2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

Simulating the Coevolution of Language and Long-Term Memory

Lan Shuai, Zhen Wang, and Tao Gong

Abstract—**Memory is fundamental to social activities such as language communications, yet it remains unclear how memory capacity and language use influence each other during language evolution, especially the early stage of language origin. Here, we proposed an evolutionary framework to address this issue. It assumed a genetic transmission of memory capacity and integrated natural and cultural selections that respectively affected the choices of parents for reproducing offspring and teaching these offspring. Simulation results obtained under this framework and relevant statistical analyses collectively traced a coevolution of language and capacity of individual long-term memory for storing acquired linguistic knowledge during the origin of a communal language in a multi-individual population. In line with the coevolutionary theory of language and related cognitive competences, this simulation study demonstrated that culturally-constituted aspects (communicative success) could drive the natural selection of predisposed cognitive features (long-term memory capacity), thus showing that language resulted from biological evolution, individual learning, and socio-cultural transmissions.**

I. Introduction

MEMORY is defined in general as the cognitive processes whereby information is encoded, stored, retrieved, or integrated. It is essential for humans and non-human animals to sense, perceive, or change the world, via individual or social activities toward the environment or among each other. In psychology, the most influential model of memory classifies memory into short-term (STM) and long-term memory (LTM) [1]. Environmental or exchanged information first enters STM for a temporary storage, which is an antechamber to more durable LTM. STM also serves as working memory (WM) for activities such as comprehension, learning, or reasoning [2]. The dominant model of WM [3] further classifies WM into: (a) central executive, to supervise other components and relevant cognitive processes; (b) phonological loop, to store recently-encountered sound stimuli; (c) visuospatial sketchpad, to hold visual and/or spatial information; and (d) episodic buffer [4], to integrate cross-domain information to form visual, spatial, and/or verbal units. LTM incorporates declarative (explicit) memory and implicit memory, the former of which comprises semantic memory recording factual information or general knowledge and episodic memory storing personal experiences [5]. Knowledge in declarative memory is often extracted from information in STM or WM [6].

These memory components are especially useful for social activities such as language. For example, the phonological loop and visuospatial sketchpad help store meaning-utterance mappings exchanged in recent communications, which serve as individual "previous experience" for acquiring lexical and grammatical knowledge of exchanged language [7][8]. In addition, empirical evidence reveals that although the STM of humans and other animals shows similar, limited capacities [9][10], the LTM of humans is superior, especially in the domain of language use. For example, other animals appear to have lower degrees (in terms of duration, number or type of stored items) of "episodic-like" memory mainly for temporarily storing information about food and its location, whereas the episodic and semantic memories of humans allow storing and constructing concrete and abstract knowledge of a variety of items, autobiographical events, as well as their correlations [11]. Furthermore, language, as the primary means of expressing semantic or conceptual knowledge, must have correlations with the memory system that stores such knowledge and participates in acquisition and processing of such knowledge. Empirical studies have gathered much evidence of memory constraints on language performance (e.g., [12][13]), yet theories of language evolution have not paid sufficient attention to the evolutionary relation between language and memory, many of which simply treat a sufficient memory capacity as a prerequisite for language evolution. Exploring evolutionary questions such as how the capacities of memory components relevant for language use are formed, or how memory capacity and language use affect each other during the evolution of language could shed important light on our understanding of the relation between language and general cognition and the process of language evolution.

We propose an evolutionary scenario of the development of human memory capacity for language use, which focuses in particular on the LTM capacity for linguistic knowledge. This scenario assumes that early hominins could temporarily store, using their STM, meaning-utterance mappings exchanged in recent communications among each other. To interpret more linguistic utterances, early hominids started to extract recurrent patterns in those temporarily-stored instances as linguistic knowledge, and allocate LTM units to store this acquired knowledge. Here, different LTM capacities could allow recording different amount of linguistic knowledge, thus causing distinct degrees of mutual understanding among each other. In this way, the LTM capacity became correlated with linguistic understanding. Following the socio-biological explanation [14] and the

Lan Shuai is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: susan,shuai@gmail.com).
Zhen Wang is with the School of Innovation Experiment, Dalian University of Technology, Dalian, Liaoning Prov. 116024 China (e-mail: wangz@dlut.edu.cn).
Tao Gong is with the Department of Linguistics, University of Hong Kong, Hong Kong (corresponding author, phone: 852-39178603; fax: 852-25467477; e-mail: gtojty@gmail.com).
This work was supported in part by the Seed Fund for Basic Research of the University of Hong Kong.

brain-language coevolution theory [15–18], communicative success could reciprocally affect the LTM capacities among language users, thus triggering a *coevolution* (a reciprocal influence between natural species or system components [19]) of language and LTM capacity.

Noting that available empirical studies cannot offer explicit answers to evolutionary questions about memory capacity and its role in language evolution, we adopt evolutionary computation techniques to evaluate this scenario. First, we adopt a multi-agent, language origin model [20][21] as the language model in our study. It simulates a simultaneous acquisition of lexical items and simple word orders out of a holistic protolanguage in a multi-individual population. There are two reasons to select this model. First, it implements the STM storing exchanged linguistic instances and the LTM recording linguistic knowledge acquired from these instances, thus enabling us to explore the coevolution of language and LTM (as well as other relevant competences). Second, it simulates instance-based learning mechanisms for acquiring both lexical and syntactic knowledge. Compared with the other studies of memory based on lexical evolution models (e.g., [22][23]), these learning mechanisms and evolved artificial languages are more comparable to the linguistic abilities of language learning children and the preliminary languages of early hominids. Second, we propose an evolutionary framework, which involves both natural and cultural selections that take effect during generation replacement and language learning. Under this framework, we conduct simulations in different conditions to systematically analyze the evolutions of language and LTM capacity. The simulation results and relevant statistical analyses collectively illustrate the coevolution of language and LTM and vividly show the roles of these selections in the coevolution process.

In the following sections, we first describe the memory system, evolutionary framework, and simulation setup, then report and analyze the simulation results, and finally discuss the coevolution of language and memory capacity.

II. Memory System, Evolutionary Framework, and Simulation Setup

A. Memory System

The major components of the language model are briefly described in the Appendix. Here, we concentrate on the memory system in this model. In this two-component system, the STM stores meaning-utterance mappings obtained in previous communications where this agent was the listener; and the LTM records linguistic knowledge (lexical and syntactic rules) extracted from the instances in the STM. These rules are used by the agent in future communications.

Both the STM and the LTM have fixed capacities, and the contents in them are updated along with communications. When the STM is full, newly-obtained mappings will replace old ones stored; when the LTM is full, newly-acquired rules will replace those having lower strengths (rule strength indicates the probability of successfully applying this rule in communications). Rules having been unsuccessfully used for many times or having zero or negative strengths due to gradual forgetting are also discarded from the LTM. For the sake of simplicity, we focus on the LTM capacity for storing lexical rules, and fix the LTM capacities for storing other types of linguistic knowledge. Lexical rules are the most fundamental linguistic knowledge in this model, on which syntactic rules are built (see Appendix).

B. Evolutionary Framework

The evolutionary framework, inspired from [18], involved *genetic transmission* (transmitting the LTM capacity for storing lexical rules from adults to offspring) during reproduction and *cultural transmission* (intra-generational communications where adults talk to each other, and inter-generational communications where adults talk to offspring). In each round of generation replacement, first, intra-generational communications take place. After that, half of the adults are chosen as parents, each producing two (in order to maintain the population size) offspring (new agents) who have no linguistic knowledge but copy the LTM capacities of their parents with occasional mutation. Then, inter-generational transmissions start. Later on, offspring replace their parents and the next generation begins. Such a punctuated setting is to explicitly trace the evolution across generations. In real cases, cultural and genetic transmissions could be intertwined.

An agent's *communicative success* (*CS*) reflects its fitness in the population. *CS* is measured as the mean percentage of integrated meanings an agent can accurately understand (using linguistic knowledge only) when others talk to him/her. In this framework, both natural and cultural selections take effect based on *CS*: natural selection selects adults who can better understand others (having higher *CS*) as parents to produce offspring; cultural selection selects adults having higher *CS* as teachers to talk to offspring. The mean *CS* of all agents is *understanding rate* (*UR*). A high *UR* indicates that the language used by agents can accurately exchange many meanings. We only address one type of cultural selection relevant for *CS* here, the roles of cultural selection in language evolution also manifest in other aspects (e.g., [24–28]).

C. Simulation Setup

We set up five sets of simulations. The *NoChange* set allows no cultural and natural selections or mutation on the LTM capacity; parents and teachers are randomly chosen and offspring directly copy their parents' LTM capacities without adjustment. The other four sets (*NoNat_NoCul*, without natural and cultural selections; *Nat_NoCul*, with natural selection but without cultural selection; *NoNat_Cul*, without natural selection but with cultural selection and *Nat_Cul*, with both natural and cultural selections) form a 2×2 design, where natural and cultural selections are two factors, each having two levels (in effect or not). When natural selection is in effect (*Nat_Cul* and *Nat_NoCul*), adults with higher *CS* have higher chances to be parents; otherwise, parents are chosen randomly. When cultural selection is in effect (*Nat_Cul* and *NoNat_Cul*), adults with higher *CS* have higher chances to be teachers; otherwise, teachers are chosen randomly. When offspring copy their parents' LTM capacities, mutation (increase or decrease the capacity with a fixed amount) may occur.

Table 1 summarizes the parameter setting in this framework (see Appendix for other parameters controlling the learning mechanisms and communications). In the simulations of this paper, there are 64 integrated meanings for agents to exchange, each having equal chances to be produced in communications. Agents in the first generation can only express eight integrated meanings using eight rules stored in their LTMs. This resembles a limited signaling system of early hominins (in fact, simulations starting from no linguistic knowledge reported similar results). The eight meanings contain all semantic constituents in all 64 integrated meanings. The simulation results are less dependent on the sizes of the semantic space and the population; if these values increase, similar results can be obtained given more rounds of cultural transmission per generation. The number of generations is set to 2000, sufficient to observe the possible coevolution and evaluate the roles of the selections in this process. In each set, we conduct 50 simulations. In each simulation, *UR* of the communal language and the LTM capacities of agents are measured at 201 sampling points evenly distributed in 2000 generations. We fix the capacities of STM and LTM for storing other types of linguistic rules to neutralize their effects on language evolution. As for the LTM capacity for lexical rules, we first analyze the effect of it on *UR* based on the *NoChange* set of simulations, and then, set the initial LTM capacity in the other sets to trace the coevolution of language and LTM capacity.

TABLE I
PARAMETER SETTING FOR THE EVOLUTIONARY FRAMEWORK.

Parameters	Values
Population size	10
Mutation rate	0.05
Amount of adjustment on LTM capacity	1
Number of intra-generational transmissions per generation	100
Number of inter-generational transmissions per generation	200
STM size	40
LTM for categories and syntactic rules	40

III. SIMULATION RESULTS

A. The NoChange Set

Simulations in this set show a correlation between the LTM capacity and *UR* of the emergent language (see Fig. 1): if the LTM capacity is below 30, *UR* remains low; once it surpasses 30, *UR* starts to increase along with the increase in it; and when it is sufficiently large, *UR* remains high. These results reveal a threshold LTM capacity (around 30), only beyond which can a communal language with high *UR* emerge.

Noting this, we set the initial LTM capacities in the other four sets of simulations to two values, one around the threshold (30) and the other sufficiently larger (60), to see how natural and cultural selections affect the evolutions of language and LTM capacity in these conditions. In those simulations, the initial LTM capacities in the first generation are randomly chosen from Gaussian distributions, whose standard deviations are fixed at 5 and means at 30 (or 60). This manipulation preserves the general characteristic of the whole population and also introduces a certain degree of individual difference, which paves the way for evolution.

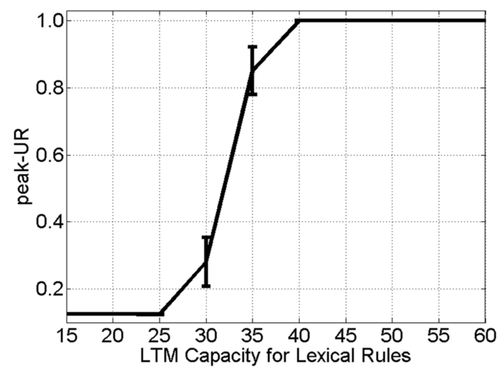


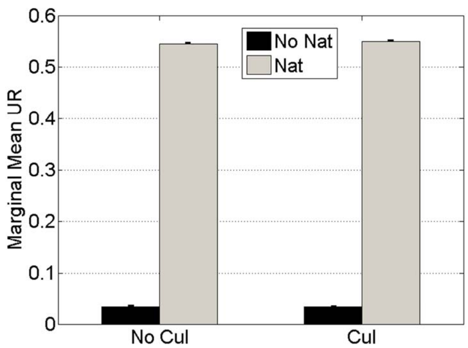
Fig. 1. Average peak-*UR* (highest *UR* throughout 2000 generations) in the *NoChange* set of simulations. Error bars denote standard errors.

B. The Other Four Sets

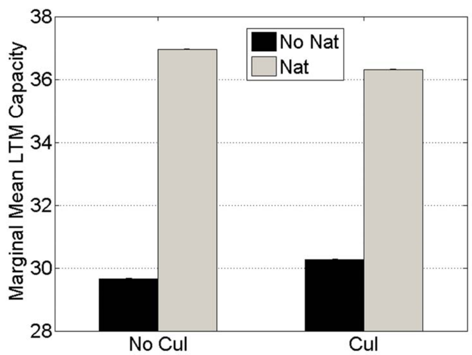
Simulations in these sets reveal a coevolution of language (indicated by *UR*) and LTM capacity.

In the condition with the initial LTM capacity being 30, for *UR*, a two-way analysis of covariance (ANCOVA [29]) (dependent variable: mean *UR* over 50 simulations; fixed factors: natural and cultural selections; covariate: sampling points throughout generations) shows that: natural selection has a significant main effect on *UR* ($F(1, 40195) = 47979.969$, $p < .001$, $\eta_p^2 = .544$), but cultural selection does not ($F(1, 40195) = 96.206$, $p = .422$, $\eta_p^2 = .000$); and there is no significant interaction between the two selections ($F(1, 40195) = 1.297$, $p = .532$, $\eta_p^2 = .051$). The covariate, generation (sampling points), also interacts significantly with *UR* ($F(1, 40195) = 8951.653$, $p < .001$, $\eta_p^2 = .182$). Using ANCOVA, not ANOVA, is to partial out the influence of the covariant.

Fig. 2(a) shows that the marginal mean *UR* (average *UR* at all sampling points of all simulations from the corresponding set; note that the two types of selection take effect throughout 2000 generations, comparing absolute values at each sampling point is inappropriate to clarify their general effects, therefore, we first compare such average values throughout simulations, the absolute values at different sampling points can be seen in Fig. 3 and Fig. 5 that trace the mean *UR* and LTM capacity throughout 2000 generations in different sets) in the sets with natural selection (*Nat_NoCul* and *Nat_Cul*) is significantly higher than that in the sets without (*NoNat_NoCul* and *NoNat_Cul*), but the marginal mean *UR* in the sets with cultural selection (*NoNat_Cul* and *Nat_Cul*) is not much different from that in the sets without (*NoNat_NoCul* or *Nat_NoCul*). These results reveal that it is natural selection, rather than cultural selection, that drives the origin of a communal language with high *UR*.



(a)

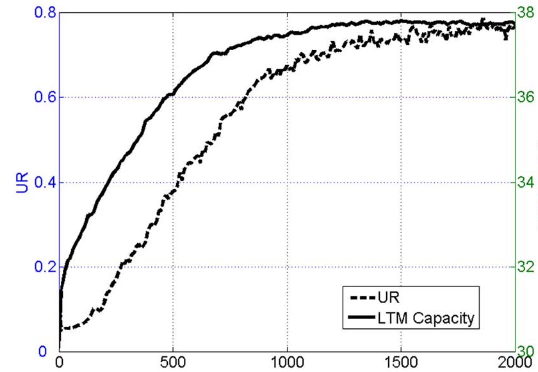


(b)

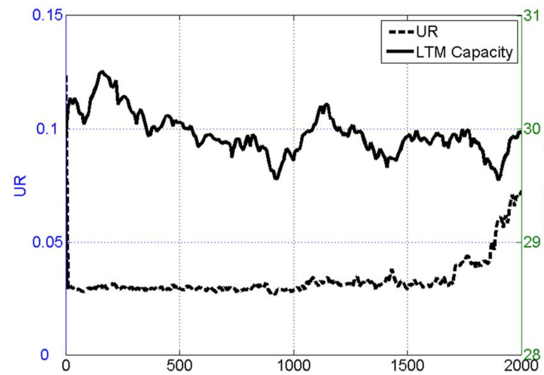
Fig. 2. (a) Marginal mean *UR* as a function of natural (*Nat*) and cultural (*Cul*) selections; (b) Marginal mean LTM capacity as a function of natural and cultural selections. Error bars denote standard errors. The initial LTM capacity is 30.

As for the LTM capacity, a similar ANCOVA (dependent variable: mean LTM capacity over 50 simulations) shows that: natural selection has a significant main effect on LTM capacity ($F(1, 40195) = 29360.665$, $p < .001$, $\eta_p^2 = .422$), but cultural selection does not ($F(1, 40195) = .128$, $p = .721$, $\eta_p^2 = .000$); and there is a significant interaction between the two selections ($F(1, 40195) = 257.788$, $p < .001$, $\eta_p^2 = .006$), but this effect is very small (due to its small η_p^2). Fig. 2(b) shows these results. Similarly, Fig. 2(b) also shows that the evolution of the LTM capacity is mainly achieved by natural selection, rather than cultural selection. In addition, the covariate also interacts significantly with LTM capacity, but this effect is small ($F(1, 40195) = 1045.447$, $p < .001$, $\eta_p^2 = .025$).

Fig. 2 and the ANCOVA analyses show that natural selection can gradually enhance an initially-low LTM capacity, and lead to the origin of a communal language with high *UR*. Such coevolution can also be observed by tracing the mean *UR* and LTM capacity throughout 2000 generations in the sets with natural selection (see Fig. 3(a)); due to natural selection, both the mean *UR* and LRM capacity increase synchronically from a low value to a high value (as for *UR*, it increases from 0.125 (due to eight initially-shared rules) to around 0.8; as for the LTM capacity, it increases from 30 to about 38). By contrast, in the sets without natural selection, such coevolution disappears; the mean *UR* and LTM capacity only fluctuate around their initial values (Fig. 3(b)).



(a)



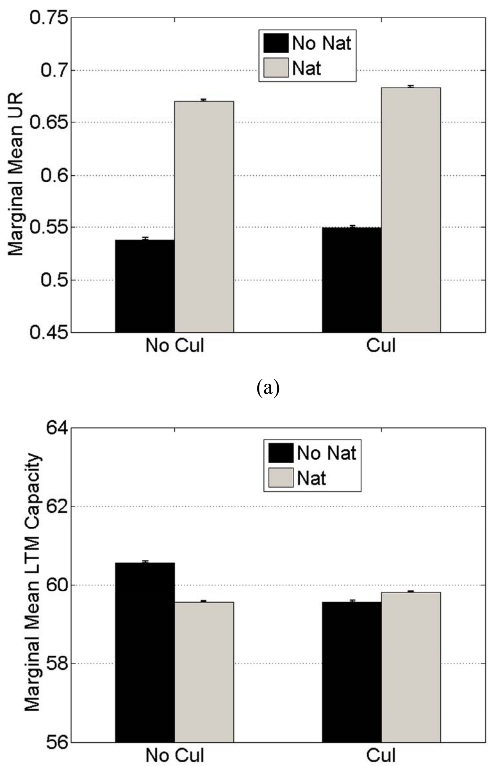
(b)

Fig. 3. (a) Mean *UR* and LTM capacity throughout 2000 generations in conditions where natural selection is in effect; (b) Mean *UR* and LTM capacity throughout 2000 generations in the conditions where natural selection is not in effect. The initial LTM capacity is 30.

In the condition with the initial LTM capacity being 60, as for *UR*, the ANCOVA analysis shows that: natural selection has a significant effect on *UR* ($F(1, 40195) = 1915.191$, $p < .001$, $\eta_p^2 = .045$), but this effect is small; cultural selection has a significant effect, but this effect is very small ($F(1, 40195) = 16.443$, $p < .001$, $\eta_p^2 = .000$); and there is no significant interaction between the two selections ($F(1, 40195) = .036$, $p = .850$, $\eta_p^2 = .000$). The covariate also interacts significantly with *UR*, but this effect is very small ($F(1, 40195) = 17.230$, $p < .001$, $\eta_p^2 = .000$).

As for the LTM capacity, the ANCOVA analysis shows that: natural selection has a significant effect, but this effect is very small ($F(1, 40195) = 95.775$, $p < .001$, $\eta_p^2 = .002$); cultural selection has a significant effect, but this effect is very small ($F(1, 40195) = 98.046$, $p < .001$, $\eta_p^2 = .002$); and there is a significant interaction between the two selections, but this effect is also very small ($F(1, 40195) = 267.043$, $p < .001$, $\eta_p^2 = .007$). The covariate has no significant effect ($F(1, 40195) = 3.482$, $p = .062$, $\eta_p^2 = .000$).

The ANCOVA analyses report several effects of natural and cultural selections, but most of them are small (due to their small η_p^2). See Fig. 4, *UR* and the LTM capacity do not change much throughout 2000 generations: the change in *UR* is within 0.15 and that in the LTM capacity is within 1, smaller than the adjusting amount of LTM capacity. These results show that when the initial LTM capacity is sufficiently large, both natural and cultural selections cannot greatly influence the communal language or further adjust the LTM capacity; in other words, the coevolution of language and LTM capacity as shown in the first condition disappears.



(a)

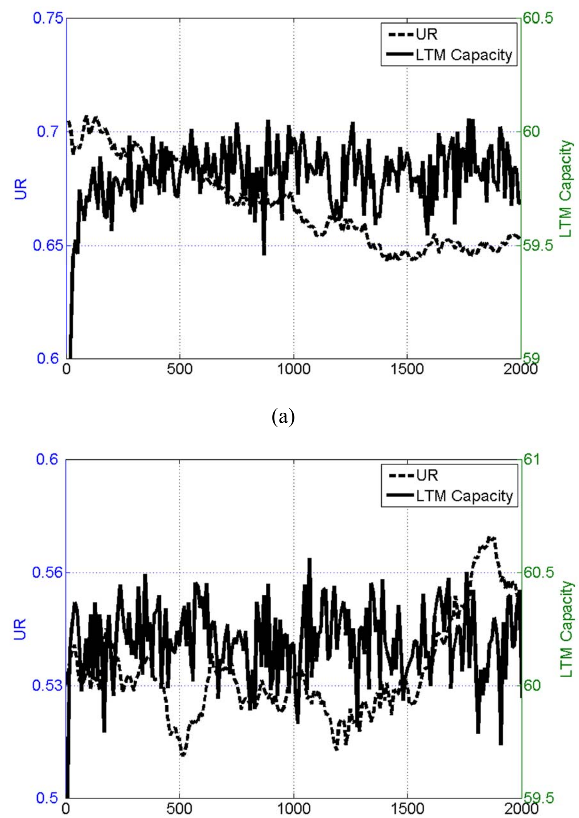
(b)

Fig. 4. (a) Marginal mean *UR* as a function of natural (*Nat*) and cultural (*Cul*) selections; (b) Marginal mean LTM capacity as a function of natural and cultural selections. Error bars denote standard errors. The initial LTM capacity is 60.

The disappearance of the coevolution can also be observed in Fig. 5 that traces the mean *UR* and LTM capacity in the sets with and without natural selection. The mean *UR* and LTM capacity in all sets fluctuate around certain values and the small differences between the mean *UR* and LTM capacities in the sets with and without natural selection reflect the small effects of natural or cultural selection on these indices. Note that the mean *UR* in Fig. 5 is lower than 1.0 as in Fig. 1. This is because that the simulations in the *NoChange* set proceed without mutations, which could accumulate the random walk effect on the LTM capacities and mean *UR* to a certain extent. Also, there are no individual differences in the initial LTM capacity in the *NoChange* set, whereas in the other sets some individuals may have quite distinct LTM capacities, thus affecting the mean *UR* and LTM capacities in the population.

IV. Discussions and Conclusions

There is no doubt that: human language is much more abundant, in quantity and type, than the artificial language simulated in our model; our language model does not take into account many aspects that are also relevant for memory (e.g., forming, storing, or retrieving semantics and utterances in different memory components) and language communications (e.g., communications using compressed signals or combing information from other domains); and the observed threshold of the LTM capacity also depends on particular simulation settings (the threshold capacity has to be bigger than the number of all semantic constituent; otherwise, agents cannot store sufficient lexical rules to express all meanings). Nonetheless, our simulations reveal that during language origin, an initially-low LTM capacity can be enhanced to serve communicative purpose better. In addition, when the evolved capacity is enough for language communications, the effect of such coevolution becomes less explicit. These results indicate an internal correlation between language and memory, thus ensuring that if linguistic complexity increases (e.g., more expressions are available, requiring storing and applying more and additional type of linguistic knowledge), capacities of related memory components will also adapt accordingly.



(a)

(b)

Fig. 5. (a) Mean *UR* and LTM capacity throughout 2000 generations in conditions where natural selection is in effect; (b) Mean *UR* and LTM capacity throughout 2000 generations in the conditions where natural selection is not in effect. The initial LTM capacity is 60.

The statistical analyses of the simulation results show that the coevolution of language and LTM capacity is achieved mainly by natural selection, rather than cultural selection or both. In our simulations, cultural selection chooses adults with higher *CS* to be teachers talking to offspring. Even if these teachers have sufficient LTM capacities for storing linguistic knowledge, without natural selection, these agents would not necessarily have higher chances to reproduce offspring and transfer their capacities to their offspring. Meanwhile, given both natural and cultural selections, adults having higher *CS* also have higher chances to reproduce offspring, but the offspring would learn only from these adults. As shown in [27], such a biased sampling affects mutual understandability within and across multi-individual populations or generations.

Although the coevolution is achieved mainly via natural selection, cultural transmissions are indispensable, which provide opportunities for agents to develop their linguistic knowledge and form their *CS* for natural and/or cultural selection to take effect. In line with previous research on the roles of cultural selection in shaping simple lexicon [23], fundamental language structures [28][30], color terms [24][26], and other aspects, our study also reflects the role of culture in human cognition in general [31].

This coevolution is also consistent with our early study [18] tracing a possible coevolution of language and a language-related learning mechanism (e.g., joint attention). Both studies show that in the context of language evolution, genetic assimilation is able to help retain and expand communicatively-effective features [32]. Similar to joint attention, LTM is not language-specific, exists before language origin, and also participates in general interactive activities. Once its capacity becomes relevant for linguistic comprehension and communicative success can offer capable individuals some functional advantage (e.g., reproduction opportunity), under the drive of communicative success, memory could piggyback on language, having its capacity adjusted along with language origin. One difference between the two studies lies in that for joint attention, the coevolution also makes the level of joint attention ratcheted at a suitable high level, whereas for memory, if the initial capacity is sufficient for language use, without natural or cultural selection, this capacity will not greatly change.

Both studies demonstrate that language evolution resulted from biological evolution (e.g., genetic assimilation of some language-related competences such as joint attention or memory), individual learning (e.g., learning mechanisms that require those competences), and socio-cultural transmission [33][34]. In addition, they also revise previous views on language evolution, such as the biological evolution leading to language readiness and the cultural evolution of modern languages take place at distinct stages of language evolution [35] and high degrees of certain cognitive characteristics must be innate to humans and serve as prerequisites for language and human communication system [36]. We suggest that there could be no clear-cut distinction between the roles of biological and cultural evolutions in language origin and distinct levels of certain language-related competences in humans could be possibly due to a coevolution with language.

Appendix: The Language Model

A. Language, Individuals and Linguistic Knowledge

This model encodes language as *meaning-utterance mappings* (*M-U mappings*). Individuals share a semantic space containing a fixed number of *integrated meanings* having a "*predicate<agent>*" or "*predicate<agent, patient>*" structure. Here, *predicate*, *agent*, and *patient* are thematic notations. *Predicates* refer to actions that individuals can conceptualize (e.g., "run" or "chase"), and *arguments* entities on or by which actions are performed (e.g., "fox" or "tiger"). Some predicates can take one argument, e.g., "run<tiger>" meaning "a tiger is running"; others can take two, e.g., "chase<tiger, fox>" meaning "a tiger is chasing a fox", where the first constituent in < >, "tiger", denotes the *agent* (action instigator) of the predicate "chase", and the second, "fox", the *patient* (entity undergoing the action) of the predicate. For the sake of simplicity, integrated meanings having identical agent and patient constituents (e.g., "fight<fox, fox>") are excluded.

Integrated meanings are encoded by *utterances*, each comprising a string of syllables from a signaling space. An utterance encoding an integrated meaning can be segmented into subparts, each mapping one or two constituents; and subparts can combine to form an integrated meaning.

Individuals are simulated as *artificial agents*, who, based on general learning mechanisms, can acquire linguistic knowledge from the M-U mappings obtained in recent communications, produce utterances to encode integrated meanings and comprehend utterances during communications with others. During reproduction, some agents are chosen as parents to produce offspring (new agents who have no linguistic knowledge but copying the LTM capacities of their parents). After learning from teachers, these offspring replace their parents.

Linguistic knowledge is encoded by *lexicon*, *syntax*, and *categories*. An individual's lexicon comprises *lexical rules* (see Fig. A1). Some are *holistic*, each mapping an integrated meaning onto an utterance (sentence), e.g., "run<tiger>" ↔ /abcd/ indicates that the meaning "run<tiger>" can be encoded by the utterance /abcd/, and also that /abcd/ can be decoded as "run<tiger>"; others are *compositional*, each mapping one or two constituent(s) onto a subpart of an utterance, e.g., "fox" ↔ /ef/ or "chase<wolf, #>" ↔ /gh*i/ ("#" denotes an unspecified constituent, and "*" unspecified syllable(s)).

A *syntactic rule* records a local order (before or after) between two lexical rules.

Categories are formed for syntactic rules acquired from some lexical items to be applicable to other lexical items having the same thematic notation (*agent*, *patient*, or *predicate*). A *category* (see Fig. A1) has a list of lexical rules and a list of syntactic rules that specify the orders in utterance between these lexical rules and those from other categories. For simplicity, we simulate a nominative-accusative language and exclude passive voice. Then, a category associating lexical rules encoding the thematic notation of *agent* is also denoted as a subject (S) category, since *agent* corresponds to the syntactic role of S. Similarly, *patient* corresponds to object (O), and *predicate* to verb (V). A local order between categories can be denoted by the syntactic roles (e.g., an order *before* between an S and a V category can be denoted by SV.

Each lexical or order rule has a *strength* (within [0.0 1.0]), denoting how often this rule is applied successfully in communications. A compositional rule also has *association weights* to categories containing this rule, indicating how often the syntactic rules in those categories can be successfully applied on this rule in communications. Rule strength and association weight enable a strength-based competition among rules in production and perception during communications and a forgetting of linguistic knowledge.

Lexical rules

Holistic rules:

(a) “chase<wolf, bear>”←→/a/ (0.5)

(b) “hop<deer>”←→/c/ (0.4)

(c) “hop<deer>”←→/d e/ (0.6)

Compositional rules:

(d) “wolf”←→/f/ (0.6)

(e) “run<#>”←→/c/ (0.7)

(f) “chase<#, bear>”←→/e f * g/ (0.7)

Syntactic rules

(1) Category 1 (S) << Category 2 (V) (0.8)

(2) Category 3 (O) >> Category 2 (V) (0.4)

Categories **Category 1 (S):**

List of lexical rules: {“fox”←→/a/ (0.5)} [0.5]
{“wolf”←→/b c/ (0.7)} [0.5]

List of syntactic rules: Category 1 (S) << Category 2 (V) (0.8)
Category 3 (O) >> Category 1 (S) (0.4)

Category 2 (V):

List of lexical rules: {“run<#>”←→/d/ (0.4)} [0.5]
{“fight<#,#>”←→/e/ (0.7)} [0.5]

List of syntactic rules: Category 1 (S) << Category 2 (V) (0.8)

Category 3 (O):

List of lexical rules: {“fox”←→/a/ (0.5)} [0.7]
{“sheep”←→/h/ (0.6)} [0.6]

List of syntactic rules: Category 3 (O) >> Category 1 (S) (0.4)

Fig. A1. Examples of lexical rules, syntactic rules, and categories. “#” denotes unspecified semantic item, and “*” unspecified syllable(s). S, V, and O are syntactic roles of categories. Numbers enclosed by ( ) denote rule strengths, and those by [ ] association weights. “<<” denotes the local order *before*, and “>>” *after*. Compositional rules can combine, if specifying each constituent in an integrated meaning exactly once, e.g., rules (c) and (d) can combine to form “chase<wolf, bear>”, and the corresponding utterance is /ehfg/. Lexical and syntactic knowledge collectively encode integrated meanings, e.g., to express “fight<wolf, fox>” using the lexical rules from the S, V, and O categories and the syntactic rules SV and SO, the resulting sentence is /bcea/ or /bcae/, following SVO or SOV.

B. General Learning Mechanisms

Agents are equipped with general learning mechanisms to acquire linguistic rules (see [20][21] for details). Lexical rules are acquired from constituent(s) and syllable(s) repetitively appearing in two or more meaning-utterance mapping in the STM. New mappings, before being stored to the STM, are compared with those already existent. If the STM is full, the newly-added mapping replaces the oldest one. For example, an agent can detect the recurrent pattern “fox” and /a/ by comparing “hop<fox>”↔/ab/ and “run<fox>”↔/acd/. If the agent has no rule recording this pattern, it will create a lexical rule “fox” ↔ /a/ and put it in the LTM for future use.

Categories and order rules are acquired based on thematic roles of lexical rules and order relations of their utterances in meaning-utterance mappings stored in the STM. If an agent notices that in some previous experiences the utterances of two or more lexical rules having the same thematic role are consistently before (or after) the utterance of another lexical rule (or the utterances of another set (category) of lexical rules all having identical thematic roles), the agent can associate these lexical rules into a category having the corresponding syntactic role, create a syntactic rule to record the local order with respect to the other lexical rule(s), and put this syntactic rule to the same category. The categorical and syntactic knowledge is also stored in the LTM. In this way, the agent can form categories linking lexical rules with local orders.

C. Communication Scenario

A linguistic communication involves two agents (a speaker and a listener), who perform a number of *utterance exchange*.

In production, the speaker (hereafter as “she”) first selects randomly an integrated meaning from the semantic space. She then activates her lexical, syntactic, and category rules to form candidate sets, each offering a sentence to encode the selected meaning. For each set, she calculates the combined strength (see [20][21] for details). After calculation, she chooses the set having the highest combined strength, builds the utterance accordingly, and transmits the utterance to the listener. If lacking enough rules to encode the meaning, she occasionally (the creation rate 0.5) creates a holistic rule to encode the meaning, and sends the utterance of this rule to the listener.

In comprehension, the listener (hereafter as “he”) receives the utterance from the speaker and an *environmental cue*. The cue, as non-linguistic information, contains an integrated meaning plus a *cue strength*. Cues are unreliable (not always containing the speaker’s intended meaning). This is to avoid explicitly transferring meanings in exchanged utterances via non-linguistic information, which would make linguistic communication unnecessary. We define *reliability of cue* (*RC*) to denote how often the listener obtains a correct cue (containing the speaker’s intended meaning) in an utterance exchange; otherwise, he receives a wrong cue (containing an integrated meaning randomly chosen from the semantic space and distinct from the speaker’s intended meaning). In the simulations of this paper, *RC* is set to 0.6. The effect of *RC* on language evolution is discussed in [18].

The listener activates his lexical, syntactic and category rules stored in his STM that can interpret the heard sentence as integrated meaning(s). He then compares the cue’s meaning with the one(s) comprehended by linguistic rules, and sets up candidate sets for comprehension. If the cue’s meaning completely or particularly matches the one interpreted by some linguistic rules, the cue and those linguistic rules form a candidate set. Otherwise, the cue itself forms a candidate set. If some linguistic rules can offer a complete interpretation, they form another set as well.

The listener calculates the combined strength of each set. For a set without a cue, its combined strength is calculated exactly the same as that in production. For a set with a cue, the cue strength is added to the combined strength. After calculation, he chooses the set having the highest combined strength for comprehension. If this combined strength exceeds a confidence threshold (0.75), the utterance exchange is deemed successful. Then, the listener stores the perceived meaning-utterance mapping to her STM, and both the speaker and the listener reward the rules in their chosen sets by adding a fixed amount (0.1) to their strengths and association weights, and penalize competing ones in the other sets by deducting the same amount from their strengths and association weights. Otherwise, the utterance exchange is deemed failed. Then, the listener discards the perceived mapping, and both the speaker and the listener penalize their rules in their chosen sets.

The cue strength equals to the confidence threshold, so that the linguistic information and non-linguistic information are treated equally. Throughout the utterance exchange, there is no direct check whether the speaker's encoded meaning matches the listener's decoded one. The adjustment on rule strength leads to conventionalization of linguistic knowledge.

For the linguistic rules stored in the LTM, agents frequently deduct a fixed amount (0.01) from their strengths and association weights. Then, lexical or syntactic rules having zero or negative strengths are discarded, lexical rules having zero or negative association weights to some categories are removed from those categories, and categories having no lexical members are discarded as well. This forgetting operation simulates the update of the LTM content.

REFERENCES

[1] P. W. B. Atkins and R. M. Shiffrin, “Human memory: A proposed system and its control processes,” in *The Psychology of Learning and Motivation: Advances in Research and Theory*, K. W. Spence, Ed. New York: Academic Press, 1968, pp. 89–195.

[2] A. D. Baddeley, “Working memory and language: An overview,” *Journal of Communication Disorders*, vol. 36, pp. 189–208, 2003.

[3] A. D. Baddeley and G. J. Hitch, “Working memory,” in *Recent Advances in Learning Motivation*, vol. 8, G. A. Bower, Ed. New York: Academic Press, 1974, pp. 47–90.

[4] A. D. Baddeley, “The episodic buffer: A new component of working memory?” *Trends in Cognitive Sciences*, vol. 4, pp. 417–423, 2000.

[5] E. Tulving, “Episodic and semantic memory,” in *Organization of Memory*, E. Tulving and W. Donaldson, Eds. New York: Academic Press, 1972, pp. 381–403.

[6] A. D. Baddeley, C. Papagno, and G. Vallar, “When long-term learning depends on short-term storage,” *Journal of Memory and Language*, vol. 27, pp. 586–595, 1988.

[7] P. W. B. Atkins and A. D. Baddeley, “Working memory and distributed vocabulary learning,” *Applied Psycholinguistics*, vol. 19, pp. 537–552, 1998.

[8] A. D. Baddeley, S. E. Gathercole, and C. Papagno, “The phonological loop as a language learning device,” Psychological Review, vol. 105, pp. 158–173, 1998.

[9] G. A. Miller, “The magic number seven, plus or minus two: Some limits on our capacity for processing information,” *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[10] C. Nelson, “The magical number 4 in short-term memory: A reconsideration of mental storage capacity,” *Behavioral and Brain Sciences*, vol. 24, pp. 87–173, 2001.

[11] J. R. Hurford, *The Origins of Meaning*. Oxford: Oxford University Press, 2007.

[12] R. Wood, P. Baxer, and T. Belpaeme, “A review of long term memory in natural and synthetic systems,” *Adaptive Behavior*, vol. 20, no. 2, pp. 81–103, 2011.

[13] C. Hulme, S. Maughan, and G. D. A. Brown, “Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span,” *Journal of Memory and Language*, vol. 30, pp. 685–701, 1991.

[14] J. R. Hurford, “Biological evolution of the Saussurean sign as a component of the Language Acquisition Device,” *Lingua*, vol. 77, pp. 187–222, 1989.

[15] W. T. Deacon, *The Symbolic Species: The Coevolution of Language and the Brain*. New York, NY: W. W. Norton, 1997.

[16] K. N. Laland, J. Odlin-Smee, and S. Myles, “How culture shaped the human genome: bring genetics and the human sciences together,” *Nature Reviews Genetics*, vol. 11, pp. 793–808, 2010.

[17] S. Számadó and E. Szathmáry, “Evolutionary biological foundations of the origin of language: The coevolution of language and brain,” in *The Oxford Handbook of Language Evolution*, M. Tallerman and K. R. Gibson, Eds. Oxford: Oxford University Press, 2013, pp. 157–167.

[18] T. Gong and L. Shuai, “Modelling the coevolution of joint attention and language,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 279, pp. 4643–4651, 2013.

[19] J. N. Thompson, *The Coevolutionary Process*. Chicago, IL: University of Chicago Press, 1994.

[20] T. Gong, *Computational Simulation in Evolutionary Linguistics: A Study on Language Emergence*. Taipei: Academia Sinica, 2009.

[21] T. Gong, “Simulating the coevolution of compositionality and word order regularity,” *Interaction Studies*, vol. 12, no. 1, pp. 63–106, 2011.

[22] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels, “Sharp transition towards shared vocabularies in multi-agent systems,” *Journal of Statistical Mechanics: Theory and Experiment*, P06014, 2006.

[23] V. Loreto, A. Baronchelli, A. Mukherjee, A. Puglisi, and F. Tria, “Statistical physics of language dynamics,” *Journal of Statistical Mechanics: Theory and Experiment*, P04006, 2011.

[24] L. Steels and T. Belpaeme, “Coordinating perceptually grounded categories through language: A case study for color,” *Behavioral and Brain Sciences*, vol. 28, pp. 469–489, 2005.

[25] T. L. Griffiths and M. L. Kalish, “Language evolution by iterated learning with Bayesian agents,” *Cognitive Sciences*, vol. 31, pp. 441–480, 2007.

[26] A. Baronchelli, T. Gong, A. Puglisi, and V. Loreto, “Modeling the emergence of universality in color naming patterns,” *Proceedings of the National Academy of Sciences of the USA*, vol. 107, 2403–2407, 2010.

[27] T. Gong, “Exploring the roles of horizontal, vertical, and oblique transmissions in language evolution,” *Adaptive Behavior*, vol. 18, 356–376, 2010.

[28] T. Scott-Phillips and S. Kirby, “Language evolution in the laboratory,” *Trends in Cognitive Sciences*, vol. 14, pp. 411–417, 2010.

[29] A. Field, *Discovering Statistics Using SPSS*, 3rd Ed. London: SAGE. 2009.

[30] M. Dunn, S. J. Greenhill, S. C. Levinson, and R. D. Gray, “Evolved structure of language shows lineage-specific trends in word-order universals,” *Nature*, vol. 473, pp. 79–82, 2011.

[31] M. Tomasello, “Human culture in evolutionary perspective,” in *Advances in Culture and Psychology*, M. Gelfand, Ed. Oxford: Oxford University Press, 2011, pp. 5–52.

[32] M. H. Christiansen, F. Reali, and N. Chater, “Biological adaptations for functional features of language in the face of cultural evolution,” *Human Biology*, vol. 83, pp. 247–259, 2011.

[33] H. Brighton, S. Kirby, and K. Smith, “Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable,” in *Language origins: Perspectives on evolution*, M. Tallerman, Ed. Oxford: Oxford University Press, 2005, pp. 291–309.

[34] S. Kirby, “Language is an adaptive system: The role of cultural evolution in the origins of structure,” In *The Oxford handbook of language evolution*, M. Tallerman and K. R. Gibson, Eds. Oxford: Oxford University Press, 2012, pp. 589–604.

[35] M. Arbib, *How the Brain Got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press, 2012.

[36] M. Tomasello, *Origins of Human Communication*. Cambridge, MA: MIT Press, 2009.