"Hands-On, Mouth-Off": The Relationship Between Hand Movements and Language Evolution

Authors: Yin Rong, Yin Rong

Date: 2020-04-25T23:15:31+00:00

Abstract

Language evolution is a significant issue in the field of evolutionary psychology. The mirror system hypothesis, tool-making hypothesis, and teaching hypothesis explain the relationship between manual actions and language evolution from different perspectives, and all three hypotheses hold that human language originated from manual action experience. Related empirical studies have found that sign language and spoken language share consistent features, that language and manual actions have a common neural basis, that gesture development can predict language development levels, and that gestures can improve the transmission efficiency of tool-making knowledge; these studies provide empirical support for the specific claims of the three hypotheses. Future research in this field needs to focus on the developmental relationship between sign language and spoken language in evolution, as well as the relationship between human language evolution and the evolution of other cognitive features.

Full Text

"Actions Speak Louder Than Words": The Relationship Between Hand Actions and Language Evolution

YIN Rong
School of Teacher Education, Weifang University, Weifang, Shandong 261061, China

Abstract

Language evolution represents a significant issue within evolutionary psychology research. The Mirror System Hypothesis, Tool-making Hypothesis, and Teaching Hypothesis each explain the relationship between hand actions and language evolution from distinct perspectives, yet all three converge on the view that human language originated from manual movement experiences. Relevant empirical studies have revealed that sign language and spoken language share consistent features, that language and hand movements possess common neural substrates, that gesture development can predict language development levels, and that gestures can enhance the transmission efficiency of tool-making knowledge. These findings provide empirical support for the specific claims of the three hypotheses. Future research in this field must focus on the evolutionary developmental relationship between sign language and spoken language, as well as the relationship between human language evolution and the evolution of other cognitive characteristics.

Keywords: hand action; gesture; sign language; language; evolution
Classification Number: B842

Introduction

In his immortal masterpiece *Hamlet*, Shakespeare extolled the beauty of humanity with passionate verse: "What a piece of work is man! How noble in reason! How infinite in faculty… the paragon of animals!" Undoubtedly, among all the features that distinguish humans from the animal world, language ability occupies a position of paramount importance. While most animals in nature possess communication systems, and even insects with the simplest biological structures exhibit effective social communication forms, the information animals can convey is limited, whereas human language possesses almost unlimited openness. A mature language system represents one of the latest features to emerge in human evolutionary history. Speaking may seem like an ordinary behavior, yet it epitomizes the concentrated expression of human mental capacities. On the scientific path of exploring human and human mind evolution, the origin of language is an unavoidable question.

However, prior to the 1980s, the scientific community harbored considerable doubts about the evolutionary adaptability of language. For instance, according to linguist Chomsky (2002), while human language ability is indeed an innate "mental organ," mental organs are not products of adaptive selection. Chomsky argued that as brains became increasingly sophisticated through evolution, humans were simply fortunate enough to "accidentally" acquire language. Similarly, biologist Gould (1991) insisted that language was a byproduct of expanded brain capacity and enhanced human thinking. This viewpoint later encountered fierce criticism. Pinker (2007), in his work *The Language Instinct*, pointed out that to produce speech and process language, humans require a series of mutually adapted and coordinated complex physiological structures, including the larynx, vocal cords, oral cavity, tongue, and language centers in the brain. These elements could not possibly appear suddenly and fuse into a unified whole. Therefore, language is by no means an evolutionary byproduct; as a complex functional system, its formation can only be explained through the theory of natural selection.

Since most human psychological mechanisms (such as aesthetic sense, mate preferences, or moral justice) produce clear and limited behavioral outcomes, this provides clues for evolutionary psychologists to study their original functions through "reverse engineering" methods. Language, however, involves virtually every aspect of human life—communicating emotions, transmitting information, cooperating reciprocally, pursuing mates, teaching offspring, describing knowledge—all are functions language can serve. Behavioral outcomes do not equate to evolutionary causes. Once complex language began to evolve, it would generate various uses unrelated to its original function, making it extremely difficult to distinguish which were results of natural selection and which were later-developed functions. Consequently, language origin has become one of the most contentious and controversial issues in evolutionary psychology research. In recent years, an increasing number of scholars have focused on this field, examining different explanatory angles such as kin communication (Nowicki & Searcy, 2014; Smit, 2014), social relationships (Dunbar, 1998), reciprocal cooperation (Tomasello, 2008), sexual resource competition (Okanoya, 2017), or knowledge transmission (Laland, 2017b). This paper focuses on the relationship between hand actions and language evolution, systematically introducing relevant theoretical and empirical research while reviewing existing problems and future research directions.

Theoretical Research on the Relationship Between Hand Actions and Language Evolution

Hand actions and language share a close and complex relationship. Although speaking is the most common communication method in daily life, vocal sound is not the only information transmission medium during conversation. Most people "talk with their hands" intentionally or unintentionally during dialogue. As a universally present auxiliary communication means, hand movements frequently appear in communication scenarios across all cultural backgrounds. Psychological research has found that although gestures and language differ in expression form, they complement each other. On one hand, gestures and language expression exhibit consistency, with gesture content always matching semantics. On the other hand, they also demonstrate complementarity, as gestures are not merely simulations of semantics but sometimes compensate for what language fails to express, enabling people to convey richer information through conversation. Additionally, gestures facilitate language processing, and this facilitation exists in both language production and comprehension processes. In other words, gestures help speakers construct their discourse expression while also making it easier for listeners to understand (Kang, Tversky, & Black, 2015; Novack & Goldin-Meadow, 2017). Experiments have shown that prohibiting gestures during communication causes speakers' language to lose fluency. Therefore, gestures can enhance mutual understanding between conversational partners and improve communication efficiency (Koppensteiner, Stephan, & Jäschke, 2016).

Neuroimaging research also reveals that communicative gestures involve the same brain regions as language comprehension (Marstaller & Burianová, 2014; Redcay, Ludlum, Velnoskey, & Kanwal, 2016; Yang, Andric, & Mathew, 2015). In summary, gestures represent an important medium in daily human communication and constitute an integral part of the human communication system. Over the past decade, numerous scholars have conducted extensive theoretical and empirical research on the cognitive characteristics of gestural communication, the brain mechanisms of sign language, and the influence of gestures on language processing, with domestic scholars having provided specialized introductions to these topics (Li & Ding, 2013; Zhang, 2019). Beyond this, however, does a deeper relationship exist between hand actions and language mechanisms? In recent years, the field of evolutionary psychology language research has begun focusing on this question, with researchers discussing the relationship between hand actions and language evolutionary origins (Arbib, 2010; Corballis, 2010; Fogassi & Ferrari, 2007; Tomasello, 2008). Representative theories include the Mirror System Hypothesis, Tool-making Hypothesis, and Teaching Hypothesis.

Mirror System Hypothesis

The Mirror System Hypothesis is directly related to the discovery of mirror neurons. In the 1990s, a team of neuroscientists led by Rizzolatti at the University of Parma attempted to implant microelectrodes into nerve cells in the ventral premotor cortex (F5 area) of macaque monkeys to record the electrical responses of these cells during monkey activities. Researchers accidentally discovered that even when monkeys themselves made no movements, if they observed other monkeys performing certain goal-directed actions—such as grasping a banana—certain nerve cells in their ventral premotor cortex F5 area would also become strongly activated (di Pellegrino et al., 1992). In other words, these neurons seemed able to "map" the actor's actions onto the observer's brain, enabling them to understand and recognize these actions. Consequently, researchers named them "mirror neurons" (Chen, 2019). Mirror neurons represent one of the most important and influential scientific discoveries in cognitive neuroscience research over the past 30 years, with numerous studies indicating that a similar "mirror system" for matching action observation and action execution exists in the human brain (Ye, 2019; Cracco et al., 2018; Salo, Ferrari, & Fox, 2019).

Based on the discovery of mirror neurons, Arbib and Rizzolatti (1997) proposed the Mirror System Hypothesis for language origins. Researchers argued that as a communication system, language must possess "parity property"—meaning the information the communicator intends to express can be accurately understood by the receiver—and parity is precisely the typical feature and function of the mirror system. Research has found that through mirror neuron activation, macaque monkeys can not only recognize other monkeys' actions but also predict their goals and subsequent actions, constituting the foundation for understanding behavioral intentions (Bonini, 2017; Ondobaka, Kilner, & Friston, 2017). Therefore, the mirror system provides the necessary neural mechanism for language emergence. The Mirror System Hypothesis posits that it was precisely the existence of the mirror system that enabled human ancestors to use hand actions to represent concrete objects, behaviors, or ideas, gradually leading to the formation of complex gestural communication abilities. Overall, modern humans' mature language originated from gestural language, and the neural circuits in the mirror system related to executing, recognizing, and imitating hand actions provided the basic architecture for the evolution of the brain's language functions (Arbib, 2012).

Based on this hypothesis, Arbib (2010, 2016) also proposed the basic process of language evolution: First, human ancestors had to develop the ability to recognize and imitate complex behaviors. This ability required organisms to decompose behaviors into combinations of familiar actions during recognition and to view certain actions as transformations of familiar actions. Subsequently, as some actions repeatedly appeared in individual interactions, human ancestors developed "pantomime" based on complex imitation abilities. They could transmit information between individuals and influence others' behaviors through hand actions alone, without any objects. At this stage, gestures were extracted from concrete actions and acquired representational functions. For example, during collective food gathering, people gradually developed pantomime gestures using "reaching" actions to express "picking."

However, the meanings of pantomime gestures remained highly ambiguous. As gestural communication expanded and became ritualized in some communities, conventionalized pantomime gestures evolved into protosign. Protosign simplified the action complexity of pantomime gestures while further clarifying their meanings. For instance, through differentiation and definition of gestures for "flying bird," separate concepts for "flying" and "bird" emerged. Finally, the protosign system provided a cognitive opportunity for the evolutionary adaptation of vocal language. Combining hand signs with vocal sounds produced protolanguage. After vocal language appeared, spoken language gradually refined into a more efficient communication method. With the expansion of early *Homo sapiens* cultural practices, mature language with rich concepts and grammatical rules was ultimately created.

Arbib (2017), the proponent of the Mirror System Hypothesis, acknowledges that many disagreements and uncertainties remain regarding the specific evolutionary course and neural mechanism transformations of human language. However, it is certain that the action-recognition mirror system provided the foundation for language evolution, and gestural language was the precursor to human language.

Tool-making Hypothesis

The Tool-making Hypothesis posits that tool manufacturing and use exerted evolutionary pressure on language, and the coupling between brain neural mechanisms related to manual operation and language functions provided the foundation for the development of language neural circuits in the brain (Kolodny & Edelman, 2018; Stout & Chaminade, 2012; Stout et al., 2011).

Fitch (2010) proposed that when exploring the causes of certain psychological mechanisms from an evolutionary perspective, one should consider the ecological context in which such mechanisms formed—namely, what kind of survival environment (social or natural) would lead to the emergence of such mechanisms. The ecological context related to language evolution must conform to the "independence principle," which specifically means that the ecological context producing language must appear independently before language itself. In other words, the ecological context that "incubated" language did not depend on language for its existence (Laland, 2017b). For example, anthropologist Dunbar (1998) proposed the social complexity hypothesis, suggesting that larger group sizes and social systems exerted evolutionary pressure on complex communication forms. However, Kolodny and Edelman (2018) pointed out that if language emerged to maintain complex social networks, how did such complex social networks appear before humans evolved language? Therefore, this hypothesis does not conform to the independence principle.

According to the Tool-making Hypothesis, the tool-using lifestyle precisely represents an ecological background that meets the above requirements and facilitates language evolution (Kolodny & Edelman, 2018). When human ancestors first began making and using tools, they did not depend on language. However, once early humans started using tools, groups that were better at transmitting tool-making technology between individuals possessed greater survival adaptability. Therefore, tool use and manufacture provided the ecological background and evolutionary driving force for language evolution (Stout & Chaminade, 2012). In this process, hand actions related to tool-making became coupled with the communication system, and the brain neural circuits executing manual operations developed new functions, evolving into brain regions supporting language functions (Hecht et al., 2015). Consequently, the cognitive characteristics of language can also reflect the cognitive characteristics of tool-making. Among these, tool-making requires decomposing tasks and performing manual operations in a certain sequence. The cognitive computational abilities for processing hierarchical structure and serial order are precisely what human language requires (Kolodny, Lotem, & Edelman, 2015). Only by combining words into sentences according to certain sequential and hierarchical rules can humans use limited symbols to express infinite meanings.

Kolodny and Edelman (2018) noted that the Tool-making Hypothesis's explanation for language origins conforms to the principle of evolutionary continuity. Chimpanzees, gorillas, and orangutans do not possess the physiological conditions for fine vocal control, but they can use simple gestures for communication. Gestures are the primary communication form for great apes, and when they use tools for foraging, they can exhibit hierarchical and sequential hand operation behaviors similar to humans. Human ancestors likely developed complex communication systems based on these abilities through a series of small, adaptive changes. Therefore, although human language differs enormously from the communication forms of other great apes, it originated from psychological and physiological mechanisms shared by human ancestors and other great apes.

Teaching Hypothesis

The Teaching Hypothesis posits that language initially evolved to facilitate teaching. During human evolution, as accumulated tools, experiences, and skills became increasingly rich, learning postnatal knowledge became increasingly important for individual and species survival adaptation. As an information communication method, language possesses the advantages of being precise, effective, and low-cost. Therefore, natural selection favored language as a psychological mechanism because early language could make teaching activities more economical and efficient (Dean et al., 2014; Laland, 2017b).

British biologist Laland (2017b), synthesizing other scholars' viewpoints, clarified seven criteria for evaluating language origin theories (Bickerton, 2009; Szamado & Szathmary, 2006). Laland argued that when addressing the evolutionary causes of language, the more criteria a theory satisfies, the stronger its explanatory power, and the Teaching Hypothesis satisfies all seven criteria. First, language origin theories must explain the authenticity of early language (Criterion 1). Human language costs very little to produce—whether gestures or vocal sounds require almost no energy expenditure. Given this, why would individuals believe information conveyed by others (Fitch, 2010)? This criterion requires that when language first emerged, information senders must transmit genuine information. If individuals send valuable, true information to others, there should be no conflict of interest between sender and receiver; information transmission must benefit both parties (Criterion 2). The Teaching Hypothesis provides a good explanation for this: since the purpose of teaching is to ensure knowledge is transmitted to others, if language was initially used for instruction, the transmitted information would naturally be true and reliable. The recipients of information are typically relatives or community companions of the teacher; when their survival competitiveness improves, the teacher's overall fitness also increases. Therefore, both parties can benefit from information transmission (Fitch, 2010).

Successful theories must also explain how primitive language possessed adaptive capabilities when it first originated—that is, what evolutionary advantages primitive language demonstrated when its conceptual scope and precision were incomplete (Criterion 3), and how primitive language acquired meaning (Criterion 4). From the perspective of knowledge transmission, crude primitive language could also play a crucial role in teaching. Through simple gestural cues, transmitters could direct observers' attention to key points for imitation learning, greatly improving knowledge transmission efficiency. Additionally, during the teaching process, language could more easily acquire meaning through actual actions. For example, a teacher could first make a gesture representing "grinding," then pick up a stone and rub its edge, allowing the observer to understand the gesture's meaning (Dean et al., 2014).

The remaining three criteria are: Why can human language conceptualize things across different domains and scopes (Criterion 5)? Why did only humans evolve language (Criterion 6)? And why does language communication require postnatal learning (Criterion 7)? The Teaching Hypothesis also satisfies these three criteria. When language first appeared, it may have been primarily used to transmit knowledge about tool-making, food processing, and hunting-gathering. This ability greatly improved human survival adaptability while accelerating knowledge updating speed. As human knowledge and skills expanded into various domains, language's scope also continuously broadened. Other animals did not evolve language because they lacked sufficiently complex tools and technologies, and their teaching needs were far smaller than humans'. Without the pressure of teaching activities, animals would not and did not need to evolve psychological mechanisms like language that could greatly improve teaching efficiency (Street et al., 2017). Additionally, after language emerged, knowledge innovation and transmission entered a positive feedback loop. The more efficiently knowledge could be transmitted, the more easily it could be iteratively updated. Correspondingly, the objects language referred to became increasingly complex and diverse. Under these circumstances, postnatal language learning could ensure that human ancestors flexibly mastered new linguistic symbols.

Although the Teaching Hypothesis does not specifically emphasize the relationship between hand actions and language origins, this hypothesis also supports the view that gestures were the earliest form of language. The reason is that from a biological structure perspective, early humans did not yet possess the conditions for vocal control and vocal learning (Fitch, 2017). For example, archaeological research has found that due to thoracic nerve bundle limitations, early *Homo erectus* living 1-2 million years ago could not produce finely varied vocal sounds (Walker & Leakey, 1993), while the earliest stone tools unearthed date back 3.3 million years (Harmand et al., 2015). Therefore, Laland (2017b) also believes that when human ancestors began the process of knowledge accumulation, they did not possess the ability to communicate through vocal sounds. Under these circumstances, teaching activities led to major transformations in humans' ability to communicate through gestures first.

It should be noted that although the Mirror System Hypothesis, Tool-making Hypothesis, and Teaching Hypothesis each emphasize different aspects, they are not actually contradictory. All three hypotheses agree that language is related to early human hand actions and share the same view on the developmental trajectory of human language—that is, human unique language evolution began with gestural language. However, the Mirror System Hypothesis focuses on explaining why gestures can represent concepts from the perspective of the mirror neuron system, clarifying the physiological basis necessary for language evolution. It answers the proximate mechanism question of "what" in language evolution. The Teaching Hypothesis emphasizes the pressure knowledge transmission (especially tool-making-related knowledge) exerted on language evolution, explaining why natural selection "chose" language as a psychological mechanism in humans. It answers the ultimate mechanism question of "why" in language evolution. The Tool-making Hypothesis incorporates both ultimate and proximate perspectives, emphasizing both that tool-making and use provided opportunities for language evolution and analyzing the role of brain neural circuits related to manual operation functions in developing language neural mechanisms. The three hypotheses describe language evolution at different levels and demonstrate good complementarity, with empirical research providing support for these hypotheses.

Consistent Features Between Sign Language and Spoken Language

Can hand movements constitute an independent communication method? Until the mid-21st century, sign language was not considered a true language. Many linguists believed that compared to the symbolic systems of spoken or written language, sign language lacked precision, subtlety, and flexibility, and could not represent abstract concepts and thinking. Hockett (1960) summarized 13 features of human language, with the first being that it is based on the sound-auditory channel. However, DeMatteo's (1977) research on American Sign Language found that sign language's linguistic structure closely resembles that of spoken language. Subsequent in-depth research by many scholars fundamentally changed the views of linguistics, psychology, and education on sign language.

After decades of systematic exploration, researchers are now convinced that hand movements themselves can constitute a complete communication method. Sign language used by deaf individuals includes grammatical rules as well as lexical, sentential, and discourse structures, and can decompose and combine symbolic codes. Therefore, sign language can express rich and complex meanings just like any other language (Brentari & Coppola, 2013; Goldin-Meadow & Brentari, 2017). In terms of learning mechanisms, no significant differences exist between sign language acquisition and spoken language acquisition. If children are exposed to sign language environments from an early age, they can effortlessly master sign language expression just as other children acquire spoken language (Garcia, 2002; Petitto, 1992). Additionally, sign language shares partially similar neural substrates with spoken language at both production and comprehension levels (Perniss, Özyürek, & Morgan, 2015). For example, when deaf individuals comprehend sign language words, Broca's area and Wernicke's area are also activated. Once deaf individuals suffer damage to Broca's area due to stroke or other accidents, their sign language expression similarly exhibits symptoms of "motor aphasia"—they can no longer produce meaningful signs (Pinker, 2007).

Survey research shows that without any adult intervention or guidance, children can spontaneously develop highly expressive natural sign language in daily interactions. In the late 1980s, the Nicaraguan government established two schools for deaf children in Managua. The government's goal was to help deaf children learn to communicate through sign language, but unfortunately, teaching by inexperienced teachers was unsuccessful. However, the students gradually created a set of sign language through daily collective interactions, which subsequent students continuously improved. After research and analysis, sign language expert Kegl (2002) confirmed that this was a very mature linguistic system, unique and created independently by Nicaraguan deaf children. Many researchers believe that these deaf children's experiences completely reproduced the evolutionary process of language from nonexistence to existence, suggesting that human ancestors may have similarly developed the ability to communicate through gestures in group living, only later gradually replacing actions with sounds (Senghas & Coppola, 2001; Senghas, Senghas, & Pyers, 2005). Other studies have also found that if deaf children do not systematically learn sign language from an early age, they will independently develop a homesign system, which also possesses grammatical structures and rules (Hunsicker & Goldin-Meadow, 2012). In summary, these studies demonstrate that humans are innately endowed with the potential to communicate through gestures. Sign language possesses the key features of human language—representational capacity, plasticity, structure, and openness. Hand movements can constitute a complete and independent communication system, providing indirect support for the gestural origin view of human language. However, more direct evidence is needed to prove the relationship between language evolution and hand actions.

Common Neural Mechanisms of Manual Operation and Language Processing

Exploring the brain neural mechanisms of language represents an important research approach to understanding language evolutionary origins. According to the Neural Reuse hypothesis, the cognitive functions that different brain regions can support are not singular. During evolution or individual development, brain regions originally supporting one function can support the development of new, higher-level cognitive functions. New cognitive functions are integrated into established brain regions, and brain neural circuits can implement various new cognitive abilities through reuse (Anderson, 2010). Both the Mirror System Hypothesis and Tool-making Hypothesis emphasize that the neural mechanisms of language functions developed from brain neural circuits related to manual operation, which aligns perfectly with the Neural Reuse hypothesis. Cognitive neuroscience research provides ample evidence for this view. In fact, the earliest mirror neurons discovered in macaques related to hand action recognition were located in the ventral premotor cortex F5 area, which is homologous to Broca's area in the human brain responsible for language functions (Arbib & Rizzolatti, 1997). In recent years, numerous studies have further demonstrated that hand movements (especially complex actions related to tool operation) and language processing share common brain neural substrates, supporting the claims of both the Mirror System Hypothesis and Tool-making Hypothesis.

For example, Farsin et al. (2003) used functional magnetic resonance imaging (fMRI) to record and analyze brain activation patterns when subjects performed action recognition, language generation, and grasping movements. The study found that subjects activated some common brain cortices when performing these tasks, including the parietal lobe, left inferior frontal gyrus, and precentral gyrus. Stout et al. (2008) conducted fMRI scans on stone tool experts and found that toolmaking activated Broca's area in their brains, which is closely related to language functions. Higuchi, Chaminade, Imamizu, and Kawato (2009) also found that Broca's area activity significantly increased when subjects listened to stories or performed tool operation tasks. Uomini and Meyer (2013) required subjects to complete two tasks in an experiment: first, making Acheulean handaxes, where subjects had to perform interfering actions at intervals before continuing to knap stone axes; second, completing word test tasks according to auditory cues. Researchers used functional transcranial Doppler ultrasonography (fTCD) to measure subjects' cerebral blood flow patterns, finding that subjects produced highly consistent cerebral blood flow patterns during the first 10 seconds of both tasks.

These studies demonstrate that manual operation and language processing share partially overlapping neural mechanisms. According to the Tool-making Hypothesis, the reason manual operation and language processing share neural mechanisms is that hand actions related to tool-making became functionally coupled with the communication system. Therefore, the cognitive characteristics of language can reflect the cognitive characteristics of tool-making. The function of Broca's area provides support for this hypothesis. Research has found that besides being involved in language structure (grammar) processing, Broca's area is also related to processing hierarchical structure and sequential order—both necessary abilities for completing complex manual operation tasks (Clerget, Andres, & Olivier, 2013; Koechlin & Jubault, 2006; Pammi et al., 2012; Ullman, 2006). In other words, Broca's area is responsible for processing both language grammar and the "action grammar" of complex hand operations. For instance, Clerget, Winderickx, Fadiga, and Olivier (2009) used repetitive transcranial magnetic stimulation to temporarily interfere with Broca's area in 13 healthy subjects while requiring them to sequence actions displayed on a computer. The results showed this manipulation affected subjects' performance on action sequencing tasks, indicating that Broca's area plays a crucial role in encoding complex human action "grammar." Alamia et al. (2016) found that applying transcranial magnetic stimulation to Broca's area prolonged subjects' processing time for mentally representing higher-order chunks in experimental tasks, further demonstrating Broca's area's role in processing hierarchical structures. Other researchers using different methods have confirmed similar conclusions (Meyer, Greenlee, & Wuerger, 2011; Pritchett et al., 2018).

However, some scholars argue that since the inferior frontal gyrus where Broca's area is located is one of the most functionally heterogeneous brain regions, merely observing that language processing and manual operation both activate Broca's area does not sufficiently prove that these two activities share common neural mechanisms. To more accurately explore the relationship between action observation/execution and language processing in the cerebral cortex, Pritchett et al. (2018) first required subjects to read normal sentences and sentences composed of meaningless words in an experiment, determining each subject's precise language neural network through a functional localizer task. They then examined the involvement of individuals' language-related neural networks during action imitation and action observation tasks. The study found that language processing networks did not show strong responses beyond baseline during action observation or imitation. Additionally, Weiss et al. (2016) studied the relationship between language disorders and action disorders by examining aphasia and apraxia patients. They found that lesions in the left inferior frontal gyrus, insula, parietal lobe, and superior temporal cortex were related to language deficits in naming, reading, writing, or auditory comprehension, while lesions in the left inferior frontal gyrus, premotor area, central region, and inferior parietal cortex were related to loss of tool operation ability. Therefore, while Broca's area damage can indeed cause both apraxia and aphasia, other brain regions causing aphasia or apraxia do not otherwise overlap.

The Relationship Between Gesture and Language Ability Development

The Mirror System Hypothesis, Tool-making Hypothesis, and Teaching Hypothesis all consistently maintain that human language originated from manual movement experiences. Early human brains first evolved the ability to understand gestural language, after which sign language developed into other language forms such as vocal sounds and writing. Since the evolution of some human psychological mechanisms can be reflected in individual development processes, if this view is correct, then it is likely that in individual development, gesture use both precedes language emergence and has an inseparable relationship with language ability development (Salo, Ferrari, & Fox, 2019).

This hypothesis has currently received support from numerous empirical studies. Research has found that infants' and toddlers' ability to use gestures can predict their future language development levels. For example, 1-year-olds who can produce more gestures during parent-child interactions (including conveying more meanings through gestures and more frequently using deictic gestures) will master more vocabulary at ages 3-4 (Rowe & Goldin-Meadow, 2009) and will master complex sentences earlier (Rowe & Goldin-Meadow, 2009). Kuhn et al. (2014) found that parents' reports of infants' gesture use at 15 months could predict their expressive abilities at ages 2-3, with other studies finding similar results (Cadime et al., 2017; Zambrana et al., 2013). Notably, this pattern also exists in children with autism or Down syndrome (Özçalışkan et al., 2017).

Lüke et al. (2017) found that children who used fewer deictic gestures at 12-14 months showed signs of language development delay at age 2. Additionally, Salo, Rowe, and Reeb-Sutherland (2018) assessed infants' gesture use frequency and joint attention levels at age 1, then tested their language comprehension and expression abilities one year later. Analysis showed that even after controlling for joint attention levels, infants' gesture use could still significantly predict their language abilities.

Salo et al. (2019) argue that these findings demonstrate that before speaking, infants can interact with the external world and produce communicative gestures through hand actions, and these early mental representations lay the foundation for their later language development. Consistent with this view, research has found that gestures can indeed help people acquire language and deepen individuals' conceptual understanding during language learning (Kontra, Goldin-Meadow, & Beilock, 2012; Mumford & Kita, 2014; Novack & Goldin-Meadow, 2017; Thill & Twomey, 2016). For instance, Dean et al. (2013) observed 23 teaching instances and found that auxiliary gestures and actions could significantly promote phonetic learning. Ianì and Bucciarelli (2017) conducted an experiment requiring subjects to listen to a performer speak target sentences, with the performer either remaining motionless or making gestures consistent with sentence content while speaking. The study showed that when subjects observed actions, they had more accurate memory for sentences. However, if subjects were busy making certain hand actions themselves while listening to sentences, this effect would be canceled out. In other words, during language comprehension, gestures affect individuals' processing of language information. Wakefield, Novack, Congdon, Franconeri, and Goldin-Meadow (2018) used eye-tracking experiments and found that in vocabulary teaching, compared to speech-only teaching modes, "speech + gesture" teaching modes produced better learning outcomes in children. Researchers believe that because gestures have intuitive and concrete features, they help learners connect self-perceived motor experiences with concepts, thus facilitating language learning. Therefore, early humans first formed the ability to represent concepts through gestures.

Additionally, Wakefield, Hall, James, and Goldin-Meadow (2018) divided child subjects into four groups in an experiment. Two groups could watch the experimenter perform actions on related objects or perform actions on objects themselves when learning verbs—for example, holding a rotatable teaching tool when learning the word "knob." The other two groups could watch the experimenter make simulated gestures above related objects or make simulated gestures above objects themselves when learning verbs—for example, rotating their hand above a teaching tool without touching it when learning "knob." A recognition test was completed one day later, requiring children to name actions presented in videos. The study found that subjects in experimental groups who watched or made gestures performed better on conceptual test tasks. Researchers believe that since gestures do not make direct contact with objects, learning through gestures makes it easier for subjects to generalize previously learned vocabulary to new scenarios. Gesture experience is an effective way to represent action concepts, and these studies also provide indirect evidence for the hypothesis that human language originated from hand actions.

However, researchers have also found that hand actions can promote processing and understanding in other learning tasks unrelated to language. For example, Goldin-Meadow, Cook, and Mitchell (2009) found that using gestures helps subjects generate new creative ideas in problem-solving tasks. Kontra, Lyons, Fischer, and Beilock (2015) found that manual operation experience can enhance students' learning effects on physical science concepts. Novack, Congdon, Hemani-Lopez, and Goldin-Meadow (2014) showed that in mathematical problem-solving tasks, gestural simulation helps third-grade students transfer mathematical knowledge to different problem forms. This indicates that the facilitative effect of gestures on cognitive processing is universal and not limited to the language domain.

The Influence of Gesture on the Transmission of Tool-Making Knowledge

The Tool-making Hypothesis and Teaching Hypothesis posit that manufacturing and using tools exerted evolutionary pressure on language, leading to the emergence of gestural language. If this view is correct, then in the process of learning to make tools, gestural communication should greatly improve information transmission efficiency compared to simple observation and imitation. In fact, archaeologists have determined that stone flakes from the *Homo erectus* era already had sharp edges, and manufacturing such flakes had strict requirements for material selection, separation, cutting, and grinding—skills difficult to master through observation and imitation alone (Chazan, 2015).

Morgan et al. (2015) tested the learning effects of five different information transmission mechanisms in an experiment. Subjects learned to manufacture stone tools through five different methods: 1) Observing finished products—learners could only see completed tools but not the manufacturing process; 2) Imitation learning—learners could see the manufacturing process but could not communicate with the instructor; 3) Guided teaching—instructors could slow down or repeat certain actions according to learners' requests for specific demonstrations but could not communicate; 4) Gestural teaching—learners and instructors could communicate through gestures; 5) Language teaching—learners and instructors could communicate through language. In each condition, experimenters arranged four short transmission chains (5-person transmission) and two long transmission chains (10-person transmission). Two hundred adults participated in this experiment, producing over 6,000 flint pieces. The study found that in all conditions, stone tool-making information was continuously lost during transmission, but communicative teaching could fully compensate for this deficiency in the transmission chain. In the gestural teaching and language teaching groups, learners' quality and efficiency in making flint pieces improved significantly, with the language teaching group producing the most refined flint pieces. These experimental results demonstrate that compared to simple imitation or observation, gestural communication can substantially improve the transmission efficiency of tool-making technology. Since archaeological research proves that when human ancestors began making tools, they did not possess the ability to communicate through vocal sounds, the pressure of technological transmission caused human ancestors to first achieve major transformations in gestural language.

Lombao, Guardiola, and Mosquera (2017) used similar research to test the impact of communication methods on stone tool manufacturing, dividing subjects into imitation, gestural communication, and language communication groups. Results showed that on efficiency measures, the imitation group's production speed was significantly lower than that of the gestural and language communication groups. On quality measures, subjects in the gestural and language communication groups could master correct manufacturing techniques and methods like stone tool experts, while the imitation group could not. The results similarly prove that compared to other information transmission mechanisms (such as simple observation), gestural and language communication can transmit more information in a short time, enabling individuals to more effectively acquire knowledge about tool-making. Therefore, once human ancestors became highly dependent on tools, natural selection could "select" for complex gestural communication abilities.

Additionally, Putt, Woods, and Franciscus (2014) found that in learning to make Acheulean stone tools, there were no differences in key indicators between stone tools made by subjects in the spoken language communication group and those in the gestural communication group. To clarify this conclusion, Cataldo, Migliano, and Vinicius (2018) divided subjects into four groups in a stone tool-making experiment: a gestural communication group (could only use sign language to communicate with experts), a spoken language communication group (could only use spoken language to communicate with experts), a complete communication group (could use both gestures and spoken language to communicate with experts), and an observation-imitation group (could only observe experts' methods). The study found that compared to the gestural communication or complete communication groups, the observation-imitation and spoken language communication groups performed worse in stone tool-making and had lower satisfaction with expert guidance. Researchers speculate that this result indicates that for manufacturing stone tools, spoken language communication alone does not have any advantage over gestural communication.

Fay, Lister, Ellison, and Goldin-Meadow (2014) also demonstrated that when subjects were asked to create symbols representing target objects, gestures were more easily understood by others than vocal sounds (self-simulated sounds, not linguistic sounds). Moreover, combining gestural symbols with vocal symbols did not improve comprehension accuracy. Therefore, hand actions may be more easily selected as information transmission methods for early humans than spoken language, making gestural language the precursor to human language. Other evolutionary pressures (such as the need for cooperative reciprocity or more complex interpersonal interactions) subsequently led to the emergence of vocal language.

Review and Outlook

Human daily life is intimately connected with language; few people can go for a continuous hour while awake without encountering any form of language. Wherever on Earth two or more people gather, they soon begin conversing. From the moment our ancestors mastered language, human evolution embarked on a different path. Through language, every individual can easily access the treasure trove of knowledge collectively created by all humanity, enabling human civilization to truly enter an era of great leaps forward. However, like other complex physiological mechanisms that have undergone long evolution, human language as a complex psychological mechanism did not emerge in its current "perfect" form but gradually experienced many different developmental stages under natural selection pressure. This paper focused on elaborating the relationship between hand actions and human language evolution, introducing related theories that explain both why humans "needed" to evolve language and why humans "could" evolve language. Simultaneously, this paper summarized and organized empirical research in this field, which has verified related hypotheses from different angles. Future research must focus on the following issues.

The Evolutionary Developmental Relationship Between Sign Language and Vocal Language First, the relationship between sign language and vocal language in evolution requires further exploration. In language research, whether the earliest language originated from gestures or vocal sounds has long been a subject of debate. Since neither gestures nor vocal sounds leave direct archaeological evidence, clarifying this disagreement presents considerable difficulty. The theories and empirical research introduced in this paper all indicate that human language origins are closely related to hand actions. However, one question this hypothesis must answer is: If language originally derived from gestures, why did it later shift to the "vocal" mode of expression during evolution? Vocal language is modern humans' primary communication form. Since sign language can also have complete grammatical rules, linguistic structures, and rich expressiveness, why did spoken language replace sign language to become the most universal language form? Speaking is not an easy task. To produce accurate and clear pronunciation, humans evolved unique larynx and vocal cord shapes, but this also causes food to easily slip into the trachea during eating, causing choking. Therefore, spoken language must have produced benefits sufficient to offset this risk. Compared to sign language, spoken language does have some obvious advantages, such as higher information transmission rates, independence from lighting conditions, and the ability to communicate with others while the hands are busy with other tasks. However, these apparent advantages may not necessarily constitute selective pressure for spoken language evolution, because in evolutionary processes, the outcomes of a psychological mechanism do not equate to its evolutionary causes. Currently, all hypotheses supporting a gestural origin for language have not provided reasonable theoretical explanations for the evolution of vocal language, and relevant empirical research is basically nonexistent.

Another derived question is: If spoken language replaced gestural language due to some selective pressure during evolution, why has the gestural communication form been preserved and frequently appears in daily conversation? Research has found that gestures are not merely epiphenomena of spoken language expression; gestures help reduce ambiguity in discourse information and can complementarily convey information that speech cannot express, thereby promoting complete semantic comprehension by information recipients (Goldin-Meadow & Brentari, 2017; Koppensteiner et al., 2016). For information senders, gestural expression also helps improve communication fluency (Brooks & Goldin-Meadow, 2016). This indicates that gestural language remains an important information representation method. McNeill (2012) suggests that gestural communication persists because human language representation requires mimetic encoding, which symbolic language cannot satisfy. Spoken or written language represents abstract symbols that need to be connected with concrete perceptual-motor experiences. In other words, abstract symbolic encoding and concrete mimetic encoding together constitute a complete language representation system. Therefore, the source of human language is not limited to gestural language alone; gestural language and spoken language experienced co-evolution and joint development.

Correspondingly, paleoanthropological research has also found that in the evolutionary history of the genus *Homo*, the ability to speak appeared earlier than the emergence of mature language systems (referring to languages with complex grammatical structures). For example, analyses of Neanderthal genes, brain structures, and hyoid bone fossils indicate that due to genetic and brain structure limitations, Neanderthals' language systems were not as complete as modern humans', but they already possessed the ability to finely control vocal sounds. If not for using spoken language for communication, they would be unlikely to have developed this characteristic (Kochiyama et al., 2018; Maricic et al., 2013). This demonstrates that humans did not first develop complex grammar on the basis of gestural language and then "graft" language ability onto the vocal expression form. In fact, many scholars now believe that when explaining language origins, gestural language and spoken language should not be opposed, assuming one communication form preceded the other. It is likely that these two language forms appeared simultaneously and experienced co-evolution (Goldin-Meadow & Brentari, 2017; Kendon, 2017)—this is the multimodal hypothesis of language origins. For instance, Kendon (2017) believes that in early language evolution, humans formed a "hand-mouth" coordinated multimodal signal system. Fröhlich et al. (2019) point out that although early humans did not yet possess the ability to finely control vocal sounds, the information content they represented was not as complex as contemporary humans'. Simple vocalization abilities could cooperate with gestures to reduce communication ambiguity. Other apes flexibly choose different communication methods such as gestures, calls, or "gesture + calls" according to situational needs, and human ancestors likely did the same. Even many researchers supporting gestural origins do not completely deny the existence of spoken language at the beginning of language evolution (Arbib, 2016; Tomasello, 2008). The specific evolutionary mechanisms require further exploration in future research.

The Relationship Between Human Language Mechanisms and Other Psychological Features in Evolution Another issue requiring consideration is the developmental relationship between human language mechanisms and other psychological features in evolution. The foundations of language evolution emphasized by the Mirror System Hypothesis, Tool-making Hypothesis, and Teaching Hypothesis—such as communicating through gestures, recognizing other individuals' action intentions, and information transmission abilities—also exist in great apes (though differing quantitatively from humans). However, why haven't other apes developed complex and complete language like humans based on these abilities?

In fact, since the 1960s, people have continuously attempted to teach great apes such as chimpanzees, orangutans, or gorillas to learn sign language. Although specific results remain controversial, it is now certain that most apes raised in human environments can learn to interact with humans using some gestures, and they can even spontaneously create some actions to express their demands to humans without special training. For example, when apes are caged, they will extend fingers or palms outside the cage toward food, gesturing for experimenters to help them obtain it. If they want to go outside the cage to play, they will also gesture toward the locked door for experimenters to open it (Pepperberg, 2016).

However, regardless of the environment in which apes grow, they cannot develop more sophisticated gestures beyond requests or commands. They will not point to an object to inform others of information they might want to know. Numerous experiments have shown that chimpanzees can understand other chimpanzees' intentions; they can imagine how others will think and act, but they do not consciously share information. In one experiment, chimpanzees could either pull a mechanism to obtain one portion of food for themselves or pull another mechanism to obtain one portion each for themselves and another chimpanzee. Although the latter behavior did not require them to pay any additional cost, the chimpanzees' choices in the experiment were quite arbitrary—they did not consider benefiting other individuals (Silk et al., 2005). In contrast, research has found that 12-month-old infants can demonstrate unconditional altruistic behavior. After detecting certain needs in adults, infants will point with their fingers to items the adults need, helping them find them (Liszkowski, Carpenter, & Tomasello, 2008). One-year-old infants will point to interesting things to share with their parents (Warneken & Tomasello, 2006). Therefore, although both apes and humans use gestures, they differ essentially in function.

The ordinary human action of "pointing with a finger" contains important evolutionary significance, demonstrating the altruistic motivation behind human communication. From an evolutionary perspective, this fact implies that sign language development is closely related to social cooperation. Apes have not developed more complex sign language systems partly because they lack cooperative reciprocity. In contrast, humans possess "shared intentionality," actively teaching each other useful things, and the mindset of sharing information constitutes an important foundation for language. In fact, the language mechanism itself is a complex system that includes various complex cognitive abilities, brain structures, and genetic foundations, with no single factor alone being able to define the human communication system—they function as an integrated whole. Therefore, when discussing language evolution and origins, it is essential to comprehensively consider other unique cognitive features of humans to explain the specificity of human language evolution compared to other animals.

Integrating Multidisciplinary Research Findings to Construct a Unified Theoretical Framework Finally, returning to the field of language evolution research itself, the current challenge is how to unify existing research findings under a complete theoretical framework. Language evolution involves multiple disciplines including genetics, archaeology, linguistics, paleoanthropology, cognitive neuroscience, psychology, and comparative zoology, with progress in each discipline providing unique contributions to unraveling the mystery of language evolution. For example, comparative zoology research shows that primate group size is closely related to communication systems. Larger group sizes, social networks, and diverse social relationships require more complex communication systems. Therefore, one pressure for language evolution may have come from the need to maintain intimate relationships, preserve group cohesion, and establish alliances (Dunbar, 1998; Feinberg, Willer, & Schultz, 2014). Cognitive neuroscience has discovered that learning to make stone tools leads to changes in individual brain connections and brain tissue, including the supramarginal gyrus of the parietal lobe and the right inferior frontal gyrus, brain regions related to advanced cognitive functions such as language communication, planning, and logical reasoning (Hecht et al., 2015; Morgan et al., 2015). Genetics research has found that gene variations such as NOTCH2NL (Fiddes et al., 2018), ARHGAP11B (Florio et al., 2015), ASPM (Mekelbobrov et al., 2005), MCPH1 (Evans, 2005; Huang et al., 2019), and GADD45G (McLean et al., 2011) in different periods of human evolutionary history dominated the increase in human brain capacity and changes in cortical structure, enabling the human brain to support complex psychological mechanisms like language. Archaeological research has discovered that between 30,000 and 40,000 years ago, humans suddenly experienced an unprecedented artistic explosion, beginning to draw pictures, carve sculptures, wear jewelry, and developing primitive religions and mythological cultures. The time when human ancestors mastered mature language could not have been much earlier than this period; mature language systems are the exclusive privilege of *Homo sapiens* (Laland, 2017a). Linguistic research proves that the farther a language is from Africa, the fewer phonemes it uses to generate vocabulary, showing a founder effect with Africa as the origin point. Therefore, all languages in the world are homologous; language diffused and differentiated along with human population migration during the process of *Homo sapiens* taking over the Earth (Atkinson, 2011). These research findings from different disciplines can mutually corroborate and compare, providing powerful tests for certain language evolution hypotheses. Only by collecting and integrating these findings from different disciplines can a complete theoretical framework be constructed to provide systematic explanations for language evolution and transformation.

References

Alamia, A., Solopchuk, O., D'Ausilio, A., Van Bever, V., Fadiga, L., Olivier, E., & Zénon, A. (2016). Disruption of Broca's area alters higher-order chunking processing during perceptual sequence learning. *Journal of Cognitive Neuroscience, 28*(3), 402–417.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral & Brain Sciences, 33*(4), 275–276.

Arbib, M. A. (2010). Mirror system activity for action and language is embedded in the integration of dorsal and ventral pathways. *Brain & Language, 112*, 12–24.

Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis.* New York, NY: Oxford University Press.

Arbib, M. A. (2016). Towards a computational comparative neuroprimatology: Framing the language-ready brain. *Physics of Life Reviews, 16*, 1–54.

Arbib, M. A. (2017). Toward the language-ready brain: Biological evolution and primate comparisons. *Psychonomic Bulletin & Review, 24*, 142–150.

Arbib, M. A., & Rizzolatti, G. (1997). Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition, 29*, 393–424.

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science, 332*(6027), 346–349.

Bickerton, A. (2009). *Adam's Tongue.* New York, NY: Hill & Wang.

Bonini, L. (2017). The extended mirror neuron network: Anatomy, origin, and functions. *The Neuroscientist, 23*(1), 56–67.

Brentari, D., & Coppola, M. (2013). What sign language creation teaches us about language. *Wiley Interdisciplinary Reviews (WIREs): Cognitive Science, 4*, 201–211.

Brooks, N., & Goldin-Meadow, S. (2016). Moving to learn: How guiding the hands can set the stage for learning. *Cognitive Science, 40*(7), 1831—1849.

Cadime, I., Silva, C., Santos, S., Ribeiro, I., & Viana, F. L. (2017). The interrelatedness between infants' communicative gestures and lexicon size: A longitudinal study. *Infant Behavior and Development, 48*, 88–97.

Cataldo, D. M., Migliano, A. B., & Vinicius, L. (2018). Speech, stone toolmaking and the evolution of language. *PLOS ONE, 13*(1), e0191071.

Chazan, M. (2015). Technological trends in the Acheulean of Wonderwerk Cave, South Africa. *African Archaeological Review, 32*(4), 701–728.

Chen, W. (2019). Is empathy equal to mirroring?—Theoretical compatibility and controversies between mirror neurons and phenomenology. *Philosophical Research, 06*, 96–107.

Chomsky, N. (2002). *On Nature and Language*. Cambridge, England: Cambridge University Press.

Clerget, E., Andres, M., & Olivier, E. (2013). Deficit in complex sequence processing after a virtual lesion of left BA 45. *PLOS ONE, 8*, 1–9.

Clerget, E., Winderickx, A., Fadiga, L., & Olivier, E. (2009). Role of Broca's area in encoding sequential human actions: A virtual lesion study. *Neuroreport, 20*(16), 1496–1499.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain & Language, 112*, 25–35.

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., . . . Brass, M. (2018). Automatic imitation: A meta-analysis. *Psychological Bulletin, 144*(5), 453–500.

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2013). Identification of the social and cognitive processes underlying human cumulative culture. *Science, 335*(6072), 1114–1118.

Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A comparative perspective. *Biological Reviews, 89*(2), 284-301.

DeMatteo, A. (1977). Visual imagery and visual analogues in American Sign Language. In L. Friedman (Ed.), *On the Other Hand* (pp. 109-137). New York: Academic Press.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91*, 176–180.

Dor, D. (2017). The role of the lie in the evolution of human language. *Language Sciences, 63*, 44-50.

Dunbar, R. (1998). *Groups, Gossip, and the Evolution of Language*. Journal of the History of the Behavioral Sciences, 34(4), 398-399.

Evans, P. D., Gilbert, S. L., Mekel-Bobrov, N., Vallender, E. J., Anderson, J. R., Baez-Azizi, L. M., . . . Lahn, B. T. (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science, 309*, 1717–1720.

Farsin, H., Michel, R., Christian, D., Volkmar, G., Cornelius, W., & Christian, B. (2003). The human action recognition system and its relationship to Broca's area: An fMRI study. *NeuroImage, 19*(3), 637–644.

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a communication system from scratch: Gesture beats vocalization hands down. *Frontiers in Psychology, 5*(354).

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science, 25*(3), 656–664.

Fiddes, I. T., Lodewijk, G. A., Mooring, M., Bosworth, C. M., Ewing, A. D., Mantalas, G. L., . . . Haussler, D. (2018). Human-specific NOTCH2NL genes affect notch signaling and cortical neurogenesis. *Cell, 173*(6), 1356–1369.

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

Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review, 24*(1), 3–33.

Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., . . . Huttner, W. B. (2015). Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science, 347*(6229), 1465–1470.

Fogassi, L., & Ferrari, P. F. (2007). Mirror neurons and the evolution of embodied language. *Current Directions in Psychological Science, 16*(3), 136–141.

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews, 94*(5), 1809–1829.

Garcia, J. (2002). *Sign With Your Baby: How to communicate with infants before they can speak*. USA: Northlight Communications.

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences, 40*, e46.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20*, 267–272.

Gould, S. J. (1991). Exaptation: A Crucial Tool for an Evolutionary Psychology. *Journal of Social Issues, 47*(3), 43–65.

Harmand, S., Lewis, J. E., Feibel, C. S., Lepre, C. J., Prat, S., Lenoble, A., . . . Arroyo, A. (2015). 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature, 521*(7552), 310.

Hecht, E. E., Gutman, D. A., Khreisheh, N., Taylor, S. V., Kilner, J., Faisal, A. A., . . . Stout, D. (2015). Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. *Brain Structure & Function, 220*(4), 2315–2331.

Higuchi, S., Chaminade, T., Imamizu, H., & Kawato, M. (2009). Shared neural correlates for language and tool use in Broca's area. *Cognitive Neuroscience and Neuropsychology, 20*, 1376–1381.

Hockett, C. F. (1960). The Origin of Speech. *Scientific American, 203*, 89–97.

Huang, J., Li, M., Lin, Q., Li, Y., Shi, L., Su, B., . . . Lu, Y. (2019). Transgenic rhesus monkeys carrying the human MCPH1 gene copies show human-like neoteny of brain development. *National Science Review, 6*(3), 480–493.

Hunsicker, D., & Goldin-Meadow, S. (2012). Hierarchical structure in a self-created communication system: Building nominal constituents in homesign. *Language, 88*(4), 732–763.

Ianì, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language, 96*, 110–121.

Kang, S., Tversky, B., & Black, J. B. (2015). Coordinating gesture, word, and diagram: Explanations for experts and novices. *Spatial Cognition & Computation, 15*(1), 1–26.

Kegl, J. (2002). Language emergence in a language-ready brain: Acquisition issues. In G. Morgan & B. Woll (Eds.), *Language acquisition in signed languages* (pp. 207–254). Cambridge: Cambridge University Press.

Kendon, A. (2017). Reflections on the "gesture-first" hypothesis of language origins. *Psychonomic Bulletin & Review, 24*, 163–170.

Kochiyama, T., Ogihara, N., Tanabe, H. C., Kondo, O., Amano, H., Hasegawa, K., . . . Bastir, M. (2018). Reconstructing the Neanderthal brain using computational anatomy. *Scientific Reports, 8*(1), 6296.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron, 50*, 963–974.

Kolodny, O., & Edelman, S. (2018). The evolution of the capacity for language: The ecological context and adaptive value of a process of cognitive hijacking. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1743), 20170052.

Kolodny, O., Lotem, A., & Edelman, S. (2015). Learning a generative probabilistic grammar of experience: A process-level model of language acquisition. *Cognitive Science, 39*(2), 227–255.

Kontra, C., Goldin-Meadow, S., & Beilock, S. L. (2012). Embodied learning across the life span. *Topics in Cognitive Science, 4*(4), 731–739.

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological Science, 26*(6), 737–749.

Koppensteiner, M., Stephan, P., & Jäschke, J. P. M. (2016). Moving speeches: Dominance, trustworthiness and competence in body motion. *Personality and Individual Differences, 94*, 254–259.

Kuhn, L. J., Willoughby, M. T., Wilbourn, M. P., Vernon-Feagans, L., Blair, C. B., & Investigators, T. F. L. P. K. (2014). Early communicative gestures prospectively predict language development and executive function in early childhood. *Child Development, 85*(5), 1898–1914.

Laland, K. N. (2017a). *Darwin's unfinished symphony: How culture made the human mind*. New Jersey: Princeton University Press.

Laland, K. N. (2017b). The origins of language in teaching. *Psychonomic Bulletin & Review, 24*, 225–231.

Li, J., & Ding, G. (2013). Comparison of brain mechanisms between sign language and spoken language comprehension and production. *Advances in Psychological Science, 21*(9), 1560–1569.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition, 108*(3), 732–739.

Lombao, D., Guardiola, M., & Mosquera, M. (2017). Teaching to make stone tools: New experimental evidence supporting a technological hypothesis for the origins of language. *Scientific Reports, 7*(1), 14394.

Lüke, C., Ritterfeld, U., Grimminger, A., Liszkowski, U., & Rohlfing, K. J. (2017). Development of pointing gestures in children with typical and delayed language acquisition. *Journal of Speech, Language, and Hearing Research, 60*(11), 3185–3197.

Maricic, T., Gunther, V., Georgiev, O., Gehre, S., Curlin, M., Schreiweis, C., . . . Lalueza-Fox, C. (2013). A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Molecular Biology & Evolution, 30*(4), 844–852.

Marstaller, L., & Burianová, H. (2014). The multisensory perception of co-speech gestures – A review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics, 30*, 69–77.

McLean, C. Y., Reno, P. L., Pollen, A. A., Bassan, A. I., Capellini, T. D., Guenther, C., . . . Kingsley, D. M. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature, 471*, 216.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution.* New York: Cambridge University Press.

Mekelbobrov, N., Gilbert, S. L., Evans, P. D., Vallender, E. J., Anderson, J. R., Hudson, R. R., . . . Lahn, B. T. (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science, 309*(5741), 1720–1722.

Meyer, G. F., Greenlee, M., & Wuerger, S. (2011). Interactions between auditory and visual semantic stimulus classes: Evidence for common processing networks for speech and body actions. *Journal of Cognitive Neuroscience, 23*(9), 2291–2308.

Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinardthuly, L., Street, S. E., Lewis, H. M., . . . Torre, I. D. L. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications, 6*(1), 6029–6029.

Mumford, K. H., & Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Development, 85*, 1181–1189.

Novack, M., Congdon, E., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science, 25*, 903–910.

Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review, 24*(3), 652–665.

Nowicki, S., & Searcy, W. A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology, 28*, 48–53.

Okanoya, K. (2017). Sexual communication and domestication may give rise to the signal complexity necessary for the emergence of language: An indication from songbird studies. *Psychonomic Bulletin & Review, 24*, 106–110.

Ondobaka, S., Kilner, J., & Friston, K. (2017). The role of interoceptive inference in theory of mind. *Brain and Cognition, 112*, 64–68.

Özçalışkan, Ş., Adamson, L. B., Dimitrova, N., & Baumann, S. (2017). Early gesture provides a helping hand to spoken vocabulary development for children with autism, Down syndrome and typical development. *Journal of Cognition and Development, 18*(3), 325–337.

Pammi, V. S. C., Miyapuram, K. P., Ahmed, Samejima, K., Bapi, R. S., & Doya, K. (2012). Changing the structure of complex visuo-motor sequences selectively activates the fronto-parietal network. *NeuroImage, 59*(2), 1180–1189.

Pepperberg, I. M. (2016). Animal language studies: What happened? *Psychonomic Bulletin & Review, 24*(1), 181–185.

Perniss, P., Özyürek, A., & Morgan, G. (2015). The influence of the visual modality on language structure and conventionalization: Insights from sign language and gesture. *Topics in Cognitive Science, 7*, 2–11.

Petitto, L. A. (1992). Modularity and constraints in early lexical acquisition: Evidence from children's early language and gesture. In M. Gunnar (Ed.), *Minnesota Symposium on Child Psychology* (Vol. 25, pp. 25–58). Hillsdale, NJ: Erlbaum.

Pinker, S. (2007). *The language instinct: How the mind creates language*. New York, NY: Harper Perennial Modern Classics.

Pritchett, B. L., Hoeflin, C., Koldewyn, K., Dechter, E., & Fedorenko, E. (2018). High-level language processing regions are not engaged in action observation or imitation. *Journal of Neurophysiology, 120*(5), 2555–2570.

Putt, S. S., Woods, A. D., & Franciscus, R. G. (2014). The role of verbal interaction during experimental bifacial stone tool manufacture. *Lithic Technology, 39*(2), 96–112.

Redcay, E., Ludlum, R. S., Velnoskey, K. R., & Kanwal, S. (2016). Communicative signals promote object recognition memory and modulate the right posterior STS. *Journal of Cognitive Neuroscience, 28*(1), 8–19.

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science, 323*, 951–953.

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science, 12*, 182–187.

Salo, V. C., Ferrari, P. F., & Fox, N. A. (2019). The role of the motor system in action understanding and communication: Evidence from human infants and non-human primates. *Developmental Psychobiology, 61*(3), 390–401.

Salo, V. C., Rowe, M. L., & Reeb-Sutherland, B. C. (2018). Exploring infant gesture and joint attention as related constructs and as predictors of later language. *Infancy, 23*(3), 432–452.

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science, 12*(4), 323–328.

Senghas, R. J., Senghas, A., & Pyers, J. E. (2005). The emergence of Nicaraguan Sign Language: Questions of development, acquisition and evolution. In S. T. Parker, J. Langer, & C. Milbrath (Eds.), *Biology and knowledge revisited: From neurogenesis to psychogenesis* (pp. 287–306). Mahwah, NJ: Erlbaum.

Silk, J. B., Brosnan, S. F., Vonk, J., Henrich, J., Povinelli, D. J., Richardson, A. S., . . . Schapiro, S. J. (2005). Chimpanzees are indifferent to the welfare of unrelated group members. *Nature, 437*(7063), 1357–1359.

Smit, H. (2014). *The social evolution of human nature*. Cambridge, UK: Cambridge University Press.

Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1585), 75–87.

Stout, D., Passingham, R., Frith, C., Apel, J., & Chaminade, T. (2011). Technology, expertise and social cognition in human evolution. *European Journal of Neuroscience, 33*(7), 1328–1338.

Stout, D., Toth, N., Schick, K., & Chaminade, T. (2008). Neural correlates of Early Stone Age toolmaking: Technology, language and cognition in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1499), 1939–1949.

Street, S. E., Navarrete, A. F., Reader, S. M., & Laland, K. N. (2017). Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proceedings of the National Academy of Sciences, 114*(30), 7908–7914.

Szamado, S., & Szathmary, E. (2006). Selective scenarios for the emergence of natural language. *Trends in Ecology and Evolution, 21*, 555–561.

Thill, S., & Twomey, K. E. (2016). What's on the inside counts: A grounded account of concept acquisition and development. *Frontiers in Psychology, 7*, 402–402.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT Press.

Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex, 42*(4), 480-485.

Uomini, N. T., & Meyer, G. F. (2013). Shared brain lateralization patterns in language and Acheulean stone tool production: A functional transcranial Doppler ultrasound study. *PLOS ONE, 8*(8), e72693.

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science, 21*(6), e12664.

Wakefield, E. M., Hall, C., James, K. H., & Goldin-Meadow, S. (2018). Gesture for generalization: Gesture facilitates flexible learning of words for actions on objects. *Developmental Science, 21*(5), e12656.

Walker, A., & Leakey, R. (1993). *The Nariokotome Homo erectus skeleton.* Netherlands: Springer.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*(5765), 1301–1303.

Weiss, P. H., Ubben, S. D., Kaesberg, S., Kalbe, E., Kessler, J., Liebig, T., & Fink, G. R. (2016). Where language meets meaningful action: A combined behavior and lesion analysis of aphasia and apraxia. *Brain Structure & Function, 221*(1), 563–576.

Ye, H. (2019). The educational value of the body: A phenomenological perspective. *Educational Research, 10*, 41–51.

Zhang, H. (2019). Cognitive characteristics of communicative gestures. *Advances in Psychological Science, 26*(5), 796–809.

Zambrana, I. M., Ystrom, E., Schjølberg, S., & Pons, F. (2013). Action imitation at 1½ years is better than pointing gesture in predicting late development of language production at 3 years of age. *Child Development, 84*(2), 560–573.

Note: Figure translations are in progress. See original paper for figures.