Evolutionary neuroscience of cumulative culture

Dietrich Stout a,b and Erin E. Hecht b,c

a Department of Anthropology, Emory University, Atlanta, GA 30322; b Center for Behavioral Neuroscience, Georgia State University, Atlanta, GA 30303; and c Yerkes National Primate Research Center, Emory University, Atlanta, GA 30320-5090

Edited by Marcus W. Feldman, Stanford University, Stanford, CA, and approved May 1, 2017 (received for review January 12, 2017)

Culture suffuses all aspects of human life. It shapes our minds and bodies and has provided a cumulative inheritance of knowledge, skills, institutions, and artifacts that allows us to truly stand on the shoulders of giants. No other species approaches the extent, diversity, and complexity of human culture, but we remain unsure how this came to be. The very uniqueness of human culture is both a puzzle and a problem. It is puzzling as to why more species have not adopted this manifestly beneficial strategy and problematic because the comparative methods of evolutionary biology are ill suited to explain unique events. Here, we develop a more particularistic and mechanistic evolutionary neuroscience approach to cumulative culture, taking into account experimental, developmental, comparative, and archaeological evidence. This approach reconciles currently competing accounts of the origins of human culture and develops the concept of a uniquely human technological niche rooted in a shared primate heritage of visuo-motor coordination and dexterous manipulation.

M odern humans live in a culturally constructed niche of artificial landscapes, structures, artifacts, skills, practices, and beliefs accumulated over generations and beyond the ability of any one individual to recreate in a lifetime (1, 2). Like the air we breathe, this cumulative cultural matrix is so immersive that it is easy to forget it is there. However, this is the medium through which we grow, act, and think, and it exerts profound influences on human life across a range of behavioral (2), developmental (3), and evolutionary (4) scales. How did our species find itself in this remarkable situation?

Niche construction is not unique to humans (5), and many animals reliably transmit behavioral traditions across generations (1). In contrast, it is controversial whether any examples of nonhuman cumulative culture exist and all can agree that no other species approaches the extent, diversity, and complexity of human culture (6, 7). Two kinds of explanations have been proposed for this disjunction. “Individual cognition” accounts (IC) propose that humans accumulate more complex cultures primarily because the biological evolution of greater intelligence in human individuals has promoted innovation and allowed mastery of more complex concepts and skills (7, 8). Alternatively, “cultural evolution” accounts (CE) propose that the difference arises from uniquely human psychological specializations for “high-fidelity” social learning [e.g., theory of mind (ToM), imitation], which have enabled the lossless “ratchet-effect” of cultural accumulation to supplant biology as humanity’s primary mode of adaptation (2, 9). Both views recognize a role for social learning in reducing the costs of knowledge and skill acquisition, but they differ on the phylogenetic uniqueness (e.g., ref. 7 vs. ref. 9) and transformative power (e.g., ref. 2 vs. ref. 8) of human high-fidelity social transmission. This plays out in starkly different visions of cumulative culture as either a fundamental evolutionary transition (cf. ref. 10) that altered the very medium of human adaptation (2), or just another “unique or extreme” biological trait comparable to “the elephant’s trunk, the narwhal’s tusk, the whale’s baleen, the platypus’s duckbill, and the armadillo’s armor” (8). Neither option, however, makes it immediately clear why this exceptional capacity for culture has evolved in humans and humans alone.

If there is one thing on which IC and CE agree, it is that cultural capacity is a good thing: It has “undeniable practical advantages” (8) that have allowed our species to have “expanded across the globe and... occupy a wider range than any other terrestrial species” (2). Indeed, the benefits are so substantial that even small “initial increments” (8) in this direction are expected to generate powerful biocultural feedback leading to further brain and cognitive evolution (2, 8). Proponents of a highly modular view of IC nevertheless argue that this feedback will lead to coordinated enhancement across multiple domains (8). CE advocates similarly suggest that evolved cognitive mechanisms (i.e., modules in a loose sense) for social learning will lead to more general brain size and intelligence increases to deal with increased amounts of “valuable cultural information” (2). So why are humans the only species to have fallen into this virtuous cycle?

Arguing from a CE perspective, Boyd and Richerson (11) suggest that cumulative culture is rare because of the evolutionary costs of requisite social-learning mechanisms. According to this argument, accumulation must begin with simple skills that are within the inventive potential of individuals. Insofar as such simple skills could be socially learned using existing “low-fidelity” mechanisms, any small increases in learning efficiency provided by new high-fidelity social-learning mechanisms would be unlikely to pay for the (presumably high) metabolic and developmental costs of those mechanisms. In this case, it would only be after accumulation had already generated a sufficient body of complex, difficult-to-learn, and useful cultural content that these expensive mechanisms would begin to pay off for themselves. Boyd and Richerson (11) thus suggest that high-fidelity social-learning capacities initially arose as a side effect (exaptation) of some other adaptation, such as behavior prediction for Machiavellian social strategizing (12, 13). Although reasonable, this hypothesis is weakened by its reliance on assumptions regarding the cost of social-learning mechanisms and the rarity of social-cognitive preadaptations. In fact, the hypothesis does not so much explain the rarity of cumulative culture as shift the puzzle to explaining the rarity of preadaptations, like behavior prediction, which might also be assumed to be quite generally beneficial.

An alternative, IC-compatible proposal is that it is large brains in general that are expensive, rather than social-learning capacities in particular. Thus, Pinker (8) argues that human cognitive exceptionalism arises from the fortuitous fact that “hominin ancestors, more so than any other species, had a collection of traits that had tilted the payoffs toward further investment in intelligence.” Whether such “intelligence” is thought to be composed of discrete but tightly coevolving innate modules (8) or a general-purpose
capacity only “secondarily” specialized by experience (14), the premise is that bigger brains are generally advantageous but can only evolve under a narrow set of conditions. Lifespans must be long enough (i.e., mortality low enough) to reward early investments in growth and additional energetic costs must be accommodated through increased intake or reallocation (4). For many species, constraints such as small body size, unstable environments, and unavoidable mortality from predation and disease may simply make large brains impractical. In this framework, high-fidelity social learning can increase the payoffs of large brains (8, 15) but is not associated with its own unique costs, as it is thought to rely on cognitive abilities that overlap (15) or evolved in tandem with (8) social learning and problem solving. This approach actually parallels the suggestion of Boyd and Richerson (11) that initial enhancements of social-learning capacities were a byproduct of selection on other capacities, and similarly shifts the question back to identifying the unique trait (4) or combination of traits (8) that pushed humans, and no others, into an auto-catalytic coevolutionary feedback loop.

The strength of both IC and CE accounts is that they explain a wide range of distinct human traits in terms of a single co-evolutionary process. However, this parsimony is offset by the perceived need to posit a uniquely exceptional [and possibly even random (16)] initial cause inaccessible to more general evolutionary explanation (8, 17, 18). An alternative is to consider cumulative culture, not as a unitary capacity that is either present or absent, but as a complex trait with a correspondingly complex history of gradual or piecemeal emergence. Although any event that altered the evolutionary cost/benefit analysis of either brain expansion generally (4, 8) or high-fidelity social learning specifically (2, 16) could theoretically have initiated runaway biocultural coevolution in our lineage, there is no reason to assume this feedback would be indefinitely self-sustaining once initiated, nor that it would necessarily produce constant increase as opposed to more complex dynamics. In fact, both comparative biological evidence (4) and cultural evolutionary models (19) indicate the potential for just such interactions and dynamics and this is entirely consistent with the emerging paleoanthropological picture of multilinear, intermittent, asynchronous change over human evolution (20, 21). This indication suggests a more contingent evolutionary history, likely involving multiple inflection points, in contrast to a hand-held hammer (typically highly specialized) to initiate controlled and predictable fracture. This means that small errors in strike execution can have catastrophic, unreversible effects. Experiments by Bril and colleagues have shown that fracture prediction and control is a demanding perceptual-motor skill reliably expressed only in expert knappers (28, 29). Building on this work, Stout and colleagues (31, 36, 37) found that even 22 mo (β = 167 h) of snapping training produced relatively little evidence of perceptual-motor improvement, in contrast to clear gains in conceptual understanding.

Knapping is a “reductive” technology involving the sequential detachment of flakes from a stone core using precise ballistic strikes with a handheld hammer (typically highly specialized) to initiate controlled and predictable fracture. This means that small errors in strike execution can have catastrophic, unreversible effects. Experiments by Bril and colleagues have shown that fracture prediction and control is a demanding perceptual-motor skill reliably expressed only in expert knappers (28, 29). Building on this work, Stout and colleagues (31, 36, 37) found that even 22 mo (β = 167 h) of snapping training produced relatively little evidence of perceptual-motor improvement, in contrast to clear gains in conceptual understanding.

We suggest that, as with technology involving the sequential detachment of flakes from a stone core using precise ballistic strikes with a handheld hammer (typically highly specialized) to initiate controlled and predictable fracture. This means that small errors in strike execution can have catastrophic, unreversible effects. Experiments by Bril and colleagues have shown that fracture prediction and control is a demanding perceptual-motor skill reliably expressed only in expert knappers (28, 29). Building on this work, Stout and colleagues (31, 36, 37) found that even 22 mo (β = 167 h) of snapping training produced relatively little evidence of perceptual-motor improvement, in contrast to clear gains in conceptual understanding.

We suggest that, as with technology involving the sequential detachment of flakes from a stone core using precise ballistic strikes with a handheld hammer (typically highly specialized) to initiate controlled and predictable fracture. This means that small errors in strike execution can have catastrophic, unreversible effects. Experiments by Bril and colleagues have shown that fracture prediction and control is a demanding perceptual-motor skill reliably expressed only in expert knappers (28, 29). Building on this work, Stout and colleagues (31, 36, 37) found that even 22 mo (β = 167 h) of snapping training produced relatively little evidence of perceptual-motor improvement, in contrast to clear gains in conceptual understanding.

We suggest that, as with technology involving the sequential detachment of flakes from a stone core using precise ballistic strikes with a handheld hammer (typically highly specialized) to initiate controlled and predictable fracture. This means that small errors in strike execution can have catastrophic, unreversible effects. Experiments by Bril and colleagues have shown that fracture prediction and control is a demanding perceptual-motor skill reliably expressed only in expert knappers (28, 29). Building on this work, Stout and colleagues (31, 36, 37) found that even 22 mo (β = 167 h) of snapping training produced relatively little evidence of perceptual-motor improvement, in contrast to clear gains in conceptual understanding.
copied by breaking them down into familiar action elements (e.g., lift, turn, twist), matching these, and reassembling (44). Such matching always requires a degree of approximation, if only because no two bodies are identical, but where low-level differences are not critical it can enable fast imitation learning (13). Unfortunately, such low-level details do matter for knapping. Learners must thus proceed by incorrectly executing an observed gesture or communicating strategy, checking the actual vs. predicted outcome, and embarking on a lengthy process of behavioral exploration to (re)discover relevant task constraints and develop corresponding internal models through a self-conscious process of deliberate practice (cf. ref. 45). These learning challenges call for an iterative approach that alternates social-learning opportunities (observation, instruction) with motivated individual practice (46), as commonly seen in coaching or apprenticeship learning. Whiten (47) describes this as a “helical curriculum.”

This complexity is not easily reconciled with the dichotomy of high-fidelity process copying (imitation) vs. low-fidelity product copying (emulation) prevalent in discussions of cumulative culture (e.g., ref. 48). For knapping and similar crafts, what is actually reproduced across individuals is a flexible skill rather than an invariant formula, and what must be learned is a structured set of relations [compare with, for example, “affordances” (49)] between effective means and appropriate goals at multiple levels of task organization (50). This is achieved through a diverse helical curriculum in which many varieties of social learning may scaffold individual development of perceptual-motor skill and strategic competence (Fig. 1). Whiten and colleagues (e.g., see ref. 12) have identified a “portfolio” of such processes, ranging from copying bodily postures and gestures to object movement reenactment and end-state emulation. In humans, these may even extend to copying cultural norms of affect, behavior, and social affiliation (30). Although such copying may be a goal in itself when learning conventional behaviors, like human dance or the “do-as-I-do” experimental paradigm used with apes (12), in instrumental skill acquisition these processes are means to the end of learning subtle affordances and complex causal relations (51). Such learning can be further facilitated by (passive or active) niche construction, increasing exposure to relevant materials and situations (30, 52) and by teaching, potentially including the provision of practice opportunities, the direction of attention, and affective feedback, as well as intentional demonstration and explicit instruction (53).

Underlying this diversity of processes, are a small number of fundamental psychological capacities required for efficient high-fidelity skill learning in a helical curriculum. These include: (i) the ability to relate fine details of actions and objects to complex goals during behavior observation and execution (46) and (ii) the prosocial motivation and ToM needed to support apprenticeship learning (9, 54, 55). In keeping with comparative (12) and archaeological (50) evidence, variation in these capacities is expected to be continuous rather than binary, and to be reflected in the complexity and difficulty of skills that can be accurately reproduced. The latter point suggests that the archaeological record of increasingly complex and difficult stone tool-making (31, 36, 50) could help adjudicate between (or reconcile) IC and CE accounts of human evolution, but only in the context of a solid inferential framework linking brains, behaviors, and evolutionary processes. Key questions include the extent and nature of overlap between processes supporting behavior execution, observation, and interpretation (e.g., ToM), and the relevance of evolutionary processes other than natural selection (e.g., CE). An emerging extended evolutionary synthesis (EES) effectively addresses both topics through its core concepts of constructive development and reciprocal causation (56).

**An Extended Evolutionary Framework**

As Deacon (57) provocatively put it, “brain evolution should be impossible.” How could random mutational changes to such a complex integrated system be anything other than catastrophic? The answer is that brain development is itself an evolutionary process of remarkable flexibility and adaptability. Basic patterning mechanisms and developmental selection (58) can produce functional systems even in the face of quite significant environmental or genetic perturbation. Such plasticity may be essential to the evolvability of larger brains, as developmental programs and network architectures are forced to accommodate shifting geometries together with massive increases in the number of potential neuronal connections to be specified. It also creates a medium for the multilevel interactions between genes, developmental mechanisms, environments both internal and external, and organisational behavior that the EES describes as “constructive development.”

![Fig. 1. A learning cycle in the helical curriculum. Social resources both passive and pedagogical (53), together with constructed learning contexts (52), provide opportunities and structure for individual practice, which can include a portfolio (12) of processes ranging from “emulative” end-state copying to the imitation of specific body movements. This behavioral continuum maps roughly onto the differing contributions of dorsal and ventral processing streams in the primate brain.](image-url)
Any viable evolutionary account of cumulative culture must address these dynamics. The primate neocortex is divided into large-scale functional networks characterized by high within-network functional and anatomical connectivity (59, 60). In humans, these networks are organized in a processing hierarchy from concrete perceptual and motor functions to abstract, domain-general processing. This arrangement is realized anatomically as a cortex-wide gradient of topological distance and connectivity patterns extending from distinct peripheral sensorimotor cortices at one end to highly interconnected central association cortices at the other (61). Along this gradient, seven widely recognized networks can be arrayed into four organizational tiers: (i) visual and somatomotor, (ii) dorsal and ventral attention, (iii) multiple demand and limbic, and (iv) default mode (61).

The tethering hypothesis of Buckner and Krienen (62) proposes that this pattern arises from disproportionate expansion of the cortical mantle during evolutionary brain enlargement, leading to gaps between the chemical signaling gradients that pattern cortical differentiation during development. Developmental selection in these gaps fosters the emergence of “noncanonical” association networks primarily interconnected with each other rather than with more developmentally constrained peripheral sensorimotor systems. As expected, these relatively unconstrained association cortices are also relatively late developing (59, 63) and variable in connectivity across individuals (64). Indeed, comparative evidence indicates human-specific changes in the rate and timing of synaptogenesis, synapse elimination, and cortical myelination, resulting in increased plasticity into adulthood (65, 66). That nonspecific selection for increased brain size in the human lineage might have indirectly driven increased plasticity is suggested by evidence of low heritability for cortical morphology (sulcal dimensions) vs. overall brain size in humans, a pattern that contrasts with high heritability of both in chimpanzees (67). In any case, the human association cortex appears particularly sensitive to environmental and behavioral influences, providing a potent evolutionary feedback mechanism between organism and environment, which the EES refers to as reciprocal causation (56).

Such phenotypic flexibility is useful but may come at a cost (e.g., investments in learning or temporary phenotype–environment mismatches). Where possible, natural selection is expected to redress these costs by realizing “plastic” responses to recurring environmental situations as automatic parts of normal development (68). The tethering hypothesis suggests that such innate specializations are most likely to be found in relatively heritable sensorimotor systems and with respect to behaviors/stimuli that have been relatively invariant over long periods of time. Because humans’ expanded association areas remain relatively plastic (64) and are both late developing (59) and phylogenetically recent (60), their derived cognitive features are less likely to be directly shaped by natural selection (phylogenetically constructed) and more likely to result from developmental side effects (developmental construction) and modifications to the structure of inputs they receive from more peripheral systems (phylogenetic or developmental inflection) (69). In theory, such modifications could arise through environmental as well as genetic inheritance, including persistent changes to the physical and social context of development brought about through niche construction (52, 70). For example, there is widespread agreement that the human brain lacks specific genetic adaptations for literacy, and yet learning to read reliably produces functional specialization for script perception in a particular region of the left ventral occipitotemporal cortex known as the “visual word form area” (3). Similar logic may apply to the enhanced mechanisms for complex action parsing and ToM that support apprenticeship learning in a helical curriculum.

Skilled actions, such as the ballistic strikes involved in stone knapping, often unfold too quickly to be guided by online sensory feedback and error correction. This limitation can be overcome through the use of internal models that predict movements and outcomes in advance (71), a simulation process supported by a distributed network of frontal, parietal, and occipitotemporal regions combining elements of the dorsal and ventral attention networks. As argued above, such models of self-action control likely form the basis for understanding and copying the observed actions of others through a process of matching, often referred to as “motor resonance” (42). The assembly of complex goal-oriented sequences from these elements is likely supported in the next tier of cortical organization by the multiple demand system (aka “frontoparietal control network”). This cognitive control network is thought to support general or “fluid” intelligence through its role in assembling structured mental programs from a series of subtasks (72), a critical process in skill learning as reviewed above. Together, these sensorimotor matching and control processes support the interactive behavioral alignment that is critical to human social learning, communication, cooperation, and bonding (73, 74).

It is debatable whether the application of increasingly sophisticated motor planning and cognitive control networks to social learning involved phylogenetic construction or is entirely explicable in terms of developmental construction and inflection (42, 75), but either scenario is consistent with the IC premise that social learning substantially overlaps with associational learning mechanisms and comes at little additional cost (15). It also fits well with the close integration of individual and social-learning processes in real-world skill acquisition, as envisioned by the helical curriculum. Advocates of CE, however, emphasize the additional importance of a specialized ToM capacity to allow truly cultural learning (9). Does this requirement imply additional evolutionary costs for cultural learning?

Although ToM is commonly thought of as a human specialization, depending on phylogenetically constructed neurocognitive mechanisms, Heyes and Frith (76) have recently argued that it is largely a product of developmental inflection and cultural evolution. On this account, low-level or “implicit” mind-reading capacities emerge directly from motor resonance properties of the action control system discussed above. Motor resonance provides the input needed to identify recurring relations between actions, outcomes, and internal states, and thus to predict behavior and infer intent. This would be largely inapplicable in terms of general mechanisms [e.g., statistical and associative learning (73, 74)] or across phylogenetically constructed systems, as favored by some IC accounts (15).

Explicit mind reading, in contrast, involves active reasoning about mental states; in other words, the “theory” part of “theory of mind.” Anatomically, this appears to be supported by regions of posterior cingulate, medial frontal, and lateral temporoparietal cortex associated with the so-called “default mode network” (DMN) (78). The DMN was initially identified as a set of regions that experience de-activation during attention-demanding tasks, but is increasingly recognized to make a positive contribution to abstract, internally directed tasks involving information retrieval and integration. Examples include introspection, social cognition, autobiographical memory, future planning, narrative comprehension, and goal-directed working memory. This functional profile reflects the fact that the DMN sits atop the cortical processing hierarchy: maximally distant from peripheral sensorimotor systems and dominated by internal connectivity with other association networks (61). As discussed above, its development and function are thus expected to be highly plastic and reliant on learning. In fact, Heyes and Frith (76) propose that learning explicit mental theories is an inherently cultural process requiring language-based instruction. Their view is supported by, among other things, evidence that individual and cross-cultural differences in caregiver use of mental state vocabulary are predictive of variation in children’s acquisition of ToM concepts, such as false belief, knowledge vs. ignorance, and difference of opinion. Insofar as mind-reading capacities are themselves seen
as critical to language acquisition (9) and both may have been important to Paleolithic skill reproduction (27, 55), this suggests a deep and densely reciprocal history of biological, cultural, and developmental interactions during the evolution of the capacities that support cumulative culture.

An EES perspective thus charts a middle path between IC and CE extremes. The action-parsing and ToM capacities that support high-fidelity transmission do have a biological basis (cf. ref. 8) and are indeed specialized and highly derived in humans (2), but they may also be “secondarily modular” (15) products of processes other than phylogenetic construction (9). A long history of reciprocal interaction between cultural and biological evolution confirms the importance of considering CE processes (2; contra ref. 8) but contradicts the view of cumulative culture origins as a single “key event” (2) involving “one and only one biological adaptation” (9). Disentangling the respective contributions of IC and CE changes will require evidence of the sequence, timing, and context of evolutionary developments.

**Evolution of Primate Action Systems**

The gold standard for reconstructing such evolutionary developments is phylogenetic inference from comparative evidence of extant species (Fig. 2). Specialized neural machinery for visuomotor integration is perhaps the quintessential primate adaptation to a diurnal life in the trees (75, 79). Success in this niche was supported by the emergence of two new brain regions. The ventral premotor cortex (PMv) appeared anterior to primary motor cortex, and allowed for the integration of visual input with new, higher-order control of movement sequences. In the temporal cortex, the middle temporal (MT) visual area (or V5) appeared as a specialized motion-processing region. These basic adaptations are present in all primates and were in place at or near the root of our clade.

From this early template, further adaptations emerged. Motion processing expanded from the MT to produce a dorsal stream of “vision for action” extending into the posterior parietal cortex (80). The internal models of movement in space processed in this stream provide the “how” needed to execute or copy (Fig. 1) bodily actions. A second ventral stream of “vision for perception” extends into the lateral and inferior temporal lobe. This “what” pathway supports the recognition of objects, individuals, and body parts critical for organizing goal-directed action and emulating observed outcomes.

Whereas the presence of both dorsal and ventral visual streams across macaques, chimpanzees, and humans reflects the ancient origins of these systems, structural and functional differences between species provide evidence of subsequent evolutionary changes. It is clear that, like other association cortices, temporal regions associated with the ventral stream underwent substantial enlargement over ape and human evolution (62) and this likely fostered new functional capacities, ranging from semantic processing to face recognition. However, we argue that it was a series of key evolutionary changes to the dorsal stream that enabled integration of increasingly fine action details and complex goals during behavior observation and execution, and thus supported the emergence of high-fidelity social learning and cumulative culture.

One such change was the emergence of new functional regions in the parietal cortex. Comparative fMRI studies with macaques and humans have identified portions of human intraparietal sulcus with novel sensitivity to 3D form-from-motion stimuli, as well as a patch of human anterior supramarginal gyrus (aSMG) specifically responsive to observed tool use (81). Both regions are likely relevant to a wide array of evolutionarily relevant object-manipulation behaviors, and both are known to be recruited by stone tool-making activities in modern humans (82, 83).

![Fig. 2. Species differences in action processing circuitry (Upper) and inferred phylogenetic origins (Lower).](image-url)
increasingly complex foraging skills and brain evolution (13). Tool-relevant innovation appears to predate the chimpanzee–human split, as might be expected from the impressive tool-use capacities of modern chimpanzees (13).

Enhanced aSMG connectivity between dorsal and ventral steam is just one aspect of a broader pattern of changes (Fig. 2) in action circuitry over ape and human evolution (84). In macaques, this circuitry is dominated by ventral stream projections to the ventrolateral prefrontal cortex (vIPFC) along the inferior longitudinal fasciculus and extreme/external capsules, with relatively little connectivity through the dorsal stream. Dorsal stream projections to the vIPFC as well as the PMv through the middle and superior longitudinal fasciculi are better developed in chimpanzees and become quite pronounced in humans. Across these three taxa, there is thus a trend toward the addition of increasing dorsal stream inputs to the vIPFC in complement to established ventral stream connectivity. We have previously proposed that this dorsal stream enhancement may underlie the progressive elaboration of action-parsing capacities from macaques to chimpanzees and then humans (84, 85), as reflected by these species’ increasingly complex foraging skills (13) and capacities/propensities for bodily imitation (12).

Motor resonance mechanisms are least-developed in macaques, which are not known to imitate manual actions. Macaque “mirror” neurons respond to observed action goals rather than detailed means of execution and are almost entirely unresponsive to actions that do not involve an object (75). In contrast, chimpanzee action observation activates nearly identical voxels to execution of the same movements, regardless of whether they produce a physical result on an object (85). This basic action-matching mechanism may thus predate the chimpanzee–human split. With respect to object-directed actions, however, chimpanzees retain a generally macaque-like pattern of brain response dominated by the “top-down” contributions of the frontal executive cortex (85, 86). Humans alone display a more distributed pattern of occipital, temporal, parietal, premotor, and prefrontal activation, reflecting an increased role for bottom-up perceptual representations incorporating kinematic and spatiotemporal details about object-directed actions (85). This functional difference, and the structural changes that support it, may be critical to the exceptional development of skill acquisition and cultural learning capacities in humans.

In humans but not chimpanzees or macaques, the core action perception circuitry includes a prominent projection to the superior parietal lobule, a region associated with awareness of one’s body in space (84). Furthermore, the third branch of the superior longitudinal fasciculus (SLFIII), which links the inferior parietal cortex with the PMv in monkeys, extends into more anterior regions of the vIPFC in humans, particularly in the right hemisphere (87). Chimpanzees again appear intermediate, with a weak but observable extension of SLFIII into the vIPFC and no evidence of right-lateralization at the population level (87). Thus, a robust extension of SLFIII into the right hemisphere homolog of Broca’s area appears to be a human-specific adaptation. This region is an element of the multiple demand system discussed above, which is thought to support the assembly of complex, multistep action plans (88). The observed extension of human SLFIII would thus provide an anatomical substrate for the integration of kinematic details into complex action goals and sequences, as required for skill learning in a helical curriculum.

Such learning in humans requires a degree of bodily awareness sufficient to match variations in kinematic detail with desired outcomes during deliberate practice. A measure of such awareness that has been applied to other animals is the Mirror Self-Recognition (MSR) test (89). Unlike enculturated humans, mirror-naive animals must discover de novo that the visual per-ception of their reflection corresponds to the sensorimotor representations of their own movements. As might be expected from the preceding review of action perception circuitry, macaques typically fail at MSR and chimpanzees are intermediate in performance, with some passing and some failing. In fact, chimpanzee MSR performance is predicted by individual variation in the degree of right-lateralization of SLFIII projections into the vIPFC. In other words chimpanzees with more human-like SLFIII connectivity show more human-like MSR behavior (90). Because attending to one’s own movements is a critical element in the hypothesized construction of implicit mind reading from motor resonance (76), this finding suggests further links between dorsal stream evolution and the cognitive prerequisites of cultural learning.

What is not yet known is the extent to which any or all of these structural and functional differences between species are classic “adaptations” in the sense of being canalized products of phylogenetic construction vs. other evolutionary processes. Even in macaques, there is some evidence that extensive tool-training can produce plastic alterations in dorsal stream connectivity (91). Enlarged ape and human brains are expected to be more developmentally plastic and subject to inflection by somatic [e.g., bipedal locomotion, hand morphology (92)] and sensorimotor adaptations, and developmental niche construction (70). In fact, research with modern humans has shown that the acquisition of Paleolithic tool-making skills elicits plastic remodeling of dorsal stream white matter connections, including SLFIII’s projection into the right vIPFC, even in adults (37). Functionally, the gray matter targeted by this projection is recruited by execution (93) and observation (83) of relatively complex tool-making sequences of the kind that appeared with Late Acheulean hand-axe technology after about 0.7 Mya (50). Such findings suggest that further experimental studies of Paleolithic tool-making may begin to fill in details of timing, mechanisms, and context of evolutionary changes that occurred since the chimpanzee–human divergence and are inaccessible to purely comparative methods.

Conclusion: An Evolving Technological Niche

The earliest stone tools (Fig. 2) predate evidence of brain expansion by hundreds of thousands of years (32, 94) during which their occurrence was extremely patchy, discontinuous, and lacking in evidence of progressive change. By 2.6 Mya, early Oldowan knapping provides some evidence for high-fidelity cultural transmission of particular methods (35), as well as increasing demands on visual, motor and attentional systems (82), but the overall impression in this early period remains one of a tenuous and expendable technology at the edge of contemporary hominin capacities. It is only after about 2.0 Mya that stone tool-making appears to become more commonplace (as indicated by site frequency and geographic distribution), at which time it is accompanied by evidence of brain- and body-size increase (20). Further episodes of apparently correlated technological (50) and brain size (20, 95) change occurred with the appearance of increasingly skill-intensive Early (1.7 Mya) and Late (0.7 Mya) Acheulean knapping. Understanding what exactly changed at these various transitions is an important priority for future research and will ultimately require an integration of CE approaches to understanding technological change (19) and IC insights into the evolutionary economies of “expensive” brains (4), with mechanistic neuroscientific perspectives on evolving brain–behavior–culture interactions.
Such an approach highlights the central importance of embodied skills for object manipulation and modification to the high-fidelity social reproduction that IC and CE agree is critical to sustained biocultural feedback. Dorsal stream action systems constitute a critical substrate for the evolutionary-developmental cascade that constructs the action-parsing and ToM capacities needed for cumulative culture. These systems are themselves sensitive to inflicted input from more peripheral somatic or sensorimotor adaptations and constructed niches that shape early object manipulation and visual experiences (96). Object manipulation and modification contribute to the construction of learning niches (Fig. 1) populated by the residues of past action (52), and provide a persistent external medium scaffolding production of the more complex and protracted action goals and sequences (46) that both require and reward cultural learning through a helical curriculum.

As we have seen, stone knapping requires the multilevel integration of bodily kinematics and object affordances into goal-oriented action sequences. This finding is supported by phylogenetic (87) and plastic (37) enhancements to action control systems and demands extended investment in deliberate practice (36). Oldowan knapping, although far from easy compared with ape tool-use (32), remains a forgiving technology that is relatively quickly acquired (36) because the limited contingency between successive actions allows substantial latitude for error (31). Given sufficient sensorimotor and action control capacities, it is likely that social-learning opportunities (15) and niche-construction processes (52) seen in socially tolerant nonhumans would be sufficient for Oldowan skill acquisition (cf. ref. 29). Nevertheless intentional demonstration and linguistic instruction are clearly helpful (27), and such teaching did eventually become a vital part of human technological reproduction (53). Archaeological evidence cannot demonstrate that a particular form of teaching was essential at a given point in prehistory but does document transmission of quite complex and demanding techniques by Late Acheulean times (41), some of which modern humans find difficult to convey without explicit verbal instruction (40, 97).

Modern human teaching relies on social affiliative bonds that support extended interaction and motivation investment by teachers. Such investments are a special case of the alloparental contribution to the humanization of human reproduction (4), and their effective substrates likely evolved as part of a more general prosocial package (98, 99) relying on strong affiliative bonds. The interindividual behavioral alignment facilitated by resonance (74) is a critical mechanism that promotes such affiliation across many species (73), and in humans also supports pragmatic collaboration, apprenticeship learning, and the development of ToM, as discussed above. In contrast to other animals, however, human affiliation routinely extends to include “ultrasocial” cooperation and sharing with nonkin enabled by the use of language to create social norms and identities, including purely symbolic affinal and fictive kinship ties (98, 99).

A substantial body of research has linked language evolution to exaptation of the same action control systems discussed here (e.g., ref. 100), often with an emphasis on gesture production and action “syntax” in the dorsal stream (55). Less emphasized has been the critical role that such high-fidelity production (whether manual, vocal, or artifact-mediated) plays in grounding semantic meaning (101). Consistent physical “tokens” anchor meaning and allow discrete recombination in a way that continuous associative networks, however complex, do not. In other words, concrete tokenization appears critical to the transition from broadly semantic to truly symbolic thought (57), as exemplified by the construction of precise mathematics from approximate numeracy using number words as tokens (102). Once learned, this externalizing symbolic “trick” can readily be internalized as a tool for thought, supporting such abstract concepts as social roles that can be filled by different actual individuals. The elaboration of action control systems supporting skilled action and behavioral alignment (74) thus emerges as a foundational adaptation enabling the modern human way of life.

This way of life might fairly be described as a human cultural niche; however, this term is already taken to refer to a somewhat more circumscribed (CE) theoretical proposal (2). We have thus suggested the term “technological niche” (36) to emphasize the critical role for object manipulation and modification indicated by evolutionary neuroscience, as well as the necessary integration of cultural evolutionary thinking with the economic logic of the expensive brain to address biocultural interactions across evolutionary time scales. As should be evident from the preceding discussion, “technological” is not meant to imply a narrow focus on artifacts apart from the broader processes that generate them. Indeed, the now-ubiquitous term “technology” did not even come into widespread use until well into the 20th century, when it was adopted to describe: (i) a newly emerging and radically transformative power going far beyond anything implied by extant conceptions of the “mechanical” or “useful” arts, and (ii) the new normativizations of social and economic relations, institutions, and regulations that come together to constitute a unitary “thing,” like “automotive technology” (103). In this sense, we find the term to be a particularly apt description of the content, scope, and power of coevolutionary relationships that conspired to produce the modern human condition.