Modelling the coevolution of joint attention and language

Tao Gong$^1$,* and Lan Shuai$^2$

$^1$Department of Linguistics, University of Hong Kong, Pokfulam Road, Hong Kong

$^2$Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MA, USA

Joint attention (JA) is important to many social, communicative activities, including language, and humans exhibit a considerably high level of JA compared with non-human primates. We propose a co-evolutionary hypothesis to explain this degree-difference in JA: once JA started to aid linguistic comprehension, along with language evolution, communicative success (CS) during cultural transmission could enhance the levels of JA among language users. We illustrate this hypothesis via a multi-agent computational model, where JA boils down to a genetically transmitted ability to obtain non-linguistic cues aiding comprehension. The simulation results and statistical analysis show that: (i) the level of JA is correlated with the understandability of the emergent language; and (ii) CS can boost an initially low level of JA and ‘ratchet’ it up to a stable high level. This coevolutionary perspective helps explain the degree-difference in many language-related competences between humans and non-human primates, and reflects the importance of biological evolution, individual learning and cultural transmission to language evolution.

Keywords: joint attention; language origin; coevolution; ratchet effect; computer simulation

1. INTRODUCTION

In psychology and cognitive sciences, joint attention (JA) is defined as the process of establishing common ground in general interactive activities, by means of socio-cognitive abilities [1]. JA usually involves two interacting human/non-human individuals and an object/event outside the two. It can be achieved by non-verbal (e.g. gaze following or pointing) and/or verbal activities (e.g. language).

JA is essential for many social activities, including language [2]. Psycholinguistic evidence reveals a positive correlation between mother–child JA and the child’s efficiency in word learning [3], and the frequency of engaging in activities involving JA becomes a reasonablenumber of candidate meanings for target words, thus assisting language acquisition. Meanwhile, comparative evidence reveals significant differences in the level of JA between humans and non-human primates: 9–12 month old human infants can easily establish common ground during interactive or collaborative activities [9], whereas the level of JA in wild/captive non-human primates of different ages is comparatively low, in particular, collaborative intentions are not easy to share among non-human primates or between human experimenters and captive primates [10].

Inspired by these findings, scholars [1,11–13] begin to ascribe the degree-difference in JA between humans and non-human primates to human-unique competences (e.g. shared intentionality and collaboration), and suggest that a fully formed high level of JA in humans is a prerequisite for language and communication [14]. To assess this claim, we need to note that many available comparative data and acquisition or simulation studies are actually insufficient or inappropriate for addressing evolutionary questions, such as whether the initial level of JA in pre-language hominins has to be high, and if not, what factors induced the contemporary degree-difference in JA and how these factors took effect in evolution. For example, based mainly on comparisons between modern humans and contemporary animals, comparative evidence offers no clues to the possible exaptation of JA from non-linguistic domains to the linguistic domain and the intermediate stages during the evolution of the JA level. Focusing mainly on normal/autistic children and word learning, available acquisition studies [15] fail to reveal the JA-language correlation at a social scale and the effect of JA on general language learning in normal children/adults. Moreover, in those simulation studies [8], JA comes to help interpret linguistic utterances only if linguistic knowledge fails to do so, and listeners get direct feedback that clarify speakers’ intended meanings. Such ‘mind-reading’ simplification [16] cannot trace when people start relying on linguistic knowledge and expressions, instead of non-linguistic cues, in comprehension. This transition is a crucial step in language origin.

From an evolutionary perspective and considering the uniqueness of both language and high JA level in humans, we propose a coevolutionary explanation for the degree-difference in JA between humans and non-human primates. Inspired by the formulaic scenario of language origin [17], we assume that in order to interpret
linguistic utterances encoding simple frequent events in the immediate environment, JA was first borrowed by early hominins from non-linguistic general interactions to linguistic communications. Via JA, early hominins developed fundamental linguistic knowledge to express these meanings. Once the JA level becomes correlated with linguistic comprehension, following the socio-biological explanation [18] and gene-cultural or brain-language coevolution hypothesis [19–21], success in linguistic communications could reciprocally cast its influence on the levels of JA among language users, thus triggering a language–JA coevolution (a reciprocal or cooperative influence between two or more natural species or system components [22]). Advocating a continuous development of the JA level, we claim that: (i) the initial level of JA in humans need not be very high; (ii) the degree-difference resulted from a coevolution with language; and (iii) during the coevolution, linguistic comprehension underwent a transition from relying on non-linguistic information to relying on emergent linguistic knowledge. Note that our explanation focuses on the function of JA in language origin, instead of the function of JA in aiding linguistic communications based on a complete set of shared linguistic knowledge.

Considering the limitations in the available data and studies, we adopt a computational model to illustrate this coevolutionary explanation. As a new research means, computer simulation can be viewed as ‘operational’ hypotheses/theories expressed in computer programs, and simulation results of those programs are empirical predictions derived from incorporated hypotheses/theories [23,24]. Motivated by empirical evidence and theoretical argumentation, our language evolution model rests upon three assumptions: (i) communications of early hominins exchange integrated meanings having simple predicate-argument structures [25,26]; (ii) comprehension is determined by either or both linguistic and non-linguistic information, and via JA, listeners can occasionally acquire, from non-linguistic information, the meanings encoded in exchanged utterances; and (iii) non-linguistic information is not always reliable [27] (otherwise, it is equivalent to explicit meaning transfer [16]).

Rather than detailed operations involved in the establishment of common ground during communications, as shown in cognitive neuroscience studies [28,29] we simply simulate JA as the availability of topics from non-linguistic information. In this way, the probability of acquiring correct topics from non-linguistic information quantifies the level of JA, and manipulating it helps reveal the correlation between language and JA. Meanwhile, bio-psychological studies have shown that genetic factors affect the level of JA in humans [30,31]. Accordingly, we assume that the level of JA can be transferred from parents to offspring during genetic transmission. Apart from genetic transmission, cultural transmission of language among individuals from different generations is also included. In this setting, communicative success could cast its influence on the level of JA via natural selection and/or cultural selection, and we can evaluate the roles of either or both types of selections in the possible language–JA coevolution.

The simulation results and relevant analysis show that without the influence of communicative success, in order to trigger a communal language with good understandability, the initial level of JA has to surpass a threshold; however, if communicative success starts affecting parent selection in generation replacement, along with the origin of a communal language with good understandability, an initially low level of JA is boosted and gets ratcheted up to a stable high level. This finding releases the prerequisite of a fully formed high JA level for language origin, and offers a new perspective on the degree-difference in many language-related competences.

2. MATERIAL AND METHODS

(a) Language evolution model

The adopted model studies whether a population of interacting individuals (artificial agents) can, based on general learning mechanisms, develop a compositional language out of a holistic signalling system [32,33]. The emergent language in this model consists of a set of common lexical items and consistent word order(s) to encode integrated meanings with simple predicate-argument structures, such as ‘chase<wolf, sheep>’ or ‘jump<lion>’. Figure 1 shows the conceptual framework of the model. In a nutshell, during iterated communications, individuals can: (i) acquire lexical items from recurrent patterns in exchanged sentences; (ii) associate lexical items with
categories according to semantic and sequential information of lexical items in exchanged sentences; and (iii) combine local orders among categories to form global orders in sentences encoding integrated meanings.

In this section, we introduce the major components of the model relevant for the language–JA coevolution, including the comprehension process, transmission framework and evaluating indices. Further details of the model (e.g. language encoding, acquisition mechanisms, communication scenario and origin of compositional structures out of holistic phrases) can be found in the electronic supplementary material, A.

In this model, non-linguistic information is simulated as environmental cues, each containing an integrated meaning. Imagine when the speaker produces an utterance encoding an ongoing environmental event, if the listener has a high level of JA, by following the speaker’s gaze or other hints, they could easily attain to the appropriate event and obtain a correct cue containing the speaker’s intended meaning encoded in the utterance. Meanwhile, early hominins, owing to restricted cognitive abilities, may have low levels of JA, therefore, common ground is not always established in communications (e.g. the listener may pay attention to another event and get a wrong cue containing a meaning distinct from the speaker’s intended one). In other words, cues are unreliable. In order to simplify the comprehension process and quantify the level of JA, we assume that a listener only obtains one cue in each round of utterance exchange (see the electronic supplementary material, A3), and define reliability of cue (RC) as the probability for a listener to obtain the correct cue. For example, if RC is 0.6, the listener has a 60 per cent chance of obtaining the correct cue in each round of utterance exchange; otherwise, they obtain a wrong cue. Both exchanged integrated meanings and cues are chosen from a predefined semantic space. In the case where a listener obtains a wrong cue, any integrated meaning distinct from the speaker’s intended one has an equal chance to be the cue.

Comprehension is determined by the non-linguistic cue and/or available linguistic rules (see the electronic supplementary material, A2 for how to acquire such rules) that can completely or partially interpret the heard utterance. This process proceeds as follows.

The listener first compares the cue’s meaning with the one(s) interpreted using its available linguistic rules. This comparison leads to three outcomes (figure 2):

— If there is an exact match between the two, the cue joins the available linguistic rules to form a candidate set for comprehension, the meaning of which is the cue’s meaning.

— If there is a partial match between the two, the cue also joins the linguistic rules for comprehension. For example, if the available rules provide an incomplete integrated meaning, ‘chase< tiger, #>’ (‘#’ denotes an unspecified constituent), and the cue says ‘chase< tiger, sheep>’, since the constituents specified by the linguistic rules match exactly those in the cue’s meaning, the cue and linguistic rules form a candidate set for comprehension, the meaning of which is the cue’s meaning, ‘chase< tiger, sheep>’.

— If there is no match between the two, or no linguistic rules are available for interpreting the heard utterance, the cue itself forms a candidate set for comprehension. For example, if the linguistic rules offer an interpretation, ‘run< tiger>’, but the cue says ‘fight< wolf, tiger>’, then, the cue forms another candidate set for comprehension, the meaning of which is ‘fight< wolf, tiger>’. This set will compete with the set formed by the linguistic rules, the meaning of which is ‘run< tiger>’.

After setting up the candidate sets for comprehension, the listener selects the set having the highest combined strength and interprets the heard utterance accordingly (see the electronic supplementary material, A3).

This comprehension scenario differs from the cross-situational learning [16,34] in several aspects. In that paradigm, the real meaning of the exchanged word is always available to the listener in each round of learning/communication, together with different distracting meanings. Such frequent advantage of the real meaning makes sure that the listener eventually forms a strong mapping between the exchanged word and its real meaning. In our model, however, it is quite possible that in some rounds of utterance exchange the listener only obtains wrong cues. In addition, the mappings between integrated meanings and utterances are evolving, instead of predefined, and the transition from relying on non-linguistic information to relying on shared reliable linguistic knowledge occurs naturally.

This model incorporates both genetic transmission, i.e. transmission of the level of JA (RC) from adults to offspring during reproduction, and cultural transmission, i.e. intra-(adults talking to each other) and inter-generational (adults talking to offspring) communications (figure 3). In each generation, intra-generational transmissions take place first. After that, half of the adults are chosen as parents, each producing two (to maintain the population size) offspring (new agents) having no linguistic knowledge but copying their parents’ RC values with occasional mutations. Then, inter-generational transmissions take place. Later on, offspring replace their parents, and a new generation begins. Such a punctuated setting is to clearly trace the
evolution across generations. In real situations, cultural and genetic transmissions could be intertwined.

An individual’s communicative success (CS) reflects its fitness in the population. CS is measured as the mean percentage of integrated meanings an individual can accurately understand (using linguistic knowledge, not cues) when others talk to this individual (see equation (2.1)):

\[
CS = \frac{\sum \text{no. accurately understood integrated meanings when agent } j \text{ talks to agent } i}{\text{size of semantic space} \times (\text{size of population} - 1)}
\]

Both natural and cultural selections take effect on the basis of CS: natural selection affects reproduction by selecting adults who can better understand others (having higher CS) as parents to produce offspring; cultural selection affects inter-generational transmissions by choosing adults having higher CS as teachers to talk to offspring. Here, we only consider one type of cultural selection that is relevant for CS, the actual roles of cultural selection in language evolution are manifest in many aspects [35–44]. The mean CS of all individuals is understanding rate (UR), a high value of which indicates that the communal language used by individuals can accurately exchange many meanings in the semantic space. In other words, UR reflects the degree of linguistic mutual understandability.

(b) Simulation setup
We set up five sets of simulations. The NoChangeRC set reveals that both cultural and natural selections, nor mutation on RC; both parents and teachers are randomly chosen and offspring copy their parents’ RC with no 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, in which natural and cultural selections are two factors, each having two levels. When natural selection is in effect, adults with higher CS have higher probability of being parents; otherwise, parents are chosen randomly. When cultural selection is in effect, adults with higher CS have higher probability of being teachers; otherwise, teachers are chosen randomly. In these four sets, offspring copy their parents’ RC with an occasional mutation (increase or decrease RC with a fixed amount).

In all sets, the initial RC values in the first generation are randomly chosen from Gaussian distributions, the s.d. of which are fixed at 0.01 but means range from 0.1 to 0.9, with a step of 0.1. This manipulation preserves the general characteristic of the population and incorporates a certain degree of individual difference. According to the means of the initial RC, we further divided each set of simulations into nine RC conditions, and conduct 50 simulations in each condition. Obviously, simulations under initially low RC values are more relevant for studying the language–JA coevolution, but those under initially high RC values are also useful for systematically evaluating the effects of natural and cultural selections on the coevolution. In each simulation, we measure UR of the communal language and RC of individuals at 51 sampling points evenly distributed in the total number of generations.

Table 1 shows the parameter setting for the natural and cultural selections (other parameters defining the semantic and signalling spaces, acquisition and communication mechanisms are shown in the electronic supplementary material, table A1). In the simulations reported in this paper, there are 64 integrated meanings for individuals to exchange, each having the same chance to be produced during communications. Individuals in the first generation can only encode eight integrated meanings. This resembles a limited signalling system of early hominins (in fact, simulations starting from no linguistic knowledge report similar results). This model does not simulate the acquisition of new semantic constituents, so those eight meanings contain all semantic constituents used to form the 64 integrated meanings. As shown in [32], if the size of the semantic space or the population increases, similar results can be obtained given more rounds of cultural transmission per generation. In order to clearly observe the coevolution and systematically evaluate the roles of relevant factors in the coevolution, we set the number of generations to 1000, 2000 and 5000.

3. RESULTS
Simulations in the NoChangeRC set reveal a correlation between the level of JA and UR of the emergent language. As shown in figure 4a, in all three numbers of generations, if the initial RC is below 0.5, UR remains low; once RC surpasses 0.5, UR starts increasing along with RC. These results reveal a threshold RC (around 0.5), only beyond which can a communal language with good understandability emerge. Meanwhile, figure 4b records a communal language which emerges under initial RC 0.7 at generation 580. This language consists of seven lexical rules and a consistent SVO order formed by three local orders, SV, VO and SO. Its high UR (0.882) indicates that this language is able to accurately exchange many integrated meanings in the semantic space. A detailed discussion on the emergent word orders in this model can be found in Gong et al. [45].
Simulations in the other four sets reveal a coevolution of language (in terms of UR) and JA level (in terms of RC). Let us take the results obtained in 1000 generations as an example (the similar results obtained in 2000 and 5000 generations are shown in the electronic supplementary material; B, C).

As for UR, a two-way analysis of covariance (ANCOVA [46]) (dependent variable: mean UR over 50 simulations; fixed factors: natural and cultural selections; random factor: RC conditions; covariate: sampling points throughout 1000 generations) shows that: natural selection has a significant main effect on UR ($F_{1,8} = 36.272, p < 0.001, \eta_p^2 = 0.819$), but cultural selection does not ($F_{1,8} = 0.913, p = 0.367, \eta_p^2 = 0.102$); and these two types of selection interact significantly ($F_{1,8} = 0.426, p = 0.532, \eta_p^2 = 0.051$). In addition, RC condition has a significant main effect on UR ($F_{8,7.817} = 5.861, p < 0.012, \eta_p^2 = 0.857$), and interacts significantly with natural selection ($F_{8,8,9} = 39.861, p < 0.001, \eta_p^2 = 0.976$). Figure 5a,b shows these results. Finally, the covariate, generation (sampling points), also interacts significantly with UR ($F_{1,1799} = 184.286, p < 0.001, \eta_p^2 = 0.093$). Here, using ANCOVA (not ANOVA) is to partial out the influence of the covariant.

As shown in figure 5a, the marginal mean UR in the set where natural selection is in effect (Nat_NoCul or Nat_Cul) is significantly higher than that in the set where natural selection is not (NoNat_NoCul or NoNat_Cul), but the marginal mean UR in the set where cultural selection is in effect (NoNat_NoCul or NoNat_Cul) is not significantly different from that in the set where cultural selection is not (NoNat_NoCul or NoNat_Cul). These results indicate that it is natural selection, rather than cultural selection, that drives the origin of a communal language with good understandability. Meanwhile, as shown in figure 5b, there is a correlation between UR and RC in the set where natural selection is not in effect (NoNat_NoCul or NoNat_Cul) (black bars), which is similar to what is shown in figure 4a. However, when natural selection is in effect (white bars), a high UR becomes available even under an initially low RC. This is more explicit when the initial RC is below the threshold (0.5), but less so when it exceeds the threshold.

As for RC, a similar ANCOVA (dependent variable: mean RC over 50 simulations) shows that: natural selection ($F_{1,8} = 41.522, p < 0.001, \eta_p^2 = 0.838$) and RC condition ($F_{8,8,023} = 73.554, p < 0.001, \eta_p^2 = 0.987$) have significant main effects on RC, but cultural selection does not ($F_{1,8} = 0.223, p = 0.649, \eta_p^2 = 0.027$); and there is no significant interaction between the two types of selection ($F_{1,8} = 0.004, p = 0.953, \eta_p^2 = 0.000$), but natural selection interacts significantly with RC condition ($F_{8,8} = 9.388, p < 0.003, \eta_p^2 = 0.900$). Figure 5c,d shows these results. The covariate also interacts significantly with RC ($F_{1,1799} = 155.785, p < 0.001, \eta_p^2 = 0.080$).

Figure 5c confirms that the evolution of RC is also achieved mainly by natural selection, not cultural selection. Figure 5d reveals two effects of natural selection on RC. If the initial RC is below the threshold (0.5), natural selection will significantly increase the marginal mean RC, e.g. the marginal mean RC increases to 0.4 under initial RC 0.2. However, if the initial RC exceeds the threshold, natural selection will preserve the marginal mean RC at that level across generations, e.g. under initial RC 0.8, with natural selection, the marginal mean RC remains around 0.8, but without natural selection, it drops to 0.6.

Noting the significant interaction between natural selection and RC condition, we further analyse the evolution of RC in different RC conditions. Two evolutions, respectively, in RC conditions 0.4 and 0.7 are shown in figure 6 (the results in the other RC conditions are shown in the electronic supplementary material). Evident in these figures, with natural selection, an initially low RC (say, 0.4) increases, whereas an initially high RC (say, 0.7) remains. Figure 6a also shows that the increasing and maintaining effects take place consecutively: with natural selection, owing to the increasing effect, the mean RC gradually increases above the threshold, and then, the maintaining effect kicks in, restricting the mean RC from further increasing and maintaining its level across generations. These results reveal the complete trajectory of the evolution of RC: with natural selection, initially low RC increases, and once exceeding the threshold, it remains at that level.

Combining all the results, these simulations clearly illustrate a language–JA coevolution. The initial stage of
the coevolution is a low level of JA, plus no or limited linguistic knowledge. Via genetic and cultural transmissions, both the level of JA and linguistic knowledge are evolving. After a number of generations, the level of JA gradually reaches a stable, relatively high level and a set of linguistic knowledge becomes available for reliably describing many integrated meanings.

4. DISCUSSIONS

(a) Threshold RC and driving force for the coevolution

The threshold RC shown in the NoChangeRC set of simulations can be explained as follows. In the language evolution model, whenever linguistic knowledge is insufficient, comprehension has to refer to cues obtained via JA. If some linguistic knowledge happens to help encode certain constituents in cues’ meanings, the strength of this knowledge will gradually increase in communications. When linguistic knowledge is sufficient to encode complete integrated meanings and its strength exceeds the cue strength, comprehension will naturally shift to relying on linguistic knowledge. With more linguistic knowledge being shared, linguistic comprehension becomes more reliable, and a communal language with good understandability emerges. During this transition, a sufficient level of JA is necessary for individuals to establish enough common ground to develop and share reliable linguistic knowledge.

The results of the NoNat_NoCul, Nat_NoCul, NoNat_Cul and Nat_Cul sets of simulations indicate that the language–JA coevolution is mainly driven by CS. During inter-generational transmissions, a higher level of JA helps an offspring acquire adults’ language; during intra-generational transmission, a higher level of JA helps adjust one’s idiolect to be similar to others’. Therefore, an individual with a higher level of JA tends to have higher CS. Then, with natural selection, such an individual will have a higher chance to reproduce and spread their level of JA in future generations, thus leading to the gradual increase in the initially low level of JA.

Two points are worth noting here. First, the JA level never reaches its maximum (1.0). During the coevolution, once JA exceeds the threshold, acquisition of reliable linguistic knowledge becomes relatively easy and the transition to relying on linguistic knowledge can be efficiently achieved, then, without further relying on non-linguistic information; the JA level stops increasing. Second, a sufficiently high JA level will not significantly drop. When all individuals already develop a communal language, JA would not greatly affect comprehension in intra-generational transmissions. But during teaching, an offspring, without any linguistic knowledge and with a reduced JA level, will fail to efficiently acquire the adults’ language. After becoming an adult, this individual would not clearly understand others. Therefore, it would not have a high CS and many chances to spread their reduced level of JA in future generations. Then, the reduced JA level will not diffuse in the population. On this point, our study illustrates a ratchet effect [47,48]. It restricts the JA level from significantly dropping and is manifest mainly during inter-generational transmissions (teaching). The ratchet effect on socio-cognitive or cultural abilities has been previously discussed from a theoretical
Finally, the language–JA coevolution shown in our study is in line with the recent theoretical discussions on the possible restrictions on language–gene coevolution [50]. In this regard, simulations demonstrate that genetic assimilation in the context of language evolution can retain and expand communicatively effective features [51], rather than abstract or arbitrary ones as hypothesized in the classic Universal Grammar [52]. JA is not language-specific, exists prior to language and takes effect during general communicative activities. As shown in our study, once JA starts aiding linguistic comprehension and CS gains some functional advantage (e.g. influencing reproduction opportunity), under the drive of CS, JA could piggyback on language, having its own level increased and ratcheted along with language evolution.

In summary, this work explicitly shows that language evolution results from an interaction of biological evolution (e.g. genetic assimilation of language-related abilities), individual learning (e.g. language-processing mechanisms) and cultural transmission (e.g. teaching) [35,43].

(c) Limitations and future directions
Simulations only show what could happen, not necessarily what must have happened [53]. A systematic analysis of a simulation study should focus not only on its major findings, but also on its major limitations. The model adopted in this study inevitably suffers from certain limitations relating to the simulation, theoretical and empirical aspects.

On the simulation aspect, the model assumes a shared semantic space among individuals. Simple semantics could emerge as shared conceptual primitives in individual mental spaces [54], and apart from humans, non-human primates can conceptualize simple predicate-argument structures [25]. Such simple meanings as simulated in our study can also be easily reconstructed via JA, without referring to much linguistic or contextual information [55]. Therefore, assuming a shared semantic space containing such simple meanings is still plausible, and such assumption has been widely adopted in simulations of syntactic evolution [35,38], but not semantic evolution or symbolic grounding [56], and the coevolution might take place mainly during the origin of a basic language as the one that emerged in this model. In addition, certain parameters (e.g. the number of generations) are arbitrarily set in the simulations. On this point, statistical analysis (e.g. ANCOVA) helps obtain general conclusions across different settings, or partial out the influence of certain parameters. Furthermore, one may question the linguistic complexity involved in this model, covering both lexical and syntactic aspects. Language is more than lexical items, the evolution of lexical items and syntax may be quite different, but still closely related. Apart from lexical items, a general model of language processing should at least cover some syntactic aspects (e.g. the basic word order as in this model). Meanwhile, as advocated by some scholars [17,26], language may originate from holophrastic expressions encoding simple meanings. Although this issue is subject to heated discussion [57–59], simulations incorporating these different hypotheses help reconstruct processes not easily observed in modern situations, and thus provide the basis for evaluating theories that incorporate such processes in their explanations of language origin.

On the theoretical and empirical aspects, there is no decisive evidence that JA is genetically transmitted.
Some scholars [60] argue that JA, like other language-related competences, could be recruited for linguistic activities to achieve high CS. The recent findings that training induces neuro-anatomical changes in monkey and human brains [61] could serve as indirect evidence for this recruitment theory. Nonetheless, the general conclusion of our study still holds, i.e. the degree-difference in JA between humans and non-human primates is owing to a selective pressure from CS. In addition, artificial agents in the model are assumed to understand communicative intentions, but can they detect such intentions before a reliable communication system is available? Non-linguistic communication games among human subjects [62] have revealed a high level of meta-cognitive ability in humans [63] for detecting communicative intentions before a reliable communication system is in place [46]. Furthermore, apart from lexical meanings, we are uncertain whether JA also assists comprehending propositional meanings having simple predicate-argument structures. We also lack empirical investigations on whether improved linguistic abilities could enhance, in a reverse manner, the level of JA in children. Apart from patient studies [30,31], we do not know how normal children, at the early stage of language acquisition, coordinate language with other types of cues and when children start solely relying on language and ignoring non-linguistic cues. All these will inspire more empirical investigations and simulations of language development, and help further evaluate language–JA coevolution.

5. CONCLUSIONS

We conducted a simulation study to explore the possible language–JA coevolution in a cultural environment involving multiple individuals. The simulation results illustrate that culturally constituted aspects (e.g. CS) can drive the natural selection of predisposed cognitive features (e.g. JA). Such a coevolutionary perspective is insightful for explaining the degree-difference in many language-related competences (e.g. memory, learning mechanisms, socio-cognitive abilities, etc.), between humans and non-human primates. This study belongs to the type of work called for in Beckner et al. [64]. It is grounded in, and in turn contributes to the current trends in language evolution, complex systems and primates–human comparisons.

This work was supported in part by the Society of Scholars in the Humanities in the University of Hong Kong. We thank Morten Christiansen from Cornell University for valuable comments on this work.

REFERENCES
