



The evolution of cognitive mechanisms in response to cultural innovations

Arnon Lotem^a, Joseph Y. Halpern^b, Shimon Edelman^c, and Oren Kolodny^{d,1}

^aDepartment of Zoology, Tel Aviv University, Tel Aviv 6997801, Israel; ^bDepartment of Computer Science, Cornell University, Ithaca, NY 14850; ^cDepartment of Psychology, Cornell University, Ithaca, NY 14850; and ^dDepartment of Biology, Stanford University, Stanford, CA 94305

Edited by Kevin N. Laland, University of St. Andrews, St. Andrews, United Kingdom, and accepted by Editorial Board Member Andrew G. Clark April 29, 2017 (received for review January 20, 2017)

When humans and other animals make cultural innovations, they also change their environment, thereby imposing new selective pressures that can modify their biological traits. For example, there is evidence that dairy farming by humans favored alleles for adult lactose tolerance. Similarly, the invention of cooking possibly affected the evolution of jaw and tooth morphology. However, when it comes to cognitive traits and learning mechanisms, it is much more difficult to determine whether and how their evolution was affected by culture or by their use in cultural transmission. Here we argue that, excluding very recent cultural innovations, the assumption that culture shaped the evolution of cognition is both more parsimonious and more productive than assuming the opposite. In considering how culture shapes cognition, we suggest that a process-level model of cognitive evolution is necessary and offer such a model. The model employs relatively simple coevolving mechanisms of learning and data acquisition that jointly construct a complex network of a type previously shown to be capable of supporting a range of cognitive abilities. The evolution of cognition, and thus the effect of culture on cognitive evolution, is captured through small modifications of these coevolving learning and data-acquisition mechanisms, whose coordinated action is critical for building an effective network. We use the model to show how these mechanisms are likely to evolve in response to cultural phenomena, such as language and tool-making, which are associated with major changes in data patterns and with new computational and statistical challenges.

tool-making | language evolution | niche construction | cognitive evolution | social learning

An open question in the study of culture and cognitive evolution is whether (and to what extent) cognitive mechanisms, especially those viewed as advanced or sophisticated, evolved in response to social-learning challenges or are merely the product of domain-general mechanisms (1–3). According to one view—still widely held in cognitive science and evolutionary psychology—cognitive adaptations take the form of specialized brain modules (or neuronal mechanisms) that evolved for specific, often social purposes, such as “imitation” (4, 5), “mind reading” (6, 7), “cheating detection” (8), or most famously, language acquisition (9, 10). These ideas have been criticized on theoretical and empirical grounds (11, 12), and the debate around them demonstrates our limited understanding of the evolution of cognition, its relationship to the evolution of social behavior and, in some organisms, culture.

The question of whether culture and social behavior shape the evolution of the brain is, in our view, best considered using the evolutionary framework of niche construction (13–16): that is, culture and social behavior change the ecological niche to which cognitive traits must adapt in the same manner that nest-building by birds changes the ecological niche in which their nestlings evolve. For example, animals’ ability to learn from each other may have initially been a by-product of domain-general associative learning mechanisms that did not evolve for social learning (1). However, as soon as these mechanisms enabled social learning and were recruited by it for regular use, social learning and its outcomes also became part of their ecological niche. From that

moment on, learning mechanisms that need not have initially been specifically social were also selected according to their ability to support social learning. If that is the case, one can certainly claim that these mechanisms were adapted or shaped to serve their new social function (although using the term “evolved for” may still be premature without knowing the degree of genetic modification and specialization).

Similarly, when social learning enables the accumulation or spread of shared group behaviors—these days recognized as the formation of “culture” (17, 18)—this culture becomes the new ecological niche for all of the learning mechanisms that contribute to it, and therefore has the potential to shape their evolution. Thus, in theory, given sufficient evolutionary time, cultural phenomena that are adaptive for the individual, and whose acquisition is supported by advanced learning or cognitive skills, such as the ability to imitate or to learn language, are expected to select for improvements in these cognitive skills (see also ref. 19). In practice, however, clear evidence showing the effect of culture on cognition is lacking, and alternative accounts for the evolution of advanced cognition and culture through domain-general learning principles cannot be ruled out (1, 2, 20). As a result, whether and how culture really shapes the evolution of cognition is still under debate.

In what follows, we first clarify some of the theoretical issues in this debate, using two recent controversies in the fields of language evolution and social learning. We then offer a process-level approach to cognitive evolution that may be useful in predicting what aspects of learning and cognition are likely to coevolve with culture. Finally, we use the model to demonstrate how cultural phenomena such as language and tool-making (each related to one of the two controversies discussed earlier) are likely to shape cognition, given their association with changes in data distribution and with new computational and statistical challenges.

Can Culture Evolve Without Shaping Cognition? On Parsimony, Likelihood, and Scientific Productivity

Evolution takes time, so it is clear that very recent cultural innovations, such as cars, computers, cellular phones, or the Internet, could not have yet generated detectable effects (or perhaps any effect at all) on the evolution of cognition. But what about relatively ancient cultural phenomena, such as song-learning in birds or tool-making and language acquisition in humans? Although there is evidence for the effect of human culture on biological traits and gene frequencies (21), evidence for specific effects of human culture on learning and cognitive mechanisms is mostly circumstantial. This evidence includes

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “The Extension of Biology Through Culture,” held November 16–17, 2016, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/Extension_of_Biology_Through_Culture.

Author contributions: A.L., J.Y.H., S.E., and O.K. designed research, performed research, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. K.N.L. is a guest editor invited by the Editorial Board.

¹To whom correspondence should be addressed. Email: okolodny@stanford.edu.

signs of selection on genes implicated in brain growth, learning, and cognition (22–24) that may be attributed to human culture (21), and differences in gene expression in the brain between human and nonhuman primates (25, 26) that may be interpreted similarly. There is also recent evidence relating structural changes in the human brain to Paleolithic tool-making abilities (27, 28), but additional work is still needed to clarify the direction of causation between culture and cognition (we will return to discuss these findings toward the end of the paper). Finally, when animal culture is considered, a recent study (29) suggests an effect of culture on learning in songbirds: a larger repertoire size is found in species that developed open-ended learning ability.

The lack of clear empirical evidence for the effect of culture on cognition is not surprising, given that cognitive mechanisms and their genetic underpinning are still poorly understood, making it difficult to track their evolution (as opposed to that of clearly defined biological traits). As a result, much of the debate is focused on theoretical arguments of plausibility and likelihood, which may be interpreted differently by psychologists and evolutionary biologists. We use two examples of such controversies to illustrate this problem and to suggest a methodologically productive resolution.

Problem 1: The Evolution of Social-Learning Mechanisms. Social learning is broadly defined as learning that is influenced by observation or interaction with other individuals or with the products of their behavior (30). This definition leaves open the question of whether social-learning mechanisms have evolved specifically to serve their social function or whether they are domain-general associative learning mechanisms that are also used to learn socially. In a series of thought-provoking papers, Heyes and her colleagues (1, 2, 31–33) have demonstrated that most mechanisms of social learning and imitation that are normally viewed as specialized adaptations for social life (4, 5, 34) can also be explained by domain-general associative learning principles. In this light, and in the absence of convincing evidence to the contrary, they also suggested that there is no need to posit that these domain-general mechanisms were shaped by their social or cultural function (35). In other words, it would appear more parsimonious to assume that these mechanisms did not evolve beyond their initial domain-general form. However, this appeal to parsimony is somewhat misleading in evolutionary contexts and time scales, where changes are actually to be expected (36, 37). In fact, for most evolutionary biologists, it would appear highly unlikely that some learning mechanisms would be used for many generations to serve social functions, yet remain unaffected by this new social niche. It would almost be like expecting the surface of the moon to remain unmarked with craters after millions of years of exposure to space debris.

To explain why, imagine a population of finches that expands its range. The beaks of these finches have a certain morphology that evolved to handle seed types present in the original habitat; the expanded habitat includes novel types of seeds. In this setting, the assumption that after many generations there would be no evolutionary change in beak morphology is not at all parsimonious. For a biological trait to remain unchanged over time, an active process of stabilizing selection is necessary, otherwise it will change through directional selection or genetic drift. That is, the probability that the new types of seeds will not affect the previous stabilizing selection regime is extremely small.

Returning to the evolution of learning, and following the same reasoning, it is difficult to imagine how adding a new function for a basic learning mechanism would not affect its evolution. As in the beak example, the change may be subtle, and merely quantitative. This is in part because a bird has only one beak that must serve many different functions (from foraging for different types of food to feather-preening and nest-building); it cannot be specialized into different kinds of beaks. However, even in this case of a single multipurpose adaptation, every new function must affect evolution; we would expect a more significant change if specialization is possible. Similarly, even if all cognitive functions are supported by the same domain-general learning mechanisms, it is unlikely that

social learning and culture could evolve without somehow affecting these mechanisms. In reality, of course, unlike in the beak example, learning mechanisms are not necessarily constrained to be uniform across all domains; there is plenty of evidence for adaptive specialization in associative learning mechanisms (e.g., refs. 38–40).

The key point in this evolutionary argument is that, even in the absence of supportive evidence, it is more parsimonious to assume that learning mechanisms were shaped by their social function and, if the relevant evidence is lacking, that we have simply failed to find it, than to assume the opposite. The assumption of no change requires us to posit a lack of genetic variance in learning mechanisms, which runs counter to substantial evidence (39, 41–45), or else to explain how the selection regime was miraculously unaffected by the new social niche. The assumption of evolutionary change may also be viewed as scientifically more productive, as it encourages further research (46). This implies that a useful working hypothesis should be that “culture did shape cognition” and that we have to find out how.

Demonstrating that social learning and the so-called “mirror neurons” phenomenon can be explained by associative learning principles (1, 2) is important and consistent with our view. But given the evolutionary argument above, the demonstration need not imply that these associative learning mechanisms did not evolve beyond their basic state (see also ref. 47). Indeed, the argument suggests that associative learning mechanisms are the building blocks of cognitive evolution, and are finely tuned to serve new social, cultural, and other advanced functions (48). This view may help us make the shift from postulating “black box” adaptations that evolve “for” particular social purposes to having credible process-level mechanistic models of cognitive evolution.

Problem 2: The Evolution of Language and Memory Constraints. The same evolutionary argument discussed above is also relevant to the question of whether or not human language shaped the evolution of cognition (16, 49–52). It implies that it is highly unlikely that the use of language for at least several thousand generations [or even more (53–55)] failed to affect some aspects of learning and cognition. The real question is not whether or not it did, but rather what aspects were affected and how. For example, Chater et al. have pointed out correctly that many aspects of human language change too fast for genetic evolution to respond (56). The authors used computer simulations to show that even in the presence of genetic variation, cultural conventions of language are like “moving targets” for natural selection, making the evolution of genetic adaptations to specific languages highly implausible. However, although this analysis makes a convincing argument against strong nativism (the claim that there is a significant innate component to human language), it also implies that genes for language can evolve if they serve general skills for language learning that are stable over time (49, 56, 57). In other words, the need to learn a language may still select for many general cognitive abilities, such as better memory, computational abilities, or greater attention to verbal input. Indeed, it does not seem possible that language could evolve without affecting such abilities.

Interestingly, in a more recent paper, Christiansen and Chater (58) have extended their view that language cannot shape the evolution of cognition by proposing that the limited sensory memory span—the window of time during which a linguistic utterance is retained in its entirety—creates a bottleneck that strongly constrains language and cannot evolve to become wider. The authors suggest that language has evolved to cope with this memory limitation, but that the evolution of this memory limitation was not affected by language. As we noted elsewhere (59), this assumption runs counter to the evolutionary argument above and to substantial evidence for genetic variance in memory parameters (60–63). In a reply to this criticism, Chater and Christiansen (64) explained that the memory bottleneck cannot be viewed as a genetically variable trait that can respond freely to selection because it “emerges from the computational architecture

of the brain.” This answer, however, merely kicks the can down the road, by moving the problem from the domain of memory to that of the computational architecture of the brain. The same arguments hold: although constrained by many factors, the computational architecture of the brain is shaped by the sum of selective pressures arising from the need to accommodate the multiple functions of the brain that influence the organism’s fitness. Even if the memory bottleneck emerges as a product of this architecture, it can still evolve as long as this architecture evolves. If the challenge of processing and using language, which affects individuals’ fitness and has been in place for thousands of generations, has played a role in shaping this architecture and its emergent properties, then, as one of these emergent properties, memory should have been affected. Specifically, as we proposed in the past (59) and explain further below, the selective pressure exerted by language learning may have acted to limit the working-memory buffer, as this may be useful for coping with the computational challenges involved in data segmentation and network construction.

Our two examples—social learning and memory bottleneck—suggest that, excluding very recent cultural innovations, it is unlikely that culture could have evolved without shaping learning and cognition. This forces us to think more specifically about how culture shapes cognition, which requires, as we claim next, adopting a process-level approach to cognitive evolution: that is, a mechanistic model that explains a behavior or an ability as the outcome of a process. Such a model may also help provide a useful structure for reexamining the two problems outlined above.

Why Do We Need a Process-Level Approach to Theorize About Culture and Cognition?

Whereas it is relatively easy to see how natural selection acts on clearly defined morphological traits, such as limbs, bones, or coloration, with cognitive traits that are not well understood, it is difficult to tell what is actually evolving. Cognition is not a physical trait, but an emergent property of processes that are carried out by multiple mechanisms, most of which involve learning. Thus, to consider how culture shapes the evolution of cognition, we must explain how such mechanisms work and how they can be modified by natural selection. The importance of using mechanistic models in the study of behavioral evolution is increasingly recognized (65–68), but most attempts to integrate evolutionary theory and cognition are still based on modeling the evolution of learning rules that are far too simple to capture complex cognition (69–74). To understand how culture shapes the evolution of cognitive mechanisms, such as those serving imitation, theory of mind, or language acquisition, it is necessary to have models that explain how such mechanisms work and how they could evolve.

Clearly, given the immense complexity of the brain, any attempt to propose a general process-level model of advanced cognition would be ambitious. However, we believe that it is possible and necessary to start by constructing models that capture some of the key working principles of advanced cognitive mechanisms in a manner that suffices to explain their evolution. An analogy that may clarify our approach is the apparent challenge in explaining the evolution of the eye. The vertebrate eye is highly complex; it is initially hard to see how it could have evolved. However, with a minimal understanding of how the eye works, the “magic” is removed (75). The basic eye model is a layer of photosensitive cells; the visual acuity it provides can gradually improve as it buckles into a ball-like shape, looking (and working) more and more like a pinhole camera. This sketch ignores many details, and is far from explaining everything about eyes and vision, but is sufficient to resolve the puzzle.

This is the kind of modeling approach that we seek for explaining cognitive evolution. Specifically, we do not seek a fully detailed neuronal-level model of brain and cognition. Instead, we want a minimal set of principles that suffice to explain how simple operational units, capable of only the most basic forms of learning, can jointly and gradually create the much more sophisticated mechanisms of advanced cognition. [Powerful algorithms using deep

learning (76, 77), with less emphasis on biological and behavioral realism (78), are being developed and applied to challenging tasks, including: perceptual parsing, associative learning, and the learning of conceptual contingencies (79, 80).]

Once we have such a minimal model, we can consider how small variations in the basic operational units and their parameters can build better, or different, cognitive mechanisms and how this evolutionary process can be shaped by culture. Over the past few years, we have developed such a model and explored its ability to explain a range of phenomena. In the following sections, we briefly describe this model and use it to consider how culture may shape the evolution of cognition and, in particular, how such a model may help resolve the abovementioned controversies regarding social-learning mechanisms and memory constraints on language.

A Process-Level Model of Cognitive Evolution

The model presented here has already been described in several of our previous papers (81–87). Some of the main aspects of the model were implemented in a set of computer simulations, demonstrating a gradual evolutionary trajectory, from simple associative learning, to chaining, to seldom-reinforced continuous learning [in which a network model of the environment is constructed (84)], to complex hierarchical sequential learning that can support advanced cognitive abilities of the kind needed for language acquisition and for creativity (85, 87). For the latter, our modeling framework had to go well beyond chaining through second-order conditioning (84, 88). The success of a computer program, originally developed to simulate the behavior of animals learning to forage for food in structured environments (85), in reproducing a range of findings in human language (86) suggests that the model may be useful in the study of cognitive evolution. Thus far the model’s implementation has been limited, for simplicity, to an unsupervised learning mode with a learning phase, and then a test phase during which the learners act based on what they have learned. Its extension to accommodate iterated cycles of learning and action, which is necessary to capture the learning of behavioral contingencies through trial and error, is straightforward. Detailed pseudocode for our model is found in the supplementary material of refs. 85 and 86.

The model is based on coevolving mechanisms of learning and data acquisition that jointly construct a complex network that represents the environment and is used for computing adaptive responses to challenges in the environment. In particular, the network is used for search, prediction, decision making, and generating behavioral sequences (including language utterances, when applicable). The extent to which learners’ use of the network produced adaptive behaviors was measured in our implementation by foraging success [in the context of animal foraging (84, 85, 87)] and by a set of language performance scores [in the context of language learning (86)]. Although the production of adaptive behaviors depends on the structure of the network, this structure is not directly coded by genes and therefore cannot evolve directly. The components that may evolve over generations are the parameters of the learning and data-acquisition mechanisms that construct the network through interaction with the environment, and whose coordinated action, as we show below, is critical for building the network appropriately. [This coevolution is very much in the spirit of the notions of constructive development and reciprocal causation in the recently proposed “extended evolutionary synthesis” (89).]

Constructing a Network. We now briefly sketch the main principles that govern how the network is constructed. (This technical description may become clearer and more intuitive after reading the simplified example outlined in the next subsection and illustrated by Fig. 1). We assume that what data are acquired by the learner is determined by its “data-acquisition mechanisms”: the collection of sensory, attentional, and motivational mechanisms that direct the learner to process and acquire whatever is deemed relevant. These mechanisms [also referred to as “input

mechanisms” (33)] determine the content and the distribution of different data items in the input. For simplicity, we assume that the input takes the form of strings of symbols (i.e., linear sequences of discrete items), which are then processed through a limited working-memory buffer, similar to the “phonological loop” in humans (90, 91), and tested for familiar segments and statistical regularities among their components. This is done by the learning mechanisms in a sequence of steps.

A data sequence is scanned for subsequences that recur within it and for previously learned subsequences, and is segmented accordingly. This results in a series of chunks, which are either previously known or are incorporated at this point into a network of nodes that represents the world as it has been learned so far (nodes stand for objects or other meaningful units; the links in the network represent their association in time and space). Weights are assigned to the nodes and to the links to reflect their frequency of occurrence: links between nodes are established whenever two nodes follow one another in the input. The weight of a node or a link is increased whenever it is encountered in the input; the weight also decreases with time if it is not encountered. This process ensures that only those units and relations that are potentially meaningful are retained in memory, and spurious occurrences are forgotten. If a node’s or link’s weight increases above a fixation threshold, its decay becomes highly improbable. The probability that a data item is learned is thus determined by how frequently it is encountered in the data, and

by the parameters of weight increase and decrease. These parameters create a window for learning, during which data can be either retained or discarded from the network.

We assume that if a data sequence reaches the threshold weight for memory fixation, it remains in memory and is not segmented any further. An intuitive example from language learning is a word such as “backpack,” which would be fixed in memory if it were heard repeatedly without prior exposure to instances of “back” or “pack” (not even within other sequences, such as “on my back” or “in the pack”). If “back” or “pack” are heard often, then their partial commonality with “backpack” would result in “backpack” being segmented into “back” and “pack” (with a directed link between them; i.e., back→pack). The fixation of long sequences may have a positive or a negative impact on a learner’s success, as discussed below and in refs. 81 and 87. Note that the fixation of “backpack” does not prevent the formation of separate nodes for “back” and “pack” following later observations.

In addition to breaking up segments to form smaller segments, a node can be formed by the concatenation of smaller segments after they are repeatedly observed in succession. Thus, nodes can be formed “top-down” directly from the raw input by segmentation, or “bottom-up” through concatenation of previously learned units, creating a hierarchical structure, with potentially multiple hierarchies that can be perceived as “sequences of shorter sequences” (see refs. 85 and 86 for more details). In both cases, the effects that memory parameters have on learning amount to a test of statistical significance: natural and meaningful patterns are likely to recur and thus pass the test, whereas spurious patterns decay and are forgotten.

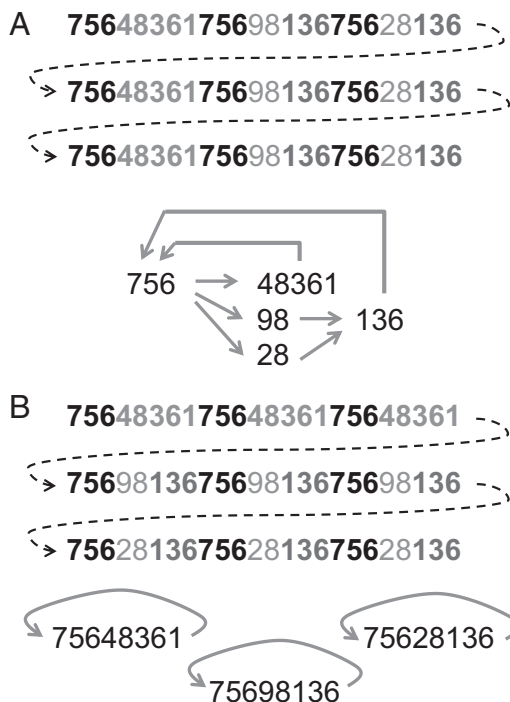


Fig. 1. Data input in the form of three strings, and the network that is constructed as a result of acquiring and processing this input using the learning mechanisms and parameter set described in the text. (A) Each data string of 24 characters is composed of three nonidentical subsequences of eight characters that share some common segments (highlighted using the same shade of gray). The three strings are identical in this case, so labeling each subsequence of eight characters as A, B, and C, respectively, allows describing the structure of the input as ABC ABC ABC. (B) The same input as in A is distributed differently over time, which can be described in short as AAA BBB CCC. This input leads to a completely different network structure due to fixation of A, B, and C as long eight-character chunks. The weights of the nodes and the links of the networks are not shown in the figure, but all of them exceed the fixation threshold of 1.0, as the weight-increase parameter was set to 0.4 per occurrence.

A Simplified Example. To better understand the process of data segmentation and network construction, a simplified example is illustrated in Fig. 1A. This example shows a network that is constructed as the result of acquiring three specific strings of data, under the assumption that the weight-increase parameter is 0.4, the fixation threshold is 1.0 (which means that a data item reaches fixation after three successive observations, because $3 \times 0.4 > 1$), and the weight-decrease (decay) parameter is 0.01 (i.e., the weight of a data item that is not yet fixated decreases by 0.01 with each symbol that enters the input). It is also assumed that the working-memory buffer can accommodate up to 24 symbols (which is the length of one data string in the example in Fig. 1A), and that the strings in this illustration are separated by 30 additional (irrelevant) characters that prevent parts of any two of these strings from being processed simultaneously in the memory buffer. The figure demonstrates that repeated sequences within each string (highlighted by shades of gray for clarity) are segmented based on their similarity and become the data units that form the nodes of the network. Directed links represent past association between these units; thus, they represent statistical regularities of the environment. For example, 98 always follows 756 and precedes 136, whereas 756 leads to 48361, 98, and 28 with equal probability. Despite the simplicity of this network, we can already observe that 98 and 28 have a similar link structure: both are preceded by 756 and followed by 136. In our earlier work (85, 86), we showed how such similarity in link structure can be used for generalization, for the construction of hierarchical representations, and for creativity (87).

We stressed earlier that, according to our model, the coordinated action of learning and data acquisition mechanisms and their evolution in response to typical input characteristics are critical for building an effective network. This point is illustrated by Fig. 1B, where the same data as in Fig. 1A are now distributed differently, leading to a radically different network representation (although the learning parameters and the working-memory buffer size remain the same). The distribution of the data input in Fig. 1B leads to the fixation of large idiosyncratic data sequences and to poor link structure, which may hamper further learning and generalization (81, 87). For example, no generalization can now be drawn for 98 and 28 because each of them is “locked” within another segment. Recognizing

segments in novel input becomes more difficult (i.e., is less likely) if the memory representation is based on large idiosyncratic units. The learner is then less likely to place novel data in context and to perform further segmentation.

Note that if we change only the weight-increase parameter from 0.4 to 0.3, then the data input of Fig. 1B would result in exactly the same network as in Fig. 1A and all of the problems that we have just described would disappear. This is because the segment “75648361” would not reach fixation after the first data string and would not decay completely before the second string is acquired, so it would be segmented when 756 is encountered in the second string and again in the third. A similar result can be achieved if we extend the working-memory buffer to include the beginning of the next string (so that the fourth occurrence of “756” can split the “75648361” and so on). This example shows how different combinations of data distribution and memory parameters can generate quite different or quite similar networks. It also demonstrates how relatively small modifications to the learning parameters or to the distribution of data input can lead to major changes in the network. As we discussed elsewhere (82, 86), it is important to bear in mind that not only may the learning parameters vary across individuals or species, but they may also evolve to differ across different sensory modalities (or different learning mechanisms) to better respond to the different distributions of data types in nature. Similarly, the learning parameters may also be modulated by physiological and emotional state, giving higher increase in memory weight to important but relatively rare observations (82).

The learning mechanism described so far is sensitive to the order in which elements appear in the data. For example, 756 and 576 are viewed as different data sequences. This may be important for some data types, such as the sequence of actions of a particular hunting technique, the phrases in a birdsong, or human speech. But for some other types of data it may be sufficient (or even better) to classify two sequences as the same by merely recognizing some of their similar components. For example, two instances of the same salad in a salad bar or a stand of mango trees in the forest may be recognized based on a combination of stimuli, ignoring their exact serial order (which may actually vary across instances). It is therefore possible that the learning mechanisms may also differ in the set of parameters that determine how sensitive they are to the exact serial order of data items. Clearly, a change in these parameters can also influence the segmentation process and the structure of the network, an issue that will become relevant again when we discuss language and tool-making, for which serial order is critically important.

Finally, as explained earlier, our model does not pretend to capture cognitive mechanisms at the neuronal level. The nodes and the links in our network do not correspond to neurons and synapses. Nevertheless, the processes described in our model at the computational level can be realized by neuronal structures and activities, and a representation of the proposed network may exist in the brain. We can assume that the neuronal structures and brain circuits that realize the network are ultimately affected by constraints of size and morphology that are at least partly determined genetically. That is, adaptive changes in the data acquisition and the learning mechanisms that can potentially lead to the construction of an extensive network in the acoustic domain, for example, may be subject to physical constraints that are also genetically determined. Over generations, genetic variants that are better in relaxing these physical constraints and in meeting the demand for larger or more appropriate neuronal structures will be favored by selection. This view of brain evolution is consistent with the “Baldwin effect” view (92, 93), according to which genes may be selected based on how well they support adaptive plastic processes, such as learning. Using this approach to address the question of how culture shapes the evolution of the brain would imply that culture exerts selective pressure that shapes learning and data-acquisition parameters, which in turn shape the structure of the constructed network. Consequently, over evolutionary time scales, brain anatomy may

be selected to better accommodate the physical requirements of the constructed network. In the next two sections we consider how this may have happened in the case of human language and stone-tool production, reexamining in this light the two problems discussed earlier regarding memory constraints and social-learning mechanisms.

The Case of Human Language and Memory Constraints

Although the question of how language has evolved is in itself the focus of extensive research (e.g., refs. 94–96), here we focus on a more specific question: Given that language has evolved, how can it shape the evolution of cognition? According to the process-level approach described above, we should address this question in terms of how the need to learn a language, or to use it, selects for possible changes in data-acquisition or learning mechanisms. The first expected change, which is quite obvious, is in the data-acquisition mechanisms: we would expect that attention to human speech, as well as to human gaze and gestures that can help to learn the meaning of spoken words, would become even more important than before. Indeed, these manifestations of social attention are very typical of human infants and young children (97, 98); impairments in such social attention skills are known to lead to problems in language learning, as in the case of autism (e.g., ref. 99). Perhaps the most significant expected impact of changes to the data distribution is on the segmentation process, and consequently on the construction of the network: because we expect the data-acquisition and learning parameters to coevolve, the evolution of language should also affect the memory parameters of the learning mechanisms. Note that this consideration takes us back to the problem of language and memory constraints discussed earlier in the paper. It is highly unlikely that the memory and learning parameters that evolved before language existed were best suited for processing linguistic data. Although certain plastic adjustment of these memory parameters on the basis of the learner’s individual experience cannot be ruled out, it is unlikely that adaptation to language learning and use over hundreds of generations did not also play a role in shaping the genetic basis of these parameters’ values. This claim is supported by the known genetic heritability component in various types of memory (60–63). The question to address next is how these parameters evolved as a result of language evolution.

Intuitively, one would expect the challenge of language acquisition to require and to select for a better working memory, leading to a view of a memory limit as a constraint rather than as an adaptation (see problem 2, above). However, according to our model, there are at least two reasons why a limited working memory may actually be adaptive. First, as explained earlier, the parameters of weight increase and decrease create a window for learning that serves as a test of statistical significance: natural and meaningful patterns are likely to recur and thus to pass the test, whereas spurious patterns decay and are forgotten. According to this view, the reason that it is typically difficult to learn a novel input from a single encounter is that the mechanism of learning has evolved to expect more evidence before deciding whether an item should be learned or ignored. The evolution of learning parameters that allow data items to reach fixation in memory after a single encounter should be possible. There are in fact examples for such “one-trial learning,” that, interestingly, occurs when rapid learning seems to be adaptive, as in the context of fear learning and enemy recognition (100, 101) or in the case of word “fast-mapping” in young children (102). However, in the case of large quantities of sequential data, where all items may be equally important, proximity of repeated occurrences in time and frequency of recurrence are the best first-resort tests of meaningfulness. The selective pressure of language in the direction of smaller buffer sizes and moderate fixation rates may explain why people are not better at memorizing sequential data verbatim (58), which would require larger buffer sizes and more rapid fixation.

Second, the limited memory buffer can be viewed as representing an adaptive trade-off between memory and computation. [Recall our earlier discussion of Christiansen and Chater's memory bottleneck (58).] We assume that a larger buffer that can accommodate more data can evolve, but whether or not it will be adaptive depends on the kind of computations needed to process the data in the buffer. In the case of linguistic input, serial order is important; the data must be segmented into words or chunks based on recurring segments. This means that the learner has to search and compare all possible chunks within the buffer, and to find possible matches between these chunks and those represented already as nodes in the network (so as to put the incoming data in context). The number of possible chunks in a linear sequence grows as the square of the length of the sequence [there are $N(N - 1)/2$ possible chunks in a linear sequence of N data items]. Thus, in a relatively small buffer of 5 data items, the learner has to find and compare only 10 possible chunks, whereas in larger buffers of, say 10 or 15 items, the learner has to find and compare 45 or 105 possible chunks, respectively. Therefore, increasing the buffer size leads to considerable computational cost that may not be justifiable. Depending on the number of items that comprise a typical meaningful chunk in the language, it might be better to process a sequence of data items by gradually scanning it with a small buffer of 5 items rather than with a large buffer of 15 items.

The computational burden is likely to be much smaller if data input does not need to be segmented accurately, but merely recognized and classified based on some characteristic features. As we mentioned earlier, recognizing a particular salad in a salad bar or a typical fruit tree in a forest may not require paying attention to the exact serial order of the data. In this case, a larger memory buffer may not lead to such a sharp increase in computation. A simple illustration of this phenomenon is given by the three nonsegmented sentences presented in Fig. 2. English speakers who cannot read Hebrew or Japanese would most certainly try to segment (almost automatically and subconsciously) the first sentence that is in English but not the next two sentences that are in Hebrew and Japanese. Those would be classified quickly as "Gibberish in a foreign language" or, more specifically, as "two sentences in Hebrew and Japanese that I can't read" (based on some distinctive features of Hebrew and Japanese letters). Clearly, readers who know Hebrew or Japanese will segment those sentences automatically. The point of this example is to demonstrate that the "ecological" context and the cultural background of a learner can determine the level of computation applied to processing incoming data.

Some level of data segmentation was clearly required even before the evolution of language; for example, when animals forage for food in structured environments (85) or need to learn or interpret observed behavioral sequences (103). However, the evolution of language almost certainly increased the proportion of data input that must be accurately segmented, thereby imposing more significant computational requirements for a given memory buffer size. Individuals with a genetic predisposition to use a smaller buffer may have been selected, which may explain why the memory bottleneck is indeed so small. This hypothesis predicts that some of the humans' close relatives that do not possess language may be endowed with a larger working-memory buffer. Indeed, a notable study on working memory for numerals in chimpanzees (104) shows that young chimpanzees have a better capacity for numerical recollection than human adults.

The number of possible chunks in a linear sequence = $N(N-1)/2$

המספר האפשרי שלצ'אנקים ברכיב ליניארי = $N(N-1)/2$

線形シーケンス内の可能なチャック数 = $N(N-1)/2$

Fig. 2. The sentence: "The number of possible chunks in a linear sequence = $N(N - 1)/2$ " written in a nonsegmented form in three different languages, English, Hebrew, and Japanese. See the explanation in the main text.

This ability may be useful for fast recognition of objects or structures in the field that does not depend on serial order. It may also improve rote learning of recent actions that might be helpful in systematically searching for food without returning to a place that was just visited. Interestingly, exceptional ability to retain in memory an accurate, detailed image of a complex scene or pattern (also known as "eidetic imagery") is more common among young children (105) and autistic savants (106). Such an ability may facilitate rote learning at the expense of effective segmentation and network representation (81). Regardless of whether or not we understand this phenomenon correctly, it clearly shows that having a larger working-memory buffer is biologically feasible and that genetic variants that possess this ability are already present in human populations. The fact that they do not spread and become the norm suggests that a small memory bottleneck is somehow more adaptive.

The Case of Tool-Making and the Evolution of Social-Learning Mechanisms

Cultural transmission of tool-making techniques depends on social-learning mechanisms. Whereas learning some advanced techniques may involve teaching and verbal instruction (107), the ability to make stone tools probably depended, initially at least, on social-learning mechanisms of the type needed to facilitate imitation or emulation (108–110). How, then, could the evolution of culturally transmitted techniques for making tools (i.e., the culture of "tool-making") affect the genetic evolution of such learning mechanisms? This question is a specific instance of the more general questions addressed earlier (*Problem 1*, above) of how using learning mechanisms for social functions shapes their evolution. To answer this question, we should first consider how imitation or emulation works. Here we try to explain it in terms of our model. We assume that for imitation, the coupling between perception and action is developed through experience (see also ref. 2). That is, when an individual repeatedly observes its own actions, it gradually—and quite automatically—associates the perception and the motor experience of those actions. Eventually, seeing another individual performing those actions activates the observer's representation of the motor experience of performing those actions [because the observed actions are perceived as being similar to the (perceptual representation of) the individual's own actions, which are already coupled with the relevant motor experience]. Thus, the first expected effect, which may be described in terms of data-acquisition mechanisms (33, 82), is an increase in attention to the behavioral patterns of other individuals (in imitation) or to the outcomes of their actions (in emulation). For imitation, it is also important to acquire much information on self-actions to create the coupling between the perception of these actions and their motor experience.

The next question is how to organize the acquired data in memory. In our framework, this amounts to asking how the network should be constructed. In the case of imitation, there seem to be two possibilities. One is to represent long sequences of observed actions or sensory experiences as large chunks: exact copies of entire sequences that can then be executed. The other possibility is to segment the observed behavior or sensory experience into smaller basic units, just as in the process of language acquisition, and then compose them again into larger sequences in the production process. The first possibility of exact imitation fits the notion of specialized imitation ability that allows copying and executing complex behaviors accurately and almost automatically. However, this approach leads to three problems. First, the expected effect of exact imitation on the evolution of learning is that both the "working-memory buffer" and the weight-increase parameter should increase in size. The working-memory buffer should be large enough to capture the long sequences of observed behaviors, and the weight-increase parameter should be high enough to allow rapid fixation in memory. This prediction does not seem to hold. As discussed earlier, the human working-memory buffer is typically small, and complex patterns require repeated encounters to be learned. The

second problem with exact imitation of novel sequences is whether or not it can work within the framework of the associative account. To create the coupling between perception and action, the learner must also produce the action, and this cannot be done before successful imitation (because the perception of a unique novel sequence does not yet have a match in motor representation that can be executed). Finally, the learning of long fixed-action sequences reduces flexibility. It would require, for example, that the initial raw stone from which a tool is to be produced be nearly identical to the raw stone that was used in the original learned sequence, a highly unlikely occurrence.

The alternative possibility, which involves sequence segmentation and reassembly, is consistent with the associative account and is also feasible. The learner first explores and practices a large repertoire of simple behavioral actions, thereby creating the necessary coupling between perception and action in a large repertoire of basic behavioral units. It can then concatenate these basic units in many possible ways; for the purpose of imitation, it can concatenate them to gradually match the complex behavior demonstrated by other individuals. The demonstrated behavior is also segmented into familiar units, which can then be associated with familiar actions, which helps in producing an imitation. This scenario is quite consistent with mounting evidence and recent views of experience-based imitation and emulation (110, 111). It also suggests that similar processes are involved both in language learning and in complex imitation, which is in line with recent views according to which tool-making possibly preadapted the brain to language learning (27).

Finally, our process-level approach may also help to explain recent new studies linking neuroanatomical changes in the brain to Paleolithic tool-making ability. These studies found that the acquisition of tool-making abilities by experimental subjects involved specific structural changes in the brain (27) and that these structures and regions in the brain are more developed in humans than in chimpanzees (28). This evidence for a short-term plastic response colocalized with structures that underwent recent evolutionary change strongly suggests a process akin to the Baldwin effect, in which genetic variants are selected based on how well they support the required plastic changes (92, 93). It is yet to be explained, however, how the observed plastic changes improve tool-making abilities. As we suggested earlier, in our view, such neuroanatomical changes have to do with the pathways and the representational systems that are recruited to serve the construction of the network. The result should be a rich network that represents sensory and perceptual experiences of

various hand movements, stone tools in various stages of completion, and the association of all these images and segments in time and space. According to our model, having a rich, well-segmented, and well-connected network helps to put new observations in context and to produce effective actions (81, 82, 87). Moreover, the ability to create a well-segmented and well-connected network likely depends on appropriate settings of the parameters of the data acquisition and the learning mechanisms that govern the dynamic process of network construction. Thus, the fine-tuning of these mechanisms by natural selection to produce the most effective network for the purpose of learning to make tools would be precisely the manner in which the culture of tool-making shapes the evolution of cognition.

Conclusions

In this paper we embraced the view that cognitive mechanisms have evolved to accommodate—among other tasks—the relatively new challenges of learning cultural constructs, such as language and tool-making techniques or, simply put, that culture shaped cognition. We claim, however, that to study how such cultural constructs shape cognitive evolution, a computationally explicit process-level mechanistic model of learning may be required. We described such a model, one that is based on coevolving mechanisms of learning and data acquisition that jointly construct a complex network, capable of supporting a range of cognitive abilities. The effect of culture on cognitive evolution is captured through small modifications of these coevolving learning and data-acquisition mechanisms, whose coordinated action improves the network's ability to support the learning processes that are involved in cultural phenomena, such as language or tool-making. Finally, we proposed that culture exerts selective pressure that shapes learning and data acquisition parameters, which in turn shape the structure of the representation network, so that over evolutionary time scales, brain anatomy may be selected to better accommodate the physical requirements of the learned processes and representations.

ACKNOWLEDGMENTS. We thank the organizers and funders of the Arthur M. Sackler Colloquium on "The Extension of Biology Through Culture." We also thank two anonymous reviewers for highly constructive comments. Funding was provided by the Israel Science Foundation Grant 871/15 (to A.L.) and by NSF Grant CCF-1214844, Air Force Office of Scientific Research Grant FA9550-12-1-0040, and Army Research Office Grant W911NF-14-1-0017 (to J.Y.H.). O.K. was supported by the John Templeton Foundation Grant ID 47981 and by the Stanford Center for Computational, Evolutionary, and Human Genomics.

- Heyes C, Pearce JM (2015) Not-so-social learning strategies. *Proc Biol Sci* 282:20141709.
- Cook R, Bird G, Catmur C, Press C, Heyes C (2014) Mirror neurons: From origin to function. *Behav Brain Sci* 37:177–192.
- Leadbeater E (2015) What evolves in the evolution of social learning? *J Zool (Lond)* 295:4–11.
- Iacoboni M, et al. (1999) Cortical mechanisms of human imitation. *Science* 286:2526–2528.
- Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2:661–670.
- Leslie AM (1987) Pretense and representation: The origins of "theory of mind." *Psychol Rev* 94:412–426.
- Baron-Cohen S (2000) Theory of mind and autism: A fifteen year review. *Understanding Other Minds*, eds Baron-Cohen S, Tagar-Flusberg H, Cohen DJ (Oxford Univ Press, Oxford), Vol A, pp 3–20.
- Cosmides L, Tooby J, Fiddick L, Bryant GA (2005) Detecting cheaters. *Trends Cogn Sci* 9:505–506, author reply 508–510.
- Chomsky N (1965) *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA).
- Tooby J, Cosmides L, Barrett HC (2005) Resolving the debate on innate ideas. *The Innate Mind: Structure and Content*, eds Carruthers P, Laurence S, Stich S (Oxford Univ Press, New York), pp 305–337.
- Anderson ML (2010) Neural reuse: A fundamental organizational principle of the brain. *Behav Brain Sci* 33:245–266, discussion 266–313.
- Bates E (1993) Modularity, domain specificity and the development of language. *Discuss Neurosci* 10:136–148.
- Scott-Phillips TC, Laland KN, Shuker DM, Dickens TE, West SA (2014) The niche construction perspective: A critical appraisal. *Evolution* 68:1231–1243.
- Laland KN, Odling-Smee FJ, Feldman MW (1999) Evolutionary consequences of niche construction and their implications for ecology. *Proc Natl Acad Sci USA* 96:10242–10247.
- Odling-Smee FJ, Laland KN, Feldman MW (2003) *Niche Construction: The Neglected Process in Evolution* (Princeton Univ Press, Princeton, NJ).
- Iriki A, Taoka M (2012) Triadic (ecological, neural, cognitive) niche construction: A scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. *Philos Trans R Soc Lond B Biol Sci* 367:10–23.
- Laland KN, Hoppitt W (2003) Do animals have culture? *Evol Anthropol* 12:150–159.
- Laland KN, Janik VM (2006) The animal cultures debate. *Trends Ecol Evol* 21:542–547.
- Stout D, Hecht EE (2017) Evolutionary neuroscience of cumulative culture. *Proc Natl Acad Sci USA* 114:7861–7868.
- Solan Z (2005) Unsupervised learning of natural languages. *Proc Natl Acad Sci USA* 102:11629–11634.
- Laland KN, Odling-Smee J, Myles S (2010) How culture shaped the human genome: Bringing genetics and the human sciences together. *Nat Rev Genet* 11:137–148.
- Williamson SH, et al. (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genet* 3:e90.
- de Magalhães JP, Matsuda A (2012) Genome-wide patterns of genetic distances reveal candidate loci contributing to human population-specific traits. *Ann Hum Genet* 76:142–158.
- Somel M, Liu X, Khaitovich P (2013) Human brain evolution: Transcripts, metabolites and their regulators. *Nat Rev Neurosci* 14:112–127.
- Cáceres M, et al. (2003) Elevated gene expression levels distinguish human from non-human primate brains. *Proc Natl Acad Sci USA* 100:13030–13035.
- Somel M, Rohlf R, Liu X (2014) Transcriptomic insights into human brain evolution: Acceleration, neutrality, heterochrony. *Curr Opin Genet Dev* 29:110–119.
- Hecht EE, et al. (2015) Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. *Brain Struct Funct* 220:2315–2331.

28. Hecht EE, Gutman DA, Bradley BA, Preuss TM, Stout D (2015) Virtual dissection and comparative connectivity of the superior longitudinal fasciculus in chimpanzees and humans. *Neuroimage* 108:124–137.
29. Creanza N, Fogarty L, Feldman MW (2016) Cultural niche construction of repertoire size and learning strategies in songbirds. *Evol Ecol* 30:285–305.
30. Shettleworth SJ (2010) *Cognition, Evolution, and Behavior* (Oxford Univ Press, New York).
31. Heyes C (2010) Where do mirror neurons come from? *Neurosci Biobehav Rev* 34:575–583.
32. Heyes C (2016) *Homo imitans?* Seven reasons why imitation couldn't possibly be associative. *Philos Trans R Soc Lond B Biol Sci* 371:20150069.
33. Heyes C (2012) What's social about social learning? *J Comp Psychol* 126:193–202.
34. Laland KN (2004) Social learning strategies. *Learn Behav* 32:4–14.
35. Catmur C, Press C, Cook R, Bird G, Heyes C (2014) Authors' response: Mirror neurons: Tests and testability. *Behav Brain Sci* 37:221–241.
36. Felsenstein J (1983) Parsimony in systematics: Biological and statistical issues. *Annu Rev Ecol Syst* 14:313–333.
37. Lotem A (1993) Secondary sexual ornaments as signals: The handicap approach and three potential problems. *Etologia* 3:209–218.
38. Mery F, Belay AT, So AK, Sokolowski MB, Kawecki TJ (2007) Natural polymorphism affecting learning and memory in *Drosophila*. *Proc Natl Acad Sci USA* 104:13051–13055.
39. Dunlap AS, Stephens DW (2014) Experimental evolution of prepared learning. *Proc Natl Acad Sci USA* 111:11750–11755.
40. Garcia J, Kimeldorf DJ, Koelling RA (1955) Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science* 122:157–158.
41. Finkel D, Pedersen NL, McGue M, McClearn GE (1995) Heritability of cognitive abilities in adult twins: comparison of Minnesota and Swedish data. *Behav Genet* 25:421–431.
42. Plomin R, Spinath FM (2002) Genetics and general cognitive ability (g). *Trends Cogn Sci* 6:169–176.
43. Briley DA, Tucker-Drob EM (2013) Explaining the increasing heritability of cognitive ability across development: A meta-analysis of longitudinal twin and adoption studies. *Psychol Sci* 24:1704–1713.
44. Pearson-Fuhrhop KM, Minton B, Acevedo D, Shahbaba B, Cramer SC (2013) Genetic variation in the human brain dopamine system influences motor learning and its modulation by L-Dopa. *PLoS One* 8:e61197.
45. Mery F (2013) Natural variation in learning and memory. *Curr Opin Neurobiol* 23:52–56.
46. Stephens DW, Krebs JR (1986) *Foraging Theory* (Princeton Univ Press, Princeton, NJ).
47. Leadbeater E, Dawson EH (2017) A social insect perspective on the evolution of social learning mechanisms. *Proc Natl Acad Sci USA* 114:7838–7845.
48. Lotem A, Kolodny O (2014) Reconciling genetic evolution and the associative learning account of mirror neurons through data-acquisition mechanisms. *Behav Brain Sci* 37:210–211.
49. Thompson B, Kirby S, Smith K (2016) Culture shapes the evolution of cognition. *Proc Natl Acad Sci USA* 113:4530–4535.
50. Christiansen MH, Chater N (2008) Language as shaped by the brain. *Behav Brain Sci* 31:489–508, discussion 509–558.
51. Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci USA* 87:1816–1819.
52. Laland KN (2017) The origins of language in teaching. *Psychon Bull Rev* 24:225–231.
53. Belfer-Cohen A, Goren-Inbar N (1994) Cognition and communication in the Levantine Lower Palaeolithic. *World Archaeol* 26:144–157.
54. d'Errico F, et al. (2003) Archaeological evidence for the emergence of language, symbolism, and music—An alternative multidisciplinary perspective. *J World Prehist* 17:1–70.
55. Mellars P (2006) Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci USA* 103:9381–9386.
56. Chater N, Reali F, Christiansen MH (2009) Restrictions on biological adaptation in language evolution. *Proc Natl Acad Sci USA* 106:1015–1020.
57. Chater N, Christiansen MH (2010) Language acquisition meets language evolution. *Cogn Sci* 34:1131–1157.
58. Christiansen MH, Chater N (2016) The Now-or-Never bottleneck: A fundamental constraint on language. *Behav Brain Sci* 39:e62.
59. Lotem A, Kolodny O, Halpern JY, Onniss L, Edelman S (2016) The bottleneck may be the solution, not the problem. *Behav Brain Sci* 39:e83.
60. Mueller ST, Krawitz A (2009) Reconsidering the two-second decay hypothesis in verbal working memory. *J Math Psychol* 53:14–25.
61. Cui J, Gao D, Chen Y, Zou X, Wang Y (2010) Working memory in early-school-age children with Asperger's syndrome. *J Autism Dev Disord* 40:958–967.
62. Blokland GAM, et al. (2011) Heritability of working memory brain activation. *J Neurosci* 31:10882–10890.
63. Vogler C, et al. (2014) Substantial SNP-based heritability estimates for working memory performance. *Transl Psychiatry* 4:e438.
64. Chater N, Christiansen MH (2016) Squeezing through the Now-or-Never bottleneck: Reconnecting language processing, acquisition, change, and structure. *Behav Brain Sci* 39:e91.
65. McNamara JM, Houston AI (2009) Integrating function and mechanism. *Trends Ecol Evol* 24:670–675.
66. Fawcett TW, Hamblin S, Giraldeau LA (2013) Exposing the behavioral gambit: The evolution of learning and decision rules. *Behav Ecol* 24:2–11.
67. Kacelnik A, Bateson M (1997) Risk-sensitivity: Crossroads for theories of decision-making. *Trends Cogn Sci* 1:304–309.
68. van den Berg P, Weissing FJ (2015) The importance of mechanisms for the evolution of cooperation. *Proc Biol Sci* 282:20151382.
69. Trimmer PC, McNamara JM, Houston AI, Marshall JA (2012) Does natural selection favour the Rescorla-Wagner rule? *J Theor Biol* 302:39–52.
70. McNamara JM, Trimmer PC, Houston AI (2012) The ecological rationality of state-dependent valuation. *Psychol Rev* 119:114–119.
71. Arbilly M, Motro U, Feldman MW, Lotem A (2010) Co-evolution of learning complexity and social foraging strategies. *J Theor Biol* 267:573–581.
72. Lange A, Dukas R (2009) Bayesian approximations and extensions: Optimal decisions for small brains and possibly big ones too. *J Theor Biol* 259:503–516.
73. Katsnelson E, Motro U, Feldman MW, Lotem A (2011) Evolution of learned strategy choice in a frequency-dependent game. *Proc Biol Sci* 279:1176–1184.
74. Hamblin S, Giraldeau L-A (2009) Finding the evolutionarily stable learning rule for frequency-dependent foraging. *Anim Behav* 78:1343–1350.
75. Nilsson DE, Pelger S (1994) A pessimistic estimate of the time required for an eye to evolve. *Proc Biol Sci* 256:53–58.
76. Schmidhuber J (2015) Deep learning in neural networks: An overview. *Neural Netw* 61:85–117.
77. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444.
78. Edelman S (2016) The minority report: Some common assumptions to reconsider in the modelling of the brain and behaviour. *J Exp Theor Artif Intell* 28:751–776.
79. Mnih V, et al. (2015) Human-level control through deep reinforcement learning. *Nature* 518:529–533.
80. Gu S, Lillicrap T, Sutskever I, Levine S (2016) Continuous deep Q-learning with model-based acceleration. *Proceedings of The 33rd International Conference on Machine Learning*, eds Balcan MF, Weinberger KQ (Proceedings of Machine Learning Research, New York), Vol 48, pp 2829–2838.
81. Lotem A, Halpern JY (2008) A data-acquisition model for learning and cognitive development and its implications for autism. *Cornell University Computing and Information Science Technical Reports*. Available at <https://ecommons.cornell.edu/handle/1813/10178>. Accessed May 11, 2017.
82. Lotem A, Halpern JY (2012) Coevolution of learning and data-acquisition mechanisms: A model for cognitive evolution. *Philos Trans R Soc Lond B Biol Sci* 367:2686–2694.
83. Goldstein MH, et al. (2010) General cognitive principles for learning structure in time and space. *Trends Cogn Sci* 14:249–258.
84. Kolodny O, Edelman S, Lotem A (2014) The evolution of continuous learning of the structure of the environment. *J R Soc Interface* 11:20131091.
85. Kolodny O, Edelman S, Lotem A (2015) Evolution of protolinguistic abilities as a by-product of learning to forage in structured environments. *Proc Biol Sci* 282:20150353.
86. Kolodny O, Lotem A, Edelman S (2015) Learning a generative probabilistic grammar of experience: A process-level model of language acquisition. *Cogn Sci* 39:227–267.
87. Kolodny O, Edelman S, Lotem A (2015) Evolved to adapt: A computational approach to animal innovation and creativity. *Curr Zool* 61:350–367.
88. Enquist M, Lind J, Ghirlanda S (2016) The power of associative learning and the ontogeny of optimal behaviour. *R Soc Open Sci* 3:160734.
89. Laland KN, et al. (2015) The extended evolutionary synthesis: Its structure, assumptions and predictions. *Proc Biol Sci* 282:20151019.
90. Baddeley A, Gathercole S, Papagno C (1998) The phonological loop as a language learning device. *Psychol Rev* 105:158–173.
91. Burgess N, Hitch GJ (1999) Memory for serial order: A network model of the phonological loop and its timing. *Psychol Rev* 106:551–581.
92. Baldwin JM (1896) A new factor in evolution. *Am Nat* 30:441–451.
93. Weber BH, Depew DJ (2003) *Evolution and Learning: The Baldwin Effect Reconsidered* (MIT Press, Cambridge, MA).
94. Dunbar R (1998) *Grooming, Gossip, and the Evolution of Language* (Harvard Univ Press, Cambridge, MA).
95. Pinker S (2003) Language as an adaptation to the cognitive niche. *Stud Evol Lang* 3:16–37.
96. Premack D (1985) "Gavagai" or the future history of the animal language controversy. *Cognition* 19:207–296.
97. Butterworth G, Jarrett N (1991) What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *Br J Dev Psychol* 9:55–72.
98. Scaife M, Bruner JS (1975) The capacity for joint visual attention in the infant. *Nature* 253:265–266.
99. Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W (2009) Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* 459:257–261.
100. Curio E, Ernst U, Vieth W (1978) Cultural transmission of enemy recognition: One function of mobbing. *Science* 202:899–901.
101. Dunsmoor JE, Murty VP, Davachi L, Phelps EA (2015) Emotional learning selectively and retroactively strengthens memories for related events. *Nature* 520:345–348.
102. Medina TN, Snedeker J, Trueswell JC, Gleitman LR (2011) How words can and cannot be learned by observation. *Proc Natl Acad Sci USA* 108:9014–9019.
103. Byrne RW (1999) Imitation without intentionality. Using string parsing to copy the organization of behaviour. *Anim Cogn* 2:63–72.
104. Inoue S, Matsuzawa T (2007) Working memory of numerals in chimpanzees. *Curr Biol* 17:R1004–R1005.
105. Conway ARA, Jarrold C, Kane MJ, Miyake A, Towse JN (2007) Variation in working memory: An introduction. *Variation in Working Memory*, eds Conway ARA, Jarrold C, Kane MJ, Miyake A, Towse JN (Oxford Univ Press, Oxford), pp 3–17.
106. Snyder AW, Mitchell DJ (1999) Is integer arithmetic fundamental to mental processing?: The mind's secret arithmetic. *Proc Biol Sci* 266:587–592.
107. Morgan TJH, et al. (2015) Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat Commun* 6:6029.
108. Whiten A (2000) Primate culture and social learning. *Cogn Sci* 24:477–508.
109. Heyes CM, Galef BG, Jr (1996) *Social Learning in Animals: The Roots of Culture* (Academic, San Diego).
110. Galef BG (2015) Laboratory studies of imitation/field studies of tradition: Towards a synthesis in animal social learning. *Behav Processes* 112:114–119.
111. Truskanov N, Lotem A (2017) Trial-and-error copying of demonstrated actions reveals how fledglings learn to 'imitate' their mothers. *Proc Biol Sci* 284:20162744.