1017/S0464793103006225

Moore, R., Estis, J., Gordon-Hickey, S., & Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *Journal of Voice, 22*, 399–407. doi: 10.1016/j.jvoice.2009.10.010

Morgan, C. L. (1896). *Habit and instinct.* London: Arnold.

Morton, E. S. (1982). Grading, discreteness, redundancy, and motivation-structural rules. In D. E. Kroodsma & E. H. Miller (Eds.), *Acoustic communication in birds* (pp. 183–212). New York: Academic Press.

Morton, E. S. (1986). Predictions from the ranging hypothesis for the evolution of long distance signals in birds. *Behaviour, 99*, 65–86. doi: 10.1163/156853986X00414

Morton, E. S. (1996). Why songbirds learn songs: An arms race over ranging? *Poultry and Avian Biology Reviews, 7*, 65–71.

Morton, E. S. (2012). Putting distance back into bird song with mirror neurons. *Auk, 129*, 560–564. doi: 10.1525/auk.2012.12072

Morton, E. S., Howlett, J., Kopysh, N. C., & Chiver, I. (2006). Song ranging by incubating male Blue-headed Vireos: The importance of song representation in repertoires and implications for song delivery patterns and local/foreign dialect discrimination. *Journal of Field Ornithology, 77*, 291–301. doi: 10.1111/j.1557-9263.2006.00055.x

Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cerebral Cortex, 23*, 1190-1197. doi: 10.1093/cercor/bhs110

Mowrer, O. H. (1952). The autism theory of speech development and some clinical applications. *Journal of Speech and Hearing Disorders, 17*, 263–268.

Mowrer, O. H. (1960). *Learning theory and the symbolic processes.* New York: John Wiley.

Mürbe, D., Friedmann, P., Hofmann, G., & Sundberg, J. (2002). Significance of auditory and kinesthetic feedback to singers' pitch control. *Journal of Voice, 16*, 44–51. doi: 10.1016/S0892-1997(02)00071-1

Murray, S. O., Mercado, E., & Roitblat, H. L. (1998). Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations. *Journal of the Acoustical Society of America, 104*, 1679–1688. doi: 10.1121/1.424380

Myers, S. A., Horel, J. A., & Pennypacker, H. S. (1965). Operant control of vocal behavior in the monkey. *Psychonomic Science, 3*, 389–390.

Naguib, M., & Wiley, H. (2001). Estimating the distance to a source of sound: Mechanisms and adaptations for long-range communication. *Animal Behaviour, 62*, 825–837. doi: 10.1006/anbe.2001.1860

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*, 422–432. doi: 10.1177/026192702237958

Nattkemper, D., Ziessler, M., & Frensch, P. A. (2010). Binding in voluntary action control. *Neuroscience and Biobehavioral Reviews, 34*, 1092–1101. doi: 10.1016/j.neubiorev.2009.12.013

Neumann, R., & Strack, F. (2000). "Mood contagion": The automatic transfer of mood between persons. *Journal of Personality and Social Psychology, 79*, 211–223. doi: 10.1037//0022-3514.79.2.211

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*, 132–142. doi: 10.1016/j.wocn.2010.12.007

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. (2000). Cultural revolution in whale songs. *Nature, 408*, 537. doi: 10.1038/35046199

Nottebohm, F., & Liu, W. C. (2010). The origins of vocal learning: New sounds, new circuits, new cells. *Brain and Language, 115*, 3–17. doi: 10.1016/j.bandl.2010.05.002

Ocampo, B., & Kritikos, A. (2011). Interpreting actions: The goal behind mirror neuron function. *Brain Research Reviews, 67*, 260–267. doi: 10.1016/j.brainresrev.2011.03.001

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology, 73*, 530–544. doi: 10.1002/ajp.20913

Palmer, C., & Drake, C. (1997). Monitoring and planning capacities in the acquisition of music performance skills. *Canadian Journal of Experimental Psychology, 51*, 369–384. doi: 10.1037/1196–1961.51.4.369

Panova, E. M., Belikov, R. A., Agafonov, A. V., & Bel'kovich, V. M. (2012). The relationship between the behavioral activity and the underwater vocalization of the beluga whale (*Delphinapterus leucas*). *Oceanology, 52*, 79–87. doi: 10.1134/S000143701201016X

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*, 190–197. doi: 10.1016/j.wocn.2011.10.001

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics, 72*, 2254–2264. doi: 10.3758/APP.72.8.2254

Parton, D. A. (1976). Learning to imitate in infancy. *Child Development, 47*, 14-31. doi: 10.1111/j.1467-8624.1976.tb03389.x

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*, 674–681. doi: 10.1038/nn1082

Patterson, D. K., & Pepperberg, I. M. (1994). A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels. *Journal of the Acoustical Society of America, 96*, 634–648. doi: 10.1121/1.410303

Payne, K., & Payne, R. S. (1985). Large scale changes over 19 years in songs of humpback whales in Bermuda. *Zeitschrift fur Tierpsychologie, 68*, 89–114.

Payne, K., Tyack, P., & Payne, R. S. (1983). Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): A detailed analysis of two seasons in Hawaii. In R. Payne (Ed.), *Communication and behavior of whales* (pp. 9–57). Boulder, CO: Westview Press.

Payne, R. S., & McVay, S. (1971). Songs of humpback whales. *Science, 173*, 585–597. doi: 10.1126/science.173.3997.585

Pepperberg, I. M. (1986). Social modeling theory: A possible framework for understanding avian learning. *Auk, 102*, 854–864.

Pepperberg, I. M. (2005). Insights into vocal imitation in African grey parrots (*Psittacus erithacus*). In S. Hurley & N. Chater (Eds.), *Perspectives on imitation, vol 1: Mechanisms of imitation and imitation in animals* (pp. 243–262). Cambridge, MA: MIT Press.

Pepperberg, I. M. (2010). Vocal learning in grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain and Language, 115*, 81–91. doi: 10.1016/j.bandl.2009.11.002

Perelberg, A., & Schuster, R. (2008). Coordinated breathing in bottlenose dolphins (*Tursiops truncatus*) as cooperation: Integrating proximate and ultimate explanations. *Journal of Comparative Psychology, 122*, 109–120.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688–691. doi: 10.1038/nn1083

Petrinovich, L. (1988). The role of social factors in white-crowned sparrow song development. In T. R. Zentall & B. G. Galef (Eds.), *Social learning: Psychological and biological perspectives* (pp. 255–278). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception, 25*, 95–115. doi: 10.1525/MP.2007.25.2.95

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-013-0401-8

Pfordresher, P. Q., & Mantell, J. T. (2009). *Singing as a form of vocal imitation: Mechanisms and deficits*. Paper presented at the Proceedings of the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music.

Pfordresher, P. Q., & Mantell, J. T. (2012). Effects of altered auditory feedback across effector systems: Production of melodies by keyboard and singing. *Acta Psychologica, 139*, 166–177. doi: 10.1016/j.actpsy.2011.10.009w

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology, 70*, 31-57. doi: 10.1016/j.cogpsych.2013.12.005

Piaget, J. (1962). *Play, dreams, and imitation in childhood*. New York: W. W. Norton.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences, 3*, 136–141. doi: 10.3389/fnhum.2012.00185

Pickering, M. J., & Garrod, S. (2006). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105–110. doi: 10.1016/j.tics.2006.12.002

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Animal behaviour: Elephants are capable of vocal learning. *Nature, 434*, 455–456. doi: 10.1038/435042b

Popper, A. N., & Edds-Walton, P. L. (1997). Bioacoustics of marine vertebrates. In M. J. Crocker (Ed.), *Encyclopedia of acoustics* (pp. 1831–1836). New York: John Wiley & Sons.

Porter, R. J. J., & Lubker, J. F. (1980). Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motoric linkage in speech. *Journal of Speech and Hearing Research, 23*, 593–602.

Poulson, C. L., Kymissis, E., Reeve, K. F., Andreators, M., & Reeve, L. (1991). Generalized vocal imitation in infants. *Journal of Experimental Child Psychology, 51*, 267–279. doi: 10.1016/0022-0965(91)90036-R

Poulson, C. L., Kyparissos, N., Andreatos, M., Kymissis, E., & Parnes, M. (2002). Generalized imitation within three response classes in typically developing infants. *Journal of Experimental Child Psychology, 81,* 341–357. doi: 10.1006/jecp.2002.2661

Price, C., & Griffiths, T. D. (2005). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences, 9*, 271–276. doi: 10.1016/j.tics.2005.03.009

Price, H. E. (2000). Interval matching by undergraduate nonmusic majors. *Journal of Research in Music Education, 48*, 360–372. doi: 10.2307/3345369

Price, J. J., & Yuan, D. H. (2011). Song-type sharing and matching in a bird with very large song repertoires, the tropical mockingbird. *Behaviour, 148*, 673–689. doi: 10.1163/000579511X573908

Prizant, B. M., & Rydell, P. J. (1984). Analysis of functions of delayed echolalia. *Journal of Speech and Hearing Research, 27*, 183–192.

Quick, N. J., & Janik, V. M. (2012). Bottlenose dolphins exchange signature whistles when meeting at sea. *Proceedings of the Royal Society B - Biological Sciences, 279*, 2539–2545. doi: 10.1098/rspb.2011.2537

Reidenberg, J. S., & Laitman, J. T. (1988). Existence of vocal folds in the larynx of Odontoceti (toothed whales). *Anatomical Record, 221*, 884–891. doi: 10.1002/ar.1092210413

Reidenberg, J. S., & Laitman, J. T. (2007). Discovery of a low frequency sound source in Mysticeti (baleen whales): Anatomical establishment of a vocal fold homolog. *Anatomical Record, 290*, 745–759. doi: 10.1002/ar.20544

Reiss, D., & McCowan, B. (1993). Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): Evidence for vocal learning. *Journal of Comparative Psychology, 107*, 301–312. doi: 10.1037/0735-7036.107.3.301

Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., . . . Ackermann, H. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology, 2*, 271. doi: 10.3389/fpsyg.2011.00271

Reiterer, S. M., Singh, N. C., & Winkler, S. (2012). Predicting speech imitation ability biometrically. In B. Stolterfoht & S. Featherston (Eds.), *Empirical approaches to linguistic theory: Studies in meaning and structure* (pp. 317–339). Berlin: De Gruyter.

Rendell, L., & Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and Brain Sciences, 24*, 309–324.

Rendell, L., & Whitehead, H. (2003). Vocal clans in sperm whales (*Physeter macrocephalus*). *Proceedings of the Royal Society B - Biological Sciences, 270*, 225–231. doi: 10.1098/rspb.2002.2239

Repp, B. H., & Williams, D. R. (1987). Categorical tendencies in imitating self-produced isolated vowels. *Speech Communication, 6*, 1–14. doi: 10.1016/0167-6393(87)90065-3

Revis, J., De Looze, C., & Giovanni, A. (2013). Vocal flexibility and prosodic strategies in a professional impersonator. *Journal of Voice*. doi: 10.1016/j.jvoice.2013.01.008

Richards, D. G. (1986). Dolphin vocal mimicry and vocal object labeling. In R. J. Schusterman, J. A. Thomas, & F. G. Wood (Eds.), *Dolphin cognition and behavior: A comparative approach* (pp. 273–288). Hillsdale, NJ: Lawrence Erlbaum Associates.

Richards, D. G., Wolz, J. P., & Herman, L. M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *Journal of Comparative Psychology, 98*, 10–28. doi: 10.1037/0735-7036.98.1.10

Ridgway, S., Carder, D., Jeffries, M., & Todd, M. (2012). Spontaneous human speech mimicry by a cetacean. *Current Biology, 22*, R860–861. doi: 10.1016/j.cub.2012.08.044

Riesch, R., Ford, J. K. B., & Thomsen, F. (2006). Stability and group specificity of stereotyped whistles in resident killer whales, *Orcinus orca*, off British Columbia. *Animal Behavior, 71*, 79–91. doi: 10.1016/j.anbehav.2005.03.026

Roitblat, H. L. (1982). The meaning of representation in animal memory. *Behavioral and Brain Sciences, 5*, 353–372. doi: 10.1017/S0140525X00012486

Roitblat, H. L., & von Fersen, L. (1992). Comparative cognition: Representations and processes in learning and memory. *Annual Review of Psychology, 43*, 671–710.

Romanes, G. J. (1884). *Mental evolution in animals.* New York: D. Appleton & Co.

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology, 52*, 453–470. doi: 10.1146/annurev.psych.52.1.453

Rothenberg, D. (2008). Whale music: Anatomy of an interspecies duet. *Leonardo Music Journal, 18*, 47–53. doi: 10.1162/lmj.2008.18.47

Russell, J. L., Hopkins, W. D., & Taglialatela, J. P. (2012). Vocal learning in captive chimpanzees (*Pan troglodytes*): Evidence of flexibility and voluntary control. *American Journal of Primatology, 74*, 66.

Salzinger, K. (1993). Animal communication. In D. A. Dewsbury & D. A. Rethlingshafer (Eds.), *Comparative psychology: A modern survey* (pp. 161–193). New York: McGraw-Hill.

Salzinger, K., & Waller, B. W. (1962). The operant control of vocalization in the dog. *Journal of the Experimental Analysis of Behavior, 5*, 383–389.

Sayigh, L. S., Tyack, P. L., Wells, R. S., & Scott, M. D. (1990). Signature whistles of free-ranging bottlenose dolphins, *Tursiops truncatus*: Mother-offspring comparisons. *Behavioral Ecology and Sociobiology, 26,* 247–260.

Sayigh, L. S., Tyack, P. L., Wells, R. S., Scott, M. D., & Irvine, A. B. (1995). Sex differences in signature whistle production of free-ranging bottlenose dolphins, *Tursiops truncatus. Behavioral Ecology and Sociobiology, 36*, 171–177. doi: 10.1007/BF00177793

Sayigh, L. S., Tyack, P. L., Wells, R. S., Solow, A. R., Scott, M. D., & Irvine, A. B. (1999). Individual recognition in wild bottlenose dolphins: A field test using playback experiments. *Animal Behaviour, 57*, 41–50. doi: 10.1006/anbe.1998.0961

Schevill, W. E., & Lawrence, B. (1949). Listening to the white porpoise (*Delphinapterus leucas*). *Science, 109*, 143–144. doi: 10.1126/science.109.2824.143

Schuler, A. L. (1979). Echolalia: Issues and clinical applications. *Journal of Speech and Hearing Disorders, 44,* 411–434.

Schusterman, R. J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. In D. K. Oller & U. Gribel (Eds.), *The evolution of communicative flexibility: Complexity, creativity, and adaptability in human and animal communication* (pp. 41–70). Cambridge, MA: MIT Press.

Schusterman, R. J., & Feinstein, S. H. (1965). Shaping and discriminative control of underwater click vocalizations in a California sea lion. *Science, 150*, 1743–1744. doi: 10.1126/science.150.3704.1743

Searcy, W. A., DuBois, A. L., Rivera-Caceres, K., & Nowicki, S. (2013). A test of a hierarchical signalling model in song sparrows. *Animal Behavior, 86*, 309–315. doi: 10.1016/j.anbehav.2013.05.019

Sewall, K. (2012). Vocal matching in animals. *American Scientist, 100*, 306–315.

Shapiro, A. D., & Slater, P. J. B. (2004). Call usage learning in gray seals (*Halichoerus grypus*). *Journal of Comparative Psychology, 118*, 447–454. doi: 10.1037/0735-7036.118.4.447

Shettleworth, S. J. (1998). *Cognition, evolution, and behavior.* New York: Oxford University Press.

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science, 1*, 305–319. doi: 10.1111/j.1756-8765.2009.01021.x

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*, 422–429. doi: 10.3758/BF03194890

Shy, E., & Morton, E. S. (1986). The role of distance, familiarity, and time of day in Carolina Wren responses to conspecific songs. *Behavioral Ecology and Sociobiology, 19*, 393–400. doi: 10.1007/BF00300541

Sigurdson, J. (1993). Frequency-modulated whistles as a medium for communication with the bottlenose dolphin (*Tursiops truncatus*). In H. L. Roitblat, L. M. Herman, & P. E. Nachtigall (Eds.), *Language and communication: Comparative perspectives* (pp. 153–174). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sjare, B. L., & Smith, T. G. (1986). The vocal repertoire of white whales, *Delphinapterus leucas*, summering in Cunningham Inlet, Northwest Territories. *Canadian Journal of Zoology, 64*, 407–415. doi: 10.1139/z86-063

Skoyles, J. R. (1998). Speech phones are a replication code. *Medical Hypotheses, 50*, 167–173. doi: 10.1016/S0306-9877(98)90203-1

Smith, J. N., Goldizen, A. W., Dunlop, R. A., & Noad, M. J. (2008). Songs of male humpback whales, *Megaptera novaeangliae*, are involved in intersexual interactions. *Animal Behaviour, 76*, 467–477. doi: 10.1016/j.anbehav.2008.02.013

Smotherman, M. S. (2007). Sensory feedback control of mammalian vocalizations. *Behavioural Brain Research, 182*, 315–326. doi: 10.1016/j.bbr.2007.03.008

Stimpert, A. K., Peavey, L. E., Friedlaender, A. S., & Nowacek, D. P. (2012). Humpback whale song and foraging behavior on an Antarctic feeding ground. *PLoS One, 7*(12), e51214. doi: 10.1371/journal.pone.0051214

Stimpert, A. K., Wiley, D. N., Au, W. W., Johnson, M. P., & Arsenault, R. (2007). 'Megapclicks': Acoustic click trains and buzzes produced during night-time foraging of humpback whales (*Megaptera novaeangliae*). *Biology Letters, 3*, 467–470. doi: 10.1098/rsbl.2007.0281

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology, 22*, 2144–2148. doi: 10.1016/j.cub.2012.09.022

Strager, H. (1995). Pod specific call repertoires and compound calls of killer whales, *Orcinus orca*, Linnaeus, 1758, in waters of northern Norway. *Canadian Journal of Zoology, 73*, 1037–1047. doi: 10.1139/z95-124

Studdert-Kennedy, M. (2000). Imitation and the emergence of segments. *Phonetica, 57*, 275–283. doi: 10.1159/000028480

Subiaul, F. (2010). Dissecting the imitation faculty: The multiple imitation mechanisms (MIM) hypothesis. *Behavioural Processes, 83*, 222–234. doi: 10.1016/j.beproc.2009.12.002

Subiaul, F., Anderson, S., Brandt, J., & Elkins, J. (2012). Multiple imitation mechanisms in children. *Developmental Psychology, 48*, 1165–1179. doi: 10.1037/a0026646

Taglialatela, J. P., Reamer, L., Schapiro, S. J., & Hopkins, W. D. (2012). Social learning of a communicative signal in captive chimpanzees. *Biology Letters, 8*, 498–501. doi: 10.1098/rsbl.2012.0113

Tayler, C. K., & Saayman, G. S. (1973). Imitative behavior by Indian bottlenose dolphins (*Tursiops aduncus*) in captivity. *Behaviour, 44*, 286–298.

Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science, 291*, 2564–2569. doi: 10.1126/science.1058522

Thomsen, F., Franck, D., & Ford, J. K. (2002). On the communicative significance of whistles in wild killer whales (Orcinus orca). *Naturwissenschaften, 89*, 404–407. doi: 10.1007/s00114-002-0351-x

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. New York: Hafner Publishing.

Thorpe, W. H. (1956). *Learning and instinct in animals*. London:Methuen and Co.

Thorpe, W. H. (1969). The significance of vocal imitation in animals with special reference to birds. *Acta Biologiae Experimentalis, 29*, 251–269.

Thorpe, W. H., & North, M. E. W. (1965). Origin and significance of the power of vocal imitation: With special reference to the antiphonal singing of birds. *Nature, 208*, 219–222. doi: 10.1038/208219a0

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*, 952–981. doi: 10.1080/01690960903498424

Troyer, T. W., & Doupe, A. J. (2000a). An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. *Journal of Neurophysiology, 84*, 1204–1223.

Troyer, T. W., & Doupe, A. J. (2000b). An associational model of birdsong sensorimotor learning II. Temporal hierarchies and the learning of song sequence. *Journal of Neurophysiology, 84*, 1224–1239.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1–25. doi: 10.1146/annurev.psych.53.100901.135114

Tyack, P. L. (1986). Whistle repertoires of two bottlenosed dolphins, *Tursiops truncatus*: Mimicry of signature whistles? *Behavioral Ecology and Sociobiology, 18*, 251–257. doi: 10.1007/BF00300001

Tyack, P. L. (1991). Use of a telemetry device to identify which dolphin produces a sound. In K. Pryor & K. S. Norris (Eds.), *Dolphin societies: Discoveries and puzzles* (pp. 319–344). Berkeley: University of California Press.

Tyack, P. L. (2000). Functional aspects of cetacean communication. In J. Mann, R. C. Connor, P. L. Tyack, & H. Whitehead (Eds.), *Cetacean societies: Field studies of dolphins and whales* (pp. 270–307). Chicago: University of Chicago Press.

Tyack, P. L. (2008). Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *Journal of Comparative Psychology, 122*, 319–331. doi: 10.1037/a0013087

Tyack, P. L., & Clark, C. W. (2000). Communication and acoustic behavior of whales and dolphins. In W. W. L. Au, A. N. Popper, & R. R. Fay (Eds.), *Hearing by whales and dolphins* (pp. 156–224). New York: Springer.

Tyack, P. L., & Sayigh, L. S. (1997). Vocal learning in cetaceans. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 208–233). Cambridge: Cambridge University Press.

Vallabha, G. K., & Tuller, B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *Journal of the Acoustical Society of America, 116*, 1184–1197. doi: 10.1121/1.1764832

van Heel, W. H. D., Kamminga, C., & van der Toorn, J. D. (1982). An experiment in two-way communication in *Orcinus orca* L. *Aquatic Mammals, 9*, 69–82.

van Santen, J. P. H., Sproat, R. W., & Hill, A. P. (2013). Quantifying repetitive speech in autism spectrum disorders and language impairment. *Autism Research, 6*, 372–383. doi: 10.1002/aur.1301

Vergara, V., & Barrett-Lennard, L. G. (2008). Vocal development in a beluga calf (*Delphinapterus leucas*). *Aquatic Mammals, 34*, 123–143. doi: 10.1578/AM.34.1.2008.123

Vesper, C., van der Wel, R. P., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 48-61. doi: 10.1037/a0028066

Wang, D., Yan, N., & Ng, L. (2012). Effects of augmented auditory feedback on pitch production accuracy in singing. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition* (pp. 1116–1119). Thessaloniki, Greece: Aristotle University of Thessaloniki.

Ward, W. D., & Burns, E. M. (1978). Singing without auditory feedback. *Journal of Research in Singing, 1*, 24–44.

Watts, C. R., & Hall, M. D. (2008). Timbral influences on vocal pitch-matching accuracy. *Logopedics Phoniatrics Vocology, 33*, 74–82. doi: 10.1080/14015430802028434

Watwood, S. L., Tyack, P. L., & Wells, R. S. (2004). Whistle sharing in paired male bottlenose dolphins, *Tursiops truncatus*. *Behavioral Ecology and Sociobiology, 55*, 531–543. doi: 10.1007/s00265-003-0724-y

Weib, B. M., Symonds, H., Spong, P., & Ladich, F. (2011). Call sharing across vocal clans of killer whales: Evidence for vocal imitation. *Marine Mammal Science, 27*, E1–E13. doi: 10.1111/j.1748-7692.2010.00397.x

Welch, G. F. (1979). Vocal range and poor pitch singing. *Psychology of Music, 7*, 13–31. doi: 10.1177/030573567972002

Westermann, G., & Reck Miranda, E. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language, 89*, 393–400. doi: 10.1016/S0093-934X(03)00345-6

Whiten, A., & Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. *Advances in the Study of Behavior, 21*, 239–283.

Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates, 50*, 56–64. doi: 10.1007/s10329-008-0117-y

Wickler, W. (2013). Understanding mimicry—with special reference to vocal mimicry. *Ethology, 119*, 259–269. doi: 10.1111/eth.12061

Wiley, D., Ware, C., Bocconcelli, A., Cholewiak, D., Friedlaender, A., Thompson, M., & Weinrich, M. (2011). Underwater components of humpback whale bubble-net feeding behavior. *Behaviour, 148*, 575–602. doi: 10.1163/000579511X570893

Williamson, V. J., Jilka, S. R., Fry, J., Finkel, S., Mullensiefen, D., & Stewart, L. (2012). How do "earworms" start? Classifying the everyday circumstances of involuntary musical imagery. *Psychology of Music, 40*, 259–284. doi: 10.1177/0305735611418553

Wilson, M. (2001a). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review, 8*, 44-57. doi: 10.3758/BF03196138

Wilson, M. (2001b). Perceiving imitatible stimuli: Consequences of isomorphism between input and output. *Psychological Bulletin, 127*, 543-553. doi: 10.1037//0033-2909.127.4.543

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*, 460–473. doi: 10.1037/0033-2909.131.3.460

Wise, K., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined 'tone deafness': Perception, singing performance, and self-assessment. *Musicae Scientiae, 12*, 3–23. doi: 10.1177/102986490801200102

Wisniewski, M. G., Mantell, J. T., & Pfordresher, P. Q. (2013). Transfer effects in the vocal imitation of speech and song. *Psychomusicology: Music, Mind, and Brain, 23*, 82–99.

Wisniewski, M. G., Mercado, E., III, Gramann, K., & Makeig, S. (2012). Familiarity with speech affects cortical processing of auditory distance cues and increases acuity. *PLoS One, 7*, e41025. doi: 10.1371/journal.pone.0041025

Witchell, C. A. (1896). *The evolution of bird-song with observations of heredity and imitation.* London: Adam and Charles Black.

Woody, R. H., & Lehmann, A. C. (2010). Student musician's ear-playing ability as a function of vernacular music experience. *Journal of Research in Music Education, 58*, 101–115. doi: 10.1177/0022429410370785

Yeater, D. B., & Kuczaj, S. A., II. (2010). Observational learning in wild and captive dolphins. *International Journal of Comparative Psychology, 23*, 379–385.

Yu, A., Abrego-Collier, C., Baglini, R., Grano, T., Martinovik, M., Otte, C., & Urban, J. (2011). *Speaker attitude and sexual orientation affect phonetic imitation*. Paper presented at the Proceedings of the 34th Annual Penn Linguistics Colloquium (Vol. 17).

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences, USA, 107*, 592–597. doi: 10.1073/pnas.0904774107

Yurk, H., Barrett-Lennard, L. G., Ford, J. K. B., & Matkin, C. O. (2002). Cultural transmission within maternal lineages: Vocal clans in resident killer whales in southern Alaska. *Animal Behavior, 63*, 1103–1119. doi: 10.1006/anbe.2002.3012

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica United with Acustica, 91*, 409–420.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology, 10*, e1001372. doi: 10.1371/journal.pbio.1001372

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6*, 37–46. doi: 10.1016/S1364-6613(00)01816-7

Zentall, T. R. (2006). Imitation: Definitions, evidence, and mechanisms. *Animal Cognition, 9*, 335–353. doi: 10.1007/s10071-006-0039-2

Zentall, T. R., & Akins, C. (2001). Imitation in animals: Evidence, function, and mechanisms. In R. G. Cook (Ed.), *Avian visual cognition* [Online]: Available: www.pigeon.psy.tufts.edu/avc/zentall.

Zetterholm, E. (2006). *Same speaker—different voices: A study of one impersonator and some of his different imitations*. Paper presented at the Proceedings of the 11th Australian International Conference on Speech Science & Technology, University of Auckland, New Zealand.

Zhang, J., Hughes, L. E., & Rowe, J. B. (2012). Selection and inhibition mechanisms for human voluntary action decisions. *Neuroimage, 63*, 392–402. doi: 10.1016/j.neuroimage.2012.06.058