Current Research in Behavioral Sciences 5 (2023) 100121

Contents lists available at ScienceDirect

Current Research in Behavioral Sciences

journal homepage: www.elsevier.com/locate/crbeha

Tooling and Construction: From Nut-Cracking and Stone-Tool Making to Bird Nests and Language

Michael A. Arbib [a,*], Dorothy M. Fragaszy [b], Susan D. Healy [c], Dietrich Stout [d]

[a] *University of California San Diego*
[b] *University of Georgia*
[c] *University of St Andrews*
[d] *Emory University*

A R T I C L E I N F O

Keywords:
“how” versus “what”
Acheulean tool making
bird nest construction
capuchin nut cracking
construction (additive/subtractive)
hafted tools
image and assemblage
language evolution
mirror system hypothesis
motor programs
Oldowan tool making
pantomime
postconditions
preconditions
technological pedagogy hypothesis
tool use
tooling

A B S T R A C T

The present paper provides an integrative theory of actions and motor programs for skill in tool use, construction, and language. We analyze preconditions for action as well as making their effects (postconditions) explicit, emphasizing the “how” of action details as well as the “what” of motor programs, aided by conceptual analysis of several brain modeling efforts. The theory is exemplified by analysis of the subtractive construction involved in percussive tooling by capuchin monkeys and Oldowan and Acheulean stone tool making by protohumans before turning to the additive construction of hafted tools. A complementary analysis focused on the construction of bird nests explores the notion of “image” and “stage” in construction. We offer a brief comparison with birdsong before arguing for a very different relation between communication and construction in humans. Pantomime lifts manipulation from practical to communicative action in protohumans, and we consider the role of pedagogy before offering hypotheses on the emergence of human language that suggest how language may have evolved from manual skills. We note that language provides an open-ended means for devising innovations in tool use and construction, but reiterate the importance of this framework for diverse future studies in ethology and comparative psychology.

1. Introduction

Although humans share many components of their abilities with other animals, they are unique in the extent and diversity of their tool use and the objects and constructs they can produce, while the use of language as a form of mental construction and sharing extends the ability to develop, share, and teach diverse innovations. However, the relations between tool use, construction, and language are poorly understood. To further that understanding, we first extend the study of *tooling* (Fragaszy & Mangalam, 2018) to provide a framework for action more generally and the study of tool use in particular. Crucially, we extend this to include an account of *construction* rooted in linking hand (or beak) and eye. We apply this framework not only tool use in monkeys and humans but also nest building by birds and human use of language.

We take from Gibson (1977) that an *affordance* enables perception of the opportunity for an action, while emphasizing that it provides *parameters* for guiding that action. Actions take place within the *action-perception cycle*: actions may provide information about the world and/or may change the world (e.g., manipulation) or a creature’s relation to it (e.g., locomotion). Different creatures have different body forms and *effectivities* (capabilities for action) and different perceptuomotor systems (ways of sensing and linking perception to action). Thus, a given environment will offer different affordances to different creatures. Moreover, different creatures have different learning capabilities. In a given situation, a creature may recognize several affordances in the environment, but they may entail actions differing in difficulty, efficiency, safety, or other important currency – details that may be important for the selection of the action. Tooling adds perceptual-motor complexity to behavior by *distalizing the end effector* (Arbib, Bonaiuto, Jacobs, & Frey, 2009), a process by which the “end-effector” is

---

[*] Corresponding author
E-mail addresses: arbib@usc.edu (M.A. Arbib), dfragaszy@gmail.com (D.M. Fragaszy), sdh11@st-andrews.ac.uk (S.D. Healy), dwstout@emory.edu (D. Stout).

https://doi.org/10.1016/j.crbeha.2023.100121
Received 20 December 2022; Received in revised form 9 June 2023; Accepted 20 June 2023

transferred from the hand to some part of the tool, and the affordances and effectivities are now with respect to the tool. Crucially, what may serve as a tool for one creature may not be usable by another that lacks relevant body parts and perceptuomotor processes.

But humans do far more than simply use tools. Adding construction to our study increases our appreciation of how actions cumulatively change the environment. Technological objects and infrastructure in turn encapsulate information and persist across generations, constituting a novel channel of cultural transmission and evolutionary inheritance (Stout, 2021), a channel complemented by the rise of language. Between them, durable artifacts, recurring activities/situations and language attuned to these and developing social structures develop an expanding spiral for the development of exceptional human capacities for skilled interaction with, and transformation of, the social and physical world.

With this we turn to a brief overview of the paper as comprising three interwoven tracks:

- Theory of Action and Construction
- Case Studies of Action, and
- From Action to Language.

1.1. Theory of action and construction

The P}A}E *Framework* (§2): To relate tooling with actions more generally and construction in particular, we emphasize the *preconditions* and *effects* (*postconditions*) for an action or behavior. Our concern is not only with affordances but also with how each action changes the relevant (small part of) the world. The notation P}A}E signifies that "If precondition P is satisfied it will be possible to execute action A (so P must include provision of affordances for A), with the likely result that effect E on the external world will be achieved." The word "likely" indicates that effect E may not be achieved and that corrective action or replanning may be required.

We will also introduce two key distinctions: that between proximal and distal goals, and that between event-level and trajectory-level processes.

Tooling Revisited (§3) extends the notion of tooling (Fragaszy & Mangalam, 2018) by making some use of the P}A}E notion and then adapts their notions to tool use. §4 "interrupts" the Theory Track to analyze capuchin nut cracking, the primary example analyzed by Fragaszy and Mangalam.

Controlling, Learning, and Recognizing Single Actions (§5). The ventral pathway for visual control of hand actions in primates analyzes the scene to enable the prefrontal cortex to determine "what" to do; whereas the dorsal pathway then fills in the details of "how to do it," passing affordance parameters to premotor cortex to adjust motor schemas for the selected actions. We emphasize the difference between recognizing an affordance as a basis for selecting actions and using the details of that affordance to guide action details. This is a general principle, as applicable to birds as to primates. We then explore how affordances and effectivities may thus be learned together to meet the "how" requirement, assessing what "innate" properties (a notion we will handle with some care) make such learning possible, a theme to be explored in several case studies.

Combining Actions into Behaviors (§6) introduces our approach to "motor programs." While an *ethogram* provides a basic *description* of how action sequences may be scheduled, it does not represent "what the brain does." We present as one alternative the notion of *opportunistic scheduling*: Here, actions "compete" and the one with highest priority "wins," where the priority of an action depends both on its *desirability* with respect to current goals and also on its *executability*.

Construction: Additive and Subtractive (§7). Here we introduce both *additive construction* (putting objects together) and *subtractive construction* (removing portions of an object to make a more desirable one) while noting that objects may also be transformed in diverse ways to complete a construction goal. We emphasize that a complex behavior, including a complex act of construction (as in building bird nests, §9), requires multiple stages, thus placing demands not only on immediate working memory for keeping track of actions within a stage, but also long-term working memory for keeping track of the stages in some overall task. We also assess the notion that, at any stage of a construction task, the actions may in some sense depend on selecting actions that bring the partially completed construct closer in form to some *image* (*not necessarily visual*, not necessarily precise, but more in terms of the combined activation of certain features linking action with multi-sensory perception) of what the construct should look or feel like when that stage is completed.

1.2. Case studies of action

Capuchin monkeys (§4) employ percussion in their act of tooling, using a "hammer stone" to strike an object (a nut) resting on an "anvil stone." The success criterion is cracking a nut "just enough" to make the kernel inside available for eating, and so the skill involves repeated hammer blows until a strike breaks the nut. We discuss the extended (multi-year) practice where one skill (noisy hammering) provides the basis for another (successful nut cracking).

Bird Nests (§8) takes us from primate to bird and from hand to beak (and from manipulation to *becculation*) to consider building bird nests as a key example of *additive construction*: deliberately placing a grasped object on or into an emerging target object/surface by first grasping an object, and then using the resultant body-plus-object system to manage spatial relation(s) between the grasped object and a target object/surface, *so that the grasped object, when released on or in the target surface, remains in contact with the target surface***.** We briefly discuss birdsong to highlight its contrast with human language and learning.

More generally, additive construction may or may not require tooling and may include processes like using adhesives, or bending objects before they can become part of an assembly.

The Oldowan-Acheulean transition and on to assembly (§9). We initiate analysis of the evolution of (proto) human skill by first assessing how Oldowan flake production became a component of Acheulean shaping of stone tools*:* The Oldowan success criterion is removing a "satisfactory" flake from the core, where "satisfactory" rests on the utility of the flake itself to serve as a tool. By contrast, the Acheulean success criterion is sculpting a tool from the core, with repeated removal of a "satisfactory" flake from the core being a repeated subgoal, but where "satisfactory" alludes to the change in the core resulting from removal of the flake. These offer examples of *subtractive construction***:** subtracting something from an object to bring it closer in form to a target object. §9 closes by bring *additive construction* into the mix, analyzing hafting.

1.3. From action to language

Language: Finally, we argue that the human capacity for *physical* construction provide a key to an understanding of how *language* evolved as a *mental* construction system that serves communication rather than directly shaping physical objects. The posited evolution of the *language-ready brain* will rest in part on the hypothesis that pantomime evolved in part to link manual skills and their pedagogy. The argument extends over 3 sections:

Skill acquisition and pedagogy (§10)
Grammars for language and action (§11)
Language emerging: The Mirror System Hypothesis (§12)

In language, we pair two acts of construction: we construct a sequence of words and we construct a meaning, guided by a grammar that supports a *compositional semantics*, in general with the intention to satisfy the *parity principle* that the meaning understood by the hearer (or observer) resembles that intended by the speaker (or signer). We adopt the *construction grammar* approach that a language combines a lexicon with a grammar defined by a large number of more or less language-specific "constructions" in the linguist's sense, each of which combines form (how to put words and/or phrases together) and meaning (how to assemble the meanings of those pieces). We ask the reader to distinguish when we use the term "construction" for the process and result of combining elements versus in the linguist's sense of "a tool for putting words and meanings together."

2. The P } A } E Framework

We make two key distinctions relevant to analyzing behaviors:

Proximal versus Distal Goals: Tooling and construction require behaviors that combine many actions. The actions are steps towards a single shared goal, the *distal goal*, yet each may have a distinct *proximal goal* that shapes that particular action. For capuchin monkeys cracking nuts, the distal goal is "eat the meat from the nut," whereas the action before the strike has the proximal goal "position yourself for a good strike."

Event-Level versus Trajectory-Level Processes: Many psychological models are *event-level*, with each event or trial or action considered an indivisible whole, and with emphasis placed on the stringing together of distinct behavioral events by decision processes. By contrast, *trajectory-level analysis* analyzes, e.g., the trajectory of the hands during an action, with mastery of the action requiring learning to adjust the parameters of the trajectory to the current situation. Learning thus operates both at the event-level and for the tuning of varied subactions.

Our concern is with how each action changes the relevant (small part of) the world. The general notation P}A}E signifies that "If precondition P is satisfied it will be possible to execute action A (so P must include provision of affordances for A), and if A is then completed successfully, effect (postcondition) E will be achieved in the external world." If E is not achieved, A will not have been successfully completed, and some "control architecture" must decide whether this signals "keep on doing A" or "abort *this* attempt at A: start A over or try something else." (The same applies to an overall behavior B, but there problems may yield changes within the strategy of B.)

[In the next 6 or 7 lines, use primes, not quotes in A'] In executing A then A', with P}A}E and P'}A}E', it is contingent whether E is related to P'. All that is required is that before A' is attempted, the environment is such that P' holds – but this may depend on other actions or a different part of the environment from that on which A acted. Contrast:

- A: Brush your teeth; A': Take a shower. Here neither sets preconditions for the other
- A: Take a shower; A' Dry your body. Here A sets the precondition for A'. Although possible, it would be "silly" to reverse the order.

The order of actions may result from planning, or be scheduled as the next step(s) in a sequence, or be opportunistic on noting that the environment now contains the affordances for a desirable action.

The top-level framing of an overall behavior might have something like "The environment (probably) can support this behavior, and the participants are motivated to perform it." For example, the behavior of knapping – the detachment of *flakes* from a stone *core* using ballistic strikes with a hand-held *hammer* to initiate controlled and predictable fracture – would only be invoked in a region in which stones and hammers can be found, and if the agent has the motivation and skill (with or without an instructor). Given these, the behavior can be executed with a fair expectation of success. Further, each such behavior is not only itself a complex of actions, but can be linked to other behaviors, e.g., quarrying, knapping, polishing. Similarly, actions for bird nest construction may be interleaved with foraging for materials.

3. Tooling revisited

Fragaszy and Mangalam (2018) – FM for short – define tooling as follows:

Definition FM: *Tooling* involves deliberately producing a mechanical effect upon a target object/surface by first grasping an object, thus transforming the body into the body-plus-object system, and then using the body-plus-object system to manage (at least one) spatial relation(s) between the grasped object and a target object/surface, creating a mechanical interface between the two.

With this, "A bout of tooling begins when the tooler acts to establish the first spatial relation in the tooling sequence and ends after the last spatial relation in the sequence is established. For example, in hammering a nail, tooling begins when the tooler places the nail against the board, the bout continues while the tooler strikes the nail with the hammer, and ends when the tooler stops striking the nail and switches to some other activity" (Fragaszy & Mangalam, 2018, p.194). Implicit here is that the grasp is maintained on the same object as a basis for exploiting the same mechanical interface.

Our key change in emphasis is to distinguish between the tooling activity per se and its employment *in the service of achieving a goal*. In the latter case, we will refer to the object first grasped in Definition FM as a *tool* that is being *used* to help achieve the goal. Given the distal goal of making a change in an object or objects, if one chooses to make that change using a tool then one must find the tool and then use it to complete the episode. In P}A}E terms, the availability of the tool is part of P, the action A involves the use of the tool, but the tool may not be part of E. We adopt the following definition of tool use on which we will play variations since no one definition suffices for this protean concept:

Definition AFHS: *Tool use* involves deliberately producing a mechanical effect upon a target object or objects by first grasping an object, known as *the tool,* and then using the body-plus-tool system to transform the target object(s) into a desired form, this constituting the goal of using the tool. In any particular bout of tool use, the intended goal may or may not be achieved. We thus must distinguish incremental progress – hammering the nail in a bit further – from a "red flag" like bending the nail, an undesired outcome that ends that particular attempt at the tooling activity or causes a modification in strategy.

A bout of tool use could itself be a subroutine in a larger behavior – as when hammering nails serves an overall task of putting multiple pieces of wood together.

3.1. Goals and goal-directed behavior

Before going further, we need to say more about the notion of a "goal." We cannot know if animals have *consciously formulated* goals but there is consensus (Gwan & McShea, 2020; Trestman, 2012) that voluntary behavior in many situations is goal-directed. The P}A}E notation is more neutral, labeling E as an effect rather than a goal. Inferring goals may be muddled but we need the concept, and assessing what is relevant in directing the behavior of a human or an animal remains an enduring challenge for studies in psychology or ethology.

3.2. Adjusting the "axioms"

We now present four "axioms" for tooling (Fragaszy & Mangalam, 2018), and will then adjust them for tool use more generally before turning to their key example of capuchin monkeys hammering to break nuts open.

Axiom FM 1. An individual perceives the potential of producing a mechanical effect upon a target object/surface with a grasped object by perceiving affordances incorporating actions with objects.

Axiom FM 2. A grasped object transforms an individual's body into a body-plus-object system, reducing or redistributing the existing degrees of freedom, and adding at least one new degree of freedom (between grasped object and target).

Axiom FM 3. An individual creates a mechanical interface with a target by establishing (at least one) spatial relation(s) between the grasped object and the target.

Axiom FM 4. Through interrelated processes an individual learns to:

(a) Perceive affordances incorporating actions with objects to produce a mechanical effect upon a target object/surface. (b) Manage the spatial relations between the grasped object and the target object/surface to create a mechanical interface between the two. (c) Coordinate and control the body-plus-object system to produce specific mechanical effects on the target object/surface.[1]

In (manual) tool use, an object becomes a tool when it is grasped to perform an action (in attempting) to transform another object or objects to achieve a specific goal. Tools are employed either to achieve a goal that could not be achieved with the unaided body, or to reduce the effort or undesirable side-effects required to achieve a goal. In the case of capuchin monkeys or Oldowan manufacture, the tool is a rock that has certain relevant characteristics. It is a "found" tool. However, the Oldowan flakes are "made" tools. Modern human tools may be designed for one purpose, but then may also be used as a tool for a different purpose, as when using a screwdriver to open a can of paint. An important challenge is to explore how the individual can learn to use an object skillfully to achieve the goal. This requires being specific about the way in which the changes in the applied mechanical forces serve the purpose for which the tool is used.

The precondition P for a tooling action A must include the availability of an affordance for making a desired change in the environment; A must include recognition of the affordance, and then pass parameters of that affordance to the action that A performs in attempting to effect that change. If the behavior includes using a tool, then positioning the tool relative to the target is part of the tooling. A may include multiple and repetitive subactions for its completion. We now adapt the FM "axioms" to get their AFHS versions (AFHS for the 4 authors here):

Axiom AFHS 1. Grasping a tool changes effectivities of, or adds new effectivities to, the agent. An individual perceives the potential of producing a desired effect upon a target object or objects with a tool by perceiving affordances for actions that can achieve the desired end result when the tool is used.

Where Definition FM stresses mechanical effects, we stress the end to which those effects are aimed. In a given situation, achieving a desired E may require (or be made more efficient by) a P}A}E where the P includes the availability of the tool and the action A involves its use. Consider the aim of using a hammer in knapping is to dislodge a flake. If you don't dislodge a flake, you have not achieved the goal. Correction implies reassessing affordances, so that action will often change accordingly.

Axiom AFHS 2. A tool transforms an individual's body into a body-plus-tool system, redistributing and possibly changing the existing degrees of freedom to provide new effectivities that require the perception of possibly distinctive affordances in the environment.

Indeed, even without tools, specific grasps (e.g., precision pinch versus power grasp) reduce the number of degrees of freedom of the hand (by coordinating synergies of available degrees of freedom, not by selecting a few degrees of freedom from the prior set) – as in the notion of a grasp-specific "opposition space" (Iberall, Bingham, & Arbib, 1986).

Axiom AFHS 3. An individual perceives new affordances in relation to effectivities made possible by establishing a mechanical interface between the tool and a target. This interface varies with the affordance selected which in turn depends on the task (and context).

Axiom FM 4 is simply a way of stating "the tooling can be learned," but where Fragaszy and Mangalam stress mechanical effects, we stress the effect of carrying out a general action:

Axiom AFHS 4. Through interrelated processes, an individual learns to use a tool to achieve a particular type of goal, mastering new effectivities and the new affordances that the environment offers when the tool is being used. Skill involves passing relevant parameters of affordances to better deploy the corresponding effectivities, matching the action with relevant observed properties of the environment.

Again, we see the transition from "tooling," in which an object may be employed in applying a mechanical force, to "tool use" in which an object is employed to develop the mechanical forces required to achieve some type of goal. We then call the object so employed a "tool" – but this is not to be confused with the words we use to label human-made objects. In the case of capuchin monkeys (to which we now turn), the monkey learns to choose rocks that will serve well the role we call "hammer" in breaking nuts – but there is no general category of "hammer" explicitly shared by a group of monkeys.

A reviewer suggests that to sustain the claim that capuchins have "no general category of 'hammer'," there "would need to be a discussion of the relationship between affordance, categorial perception ('seeing as') and conceptualization ('seeing that'). The issue … bears on the question of language precisely because (proto-) words … [serve] to guide attention, memory and voluntary action through semiotic mediation." This is indeed consistent with the view of language evolution offered in §§10-12 – the blend of action and language that humans have achieved holds an important clue, as "protowords" and "protoconstructions" co-evolve culturally to take on a broader range of meanings and combination of meanings, respectively. Once a word comes into limited use, one may also find its use apposite, though non-standard, in situations where no pre-existing word or phrase seems applicable. Thus metaphor is born, and a word that had an established meaning in one domain now becomes available in other domains. In some cases, the use of the word evolves so that it comes to have truly distinct meanings. In the case mentioned here, the initial use of *hammer* for a specific type of object by humans informs the naming of the attendant action – but then each becomes separately untethered from the early specifics. Lacking the equivalent of the English word *hammer,* adult capuchins cannot, for example, articulate criteria for what constitutes a "hammer," but are limited to developing individual criteria for judging whether a rock helps satisfy the preconditions for attempts at cracking a nut. With practice, the monkey better adapts details of action execution to affordance details of the objects involved.

In summary, the non-human brain offers no mechanisms of language-assisted transfer of skill (whether in [multi-modal] perception or action), or attendant generalizations from one domain to another. However, in the next section we shall see that "culturally-mediated" generalization can be available. For example, capuchin monkeys take years of practice to successfully acquire the skill of cracking palm nuts, but can adapt that skill even while it is still imperfect to crack softer cashew nuts.

4. Capuchin monkeys hammering to break nuts open

We now consider the example of wild bearded capuchin monkeys (*Sapajus libidinosus*) at Fazenda Boa Vista in Brazil hammering palm nuts to break them open (Visalberghi & Fragaszy, 2013 offer a broad description), calibrating the revised theory of tooling to point the way to the theory of "extended manual action" in the next section.

4.1. The skill and its acquisition

A capuchin monkey finds a palm nut that it must crack open to get at the kernel. To this end, it places the nut on a flat stone or log ("anvil"), usually in a pit on the surface of that anvil, and places a "hammer" stone – the heavy stone used to crack resistant palm nuts – on the anvil between itself and the nut. The monkey positions itself on the anvil behind the hammer stone while facing the nut, grasps the stone in both hands, and in one continuous motion, raises the hammer to about head or shoulder level and lowers it rapidly downward to strike the nut (Liu et al., 2009). This may simply displace the nut, or it may hit the nut without cracking it open. But, in due course, the skilled monkey delivers a strike that cracks the nut open to make the kernel accessible. Extended practice increases skill by mastering not only spatial relations but also

---

[1] There is also an Axiom FM 5, "The component processes in tooling (perceiving affordances, managing spatial relations, and coordinating movements in action) demand perceptual-motor resources," but we see this as part of our general action-perception framework, rather than being specific to tooling.

the bodily forces appropriate to the current details of affordances.

The tooling event comprises the actions from the placement of the nut on the anvil until the monkey stops striking. For us, this may indeed be considered “tool use” if we replace “until the monkey stops striking” with “until the monkey cracks the nut open.” Of course, the monkey may fail, a case of “unsuccessful tool use.” The key point is to specify a goal that sets the success criterion – so that we may distinguish “percussion” from “nut breaking” even if the same tooling is involved.

Learning operates both at the event-level, and for the tuning of varied subactions. Observation of the capuchin monkeys show an accumulation of relevant skill rather than learning of nut cracking *ab initio.* Young monkeys in their first two years begin to perform an event-level behavior –grasp an object, strike it on a surface – that will later provide a key component of nut-breaking. They do this in their first year of life in playful settings independent of cracking nuts, usually well above the ground surface. When they begin to spend time on the ground, where most anvils are found, they begin striking nut shells or small stones on other nut shells. This behavior seems not to be directed to the goal of nut-cracking. However, usually around two to three years of age, they place and release a nut shell on an anvil surface, then strike it with another shell or small stone, mastering the required spatial relations and the required event-level behavior (position nut on anvil, strike nut with stone) (Eshchar, Izar, Visalberghi, Resende, & Fragaszy, 2016).

At Fazenda Boa Vista, monkeys take two or more additional years of practice to begin striking with a stone forcefully enough to crack a palm nut. Monkeys at a site with different species of palm trees, the nuts of which are smaller and less resistant to fracture, begin to place nuts on an anvil and strike them with a stone around the same age as monkeys in Boa Vista, but they can crack nuts before they are three (Resende, Ottoni, & Fragaszy, 2008) suggesting that beginning mastery of action tuning (trajectory level) can appear by age 3 years. The long period of ineffective effort at Fazenda Boa Vista reflects the resistance of the palm nuts at that site, leading to the requirement to use heavy stones relative to the monkeys’ body mass (Fragaszy et al., 2016). Young monkeys at Fazenda Boa Vista do not achieve adult proficiency at cracking palm nuts until 6 years or older.

But why persist for years before success?

1) Striking objects on a substrate is a common species-typical foraging action in capuchin monkeys. Striking, per se, is a pre-potent behavior that capuchins perform with virtually any object held in the hand.
2) Until about two years of age, young monkeys are permitted to “scrounge” bits of nut remaining on the anvil while or just after an adult has cracked a nut (Coelho et al., 2015). Other monkeys cracking nuts are noisy and visually interesting, and they potentially provide a source of tasty bits of nuts. It is not surprising that they are closely observed by young monkeys. Occasionally they obtain a partially-cracked nut, and may be able to crack it into smaller pieces, providing added reinforcement for subsequent efforts to crack nuts.
3) Nut-cracking activity by others facilitates (increases the likelihood of) performance of actions with nuts by young monkeys while others are cracking and for a period of seconds to minutes after others stop cracking. This effect may support the development of sustained attention by young monkeys to their own activities with nuts (Fragaszy et al., 2017).
4) In a study over three annual observation periods, Fragaszy et al. (2023) assigned young and adult bearded capuchins to novice, intermediate or expert classes in accord with their success at cracking nuts. Their findings suggest practice using the body-plus-tool system for cracking palm nuts supports affordance learning and results in gradual mastery of this skill and that changing body mass plays a small role in this process.

In humans, social interactions associated with joint activity and experienced repeatedly from the first year of life influence the development of attention (to actions), shaping learning processes in culturally relevant ways. Indeed, learning processes (not just what is learned) are culturally variable (Flynn, Laland, Kendal, & Kendal, 2013; Rogoff, 1991; Yu & Smith, 2016). On this view, capuchins learn to crack nuts in part through the indirect influence of others helping them to perform particular (initially uncommon) actions, and to extend the time spent doing them, thus helping them to acquire the sustained attention for the particular actions needed to master nut-cracking – even though mastery of percussion long precedes mastery of the skill of nut-cracking. The tradition of nut-cracking is accompanied, we believe, by tradition-specific learning processes. Moreover, the presence of anvil stones with hammers and the durable shells of cracked nuts on and around anvils provide long-lasting social cues at times when others are not cracking nuts (Fragaszy, 2011). Thus the socially-constructed niche provides many avenues for support of young monkeys’ continued interest in nuts, hammer stones, and anvils, and in practicing striking nuts.

A further source of support derives from the monkeys’ actions with cashew nuts. These are far less resistant than palm nuts, and small, light objects are adequate to crack them. The youngest monkey to open a cashew nut by percussion in one study (Visalberghi, Barca, Izar, Fragaszy, & Truppa, 2021) was less than three years old. Thus young monkeys attempting to crack palm nuts in their fourth year and beyond likely have experienced success cracking cashew nuts with a similar action set, and they have frequently practiced the same sequence of actions (collect nut, travel to anvil site, find a suitable hammer, place the nut, and proceed to crack it) in another context.

The ability to master new skills varies from species to species. We have seen that the capuchins’ skill involves behavior on (at least) two spatial scales and associated timelines. We focus here on the small scale (a meter or two) and short timeline (a few minutes) of establishing spatial relations between nut, anvil, hammer stone, the monkey’s stance during striking, and the action of striking. The other is the large scale (hundreds of meters; many minutes) of knowledge of the territory – to know where hammers and anvils are located, to know where to search for nuts and, once nuts are collected, to know how to go to an anvil.

Even in the small-scale behavior, there are four tasks after the hammerstone has been placed on the anvil:

Task 1: The hand is the effector, and the target is the nut, and *the nut* is grasped in a way that prepares for task 2.

Task 2: The nut becomes the end-effector, and the task is to position it to sit securely in the pit *of the anvil* and then release it. Then [here or before 1] the monkey positions itself behind the stone facing the nut.

Task 3: The hammer stone is handled until both hands grasp the hammer in a firm grip.

Task 4: The stone is lifted and lowered using a whole body motion, lifting from the ankles, hips, and knees. Here the hammer is the end-effector, but it takes the whole body to move it.

Step 4 may or may not crack the nut. The P for the overall behavior involves the elaborate preparation provided by Tasks 1 to 3, while the E involves the state in which the nut is cracked and the kernel is accessible. The overall behavior includes a “repeat until nut is cracked” loop. In general, a repeated attempt will involve the whole sequence 1-4, unless the nut remains satisfactorily placed, in which case Tasks 1 to 2 may be omitted. Our discussion of the young monkey suggests that for a few years, the E for its action may instead have been *make a loud noise* or *hit something on a substrate*, and the “best” affordance in Task 2 would still need to be learned. Again, increasing skill in Task 3 becomes part of achieving success in Task 4.

Shared characteristics of capuchin monkeys that support learning to crack nuts include

1 a predilection for pounding,
2 generative manipulative behavior (generating new combinations of actions, objects and surfaces),
3 attraction to nuts as a valued food (worth trying to open),
4 good navigational abilities and memory for landscape features,

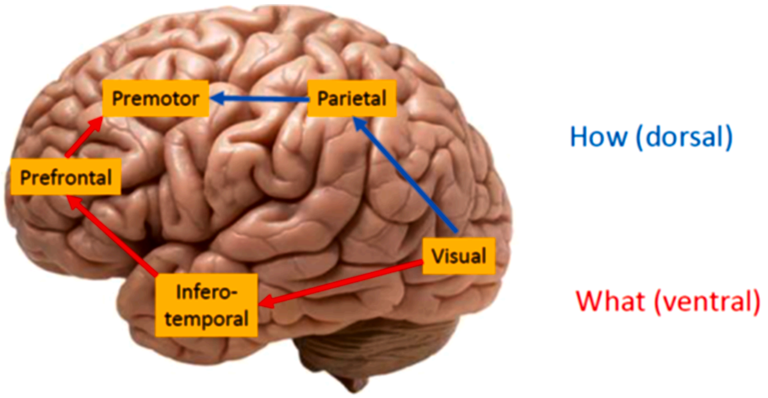

Fig. 1. The ventral and dorsal pathways for visual control of hand actions. The ventral pathway analyzes the scene to enable the prefrontal cortex to determine "what" to do; the dorsal pathway then fills in the details of "how to do it," passing affordance parameters to premotor cortex to adjust motor schemas for the selected actions.

5. facultative bipedal stance and the mastery of dynamic balance that entails,
6. ability to develop joint synergies supporting the controlled movement of stones,
7. attentional processes that allow them to work on the target of their interest (a nut) indirectly with a stone while they are NOT holding the nut, and
8. haptic sensitivity to the movement of an object on a surface.

Of this list, 1, 2, 5, and 7 are most likely to differentiate capuchins from other monkeys, but in fact we have little to no specific knowledge (from empirical study) of these characteristics in most species, including capuchins. For our discussion of construction (§7) and language (§12), 2 and 7 may be most relevant: Generative action and attention to movements of objects distant from the body. Attention to object(s) during movement is expandable in time and space (through experience), and the "workspace" for action is expandable (through generativity).

All this exemplifies our aim to create a framework adequate for generating hypotheses both for ethology and for human evolutionary studies. An ongoing challenge here as in all ethology is "cutting" the ongoing behavior into distinct actions that combine to yield behavior, and minding the relevant Ps and Es the way the animal does, rather than being misled by how humans think the animal might do it.

4.2. General Implications

All this exemplifies the notion that a new skill requires both mastery of the sequence (or more general conditional relationship) of actions and the tuning of actions through practice to reliably achieve the goal. Among the relevant processes are:

(a) Recognizing that an object can be acted upon to get something with desirable properties (recognize affordances/preconditions to achieve a desired effect/reach a goal).
(b) Attending to salient actions of someone achieving the goal, accompanied by experiencing facilitation to perform similar actions (this does not require a further step of recognition of a specific action).
(c1) Improving control over one's own attention to achieving the goal, resulting in increasingly consistent effort toward solving a problem.
(c2) Recognizing affordances more precisely. This involves perceptual learning about the objects, and about outcomes of acting with them and on them, but this may also engage motor learning, tuning actions that have been crudely incorporated at the event level to better match observed affordances.
(c3) Increasing the speed of performing all the relevant actions.
(c4) Tuning these actions by trial and error and/or (in humans only) purposefully modifying them to improve skill.

In §9 we will turn from how smashing one rock on another object is *used by capuchin monkeys* to get the meat out of a nut, to stone knapping by (proto) humans where flakes are detached from a core either to be useful in themselves (Oldowan) or to shape the core into a tool (Acheulean). Thus the present study of capuchins will provide another perspective on s(proto)human tool use directed at tool making.[2]

As we discuss in §10, humans may have additional sources of influence on learning a motor skill, prominently including what we call *complex imitation* (other species, we suggest, have simpler forms of imitation) and *pedagogy*.

5. Controlling, learning, and recognizing single actions

Our focus has been on *tool use that exploits manual skill*. We turn to a general account of manual skill in monkeys and humans, but with clear implications for the study of bird nest construction below. In this section, we focus on the control, learning, and recognition of single actions; §6 then discusses ways of combining actions into behaviors.

These sections will offer conceptual overviews, rather than details, of several computational models based on brain and behavioral data from macaques and humans with the aim of enriching our study of tooling and construction. The models shown here are but a small and personal sampling of a rich literature that can offer new insights and techniques for the study of cognition and behavior.

5.1. Two pathways for affordances

The **FARS (Fagg-Arbib-Rizzolatti-Sakata) Model** (Fagg & Arbib, 1998), based in part on macaque neurophysiology but related to humans as well, explains how the brain may use visual information to guide the hand in grasping an object, addressing the need to make decisions between multiple affordances. For example, one may grasp a mug by the handle or by the rim or by a grasp around the body of the mug if one wishes to lift it (*proximal* goal). Moreover, the choice (possibly nonconscious) of which affordance to exploit may depend on the *distal* goal – one may be more likely to grasp the handle if one plans to drink rather than if one just wishes to move the mug elsewhere. Moreover, deciding which affordance to exploit is not enough – details of size, shape and location must be perceived and passed to the "motor

---

[2] Alice Auersperg's lab in Vienna has contributed to the theory of tooling by studying a range of novel behaviors in cockatoos (Colbourne, Auersperg, Lambert, Huber, & Völter, 2021).

schemas" that control the movement (Arbib, 1981).[3]

We distinguish the dorsal and ventral pathways for vision related to manual control (Fig. 1). The model exploits neuropsychological data (Goodale, Milner, Jakobson, & Carey, 1991; Jeannerod, Decety, & Michel, 1994) to apportion computations between two visual pathways: The *ventral ("what") pathway* – from primary visual cortex via inferotemporal cortex – supports object recognition to provide input that the prefrontal cortex can combine with motivation and working memory to plan a sequence of actions, but without the fine details needed for graceful execution. The *dorsal ("how") pathway* – from primary visual cortex via parietal cortex – provides metrical details relevant to the current action (and possibly the transition to the next) to support tuning of relevant motor schemas.

This suggests a general principle, as applicable to birds as to primates, that emphasizes the difference between

(i) recognizing an affordance as a basis for selecting actions and
(ii) using the details of that affordance to guide an action that exploits it.

This dichotomy is not limited to visual processing but applies to other senses as well, including touch, proprioception and hearing. In general, perception for action is multi-modal and involves activation of diverse perceptual schemas in different modalities to assess objects, actions and relations in the environment.

The latter requires learning (probably nonconscious) that supports automatized feedforward control. Clearly, the choice of an affordance depends on the effectivities of the agent, and its current goals and motivation. For example, even in Acheulean technology (shaping a core stone into a usable tool), the core must be examined to determine where to remove the next flake. To this end, diverse possible targets for the next hammer blow must be examined before the strike is made. Similarly, a bird inserting twigs into a nest it is building must assess where and how to insert the next twig.

5.2. *Mastering affordances and effectivities*

Having noted the importance of matching affordances and effectivities, we view a computational model of how they may be mastered together – in this case matching the shape of *part of* an object (a potential affordance, such as offered by the rim, the handle, or the body of a mug) to mastery of a successful way to grasp it. This model builds on efforts (Oztop, Arbib, & Bradley, 2006; Oztop, Bradley, & Arbib, 2004) that include a review of studies of infants learning to grasp, rejecting the notion of an innate maturational timetable in favor of one that addresses the cumulative impact of learning.

ILGA, our model of Integrated Learning of Grasps and Affordances (Bonaiuto & Arbib, 2015), shows that well-known distinctions in the literature concerning the forms of grasping, like "precision pinch" versus "power grasp," can be formed by learning mechanisms that suffice to explain a variety of other grasps adapted to, for example, the use of particular tools. We showed, through computational modeling, how an existing behavior (reaching) may yield a behavior (grasping) that is more complex (in the sense of more precise adaptation to the current external circumstance) through interactive goal-directed trial and error learning. A range of actions and the detailed affordances that support their execution are acquired together, based on a reinforcement learning mechanism ("joy of grasping") that reinforces the formation of a hand shape together with recognition of the shape of part of an object if the grasp is stable (the object does not slip from the grasp).

This draws attention to the general issue: What developmental starting point of perceptuomotor competence (a "rough program") and what reinforcement criteria must exist to yield the emergence and subsequent coordination of the behaviors we study? The imprecise notion of "rough program" here is meant to convey the observation that the infant seems to have an "innate" propensity to reach toward objects it sees, but this is initially highly inaccurate. However, with successful contact, the visuomotor transformation from visual input to arm movement becomes well-tuned (Kuperstein, 1988). A "rough" program becomes increasingly "smooth." Similarly, ILGA models the stage where the infant is able to bring the hand into contact with the object and may reflexively close the fingers around the object if it contacts the palm, leading occasionally to a stable grip. The model demonstrates how, with only achievement of a stable grasp for reinforcement, the child may come not only to successfully grasp an object but also to recognize the visual affordance that a part of the object offers for that novel grasp, and learn to adaptively preshape the hand to successfully grasp the object there.

We seek a framework in which the quest for both "rough programs" and learning principles can proceed.

5.3. *Recognizing manual actions*

Some skills are acquired based in part by learning through observation of the actions, including possible goals, of others. One small but important part of the quest to understand the underlying brain mechanisms was the discovery of mirror neurons for grasping in the macaque premotor cortex. Many writers ascribe too large a range of cognitive functions to such neurons, and our later discussion of imitation will show that mirror neurons alone cannot support it but must work with neural systems "beyond the mirror." In this section we address the tightly focused question: how can one go from an action mastered through trial and error to the ability to recognize that someone else is executing a similar action? In §6, we will suggest that mirror neurons play a role in learning motor skills in part through observing *one's own* actions, and not just observation of others.

Mirror neurons have been characterized as neurons that fire in relation to the same limited set of actions in the agent's repertoire, whether those actions are being executed or observed (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). Mirror neurons in macaque monkeys have been found for such manual tasks as tearing paper and breaking peanuts (Keysers et al., 2003), making clear that mirror neurons are in general the result of learning. However, mirror neurons may become tuned as part of the execution of the movements before or after they become tuned to the recognition of similar actions executed by others. Here, we focus on the former, *before*, case; the latter must hold when learning novel actions through observation.

The **MNS** (Mirror Neuron System) model (Oztop & Arbib, 2002) suggests how mirror neurons for manual actions might be shaped by learning during observation of one's own actions by forming associations between motor commands and the trajectory observed during their execution. During training of mirror neurons for actions already in the repertoire, canonical neurons (modeled by FARS) controlling the monkey's grasp are posited to activate, via corollary discharge, a set of potential mirror neurons. This "canonical code" for a grasp serves as the training signal for the latter neurons to learn to recognize the corresponding *trajectories* of the hand moving in relation to the selected object affordance via features coding the movement of the hand and its preshape relative to that affordance. Eventually, the synapses formed under this training become powerful enough that these now-established mirror neurons can fire on observation of an appropriate hand-state trajectory *even if it belongs to someone else* – thanks to the crucial encoding of trajectory relative to the object, not the actor.

Thus, the MNS model emphasizes recognizing *trajectories.* Moreover, simulations demonstrated how, as learning progresses, recognition of the grasp may occur earlier and earlier in the trajectory, an important

---

[3] The 1981 model, the first attempt to transport an earlier account of perceptual and motor schemas in frog visuomotor coordination to primate hand actions, was inspired by a 1979 presentation by Marc Jeannerod (published as Jeannerod & Biguer, 1982)

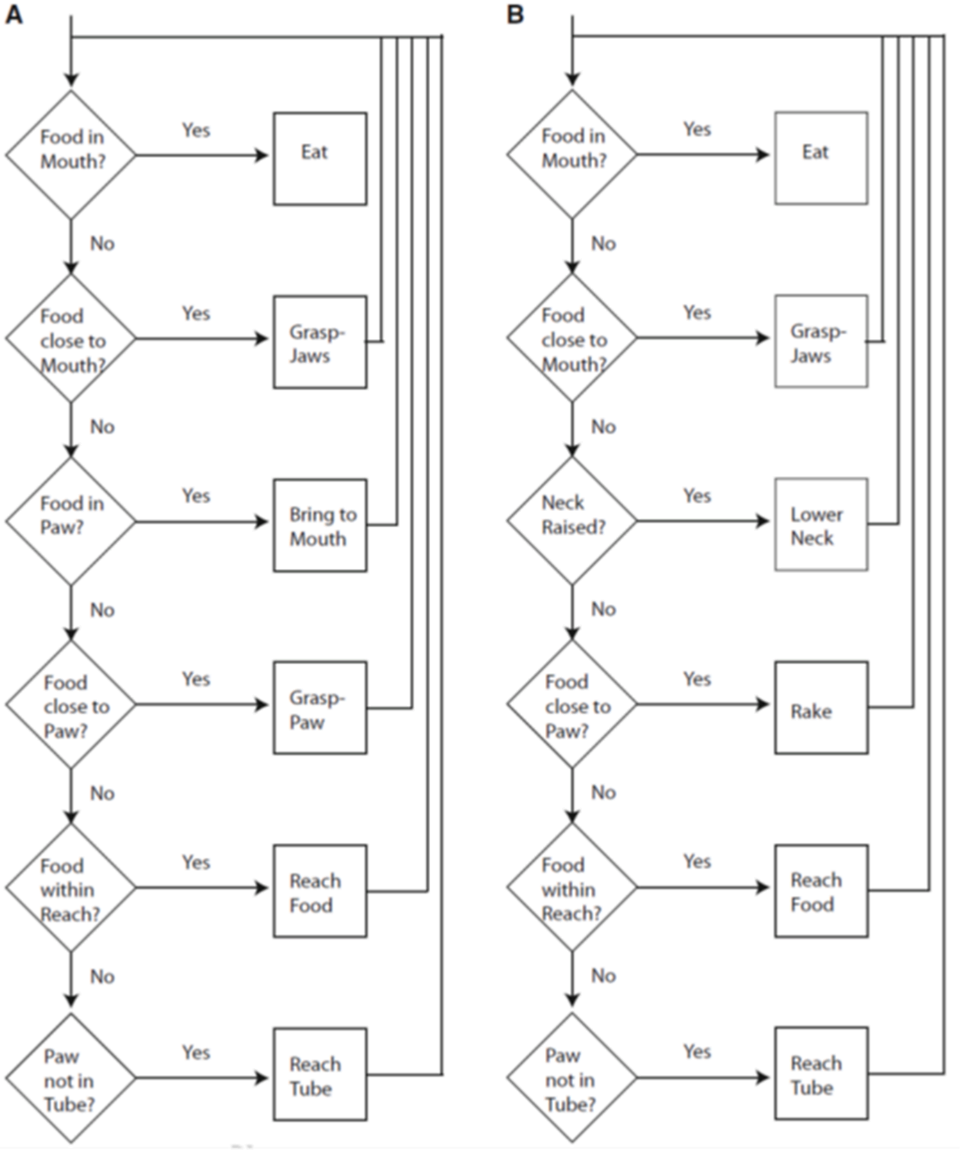

Fig. 2. A. The original ethogram for eating a piece of food initially in a horizontal tube. B. The ethogram that describes the behavior that is learned after the Grasp-Paw motor schema is lesioned (Bonaiuto & Arbib, 2010).

property for social interaction.

Turning from MNS to skill acquisition informed by observation of the actions of others, we note (as in developing nut cracking skill by young capuchin monkeys) that much practice is required to transform an overall pattern of movement into a skilled behavior. All of this requires individuals to be around conspecifics for some time, and to pay attention to them. However, aspects of an action will escape the novice's attention, as in positioning the nut in the pit in capuchins' cracking, or rotating the core in order to knap it.

6. Combining actions into behaviors

With this, we turn to overall behaviors, offering two approaches to "motor programs" that combine actions at the event level. Nonetheless, mastering the overall behavior will depend on fine-tuning the various components and the coordination between them.

6.1. *Opportunistic scheduling*

An ethogram summarizes field workers' observations of some related patterns of animal behavior, but may not represent "what goes on in the brain." How is the described skill learned? Byrne (2003) has characterized one way young apes may acquire a skill over many months as *imitation as behavior parsing:*

- certain subgoals become evident from repeated observation as being common to most performances, but
- detailed actions for achieving the subgoals are achieved through trial and error.[4]

By contrast, human mastery of language can support rapid learning of the "overall program" of a not-too-complex behavior – but the sensorimotor tuning of the actions and their transitions can still require months of practice, expanding the subtlety of the affordances to be

---

[4] Lind (2018) combined Pavlovian and instrumental conditioning in a sequence learning model used to simulate planning studies of apes saving tools for future use (Mulcahy & Call, 2006) and ravens planning for tool-use and bartering (Kabadayi & Osvath, 2017). Lind concluded that these studies of flexible planning in apes and corvids might be accounted for by associative learning.

recognized, and mastering the muscle control that converts affordance details into desired end-results.

Action level imitation (Byrne & Russon, 1998) can speed up the process of achieving a subgoal by copying some movement details of the manual actions another individual uses– without denying that, even here, much practice may be needed to fine-tune the skill. We present our variant, *complex imitation,* in §10.

Our *Opportunistic Scheduling* model (the Augmented Competitive Queuing model, ACQ, of Bonaiuto & Arbib, 2010) offers another approach to the flexible scheduling of actions (and these need not be manual) to achieve some overall goal. It offers a form of learning without imitation, and focuses on the opportunistic scheduling of known actions, not on how actions are added to the repertoire. Crucially, Opportunistic Scheduling incorporates a mirror system which, in particular, can monitor *self*-actions. When an *intended action* is unsuccessful, it may appear similar to an *unintended* action – and then the mirror neurons for the *apparent action* can serve a "what did I just do?" function, supporting recognition of when this unintended action is helpful in achieving positive reinforcement.

As a comparison point with the ethograms, we consider how Opportunistic Scheduling modeled the observation of a cat reaching to grab a piece of food from a glass tube and bringing it to its mouth. After a lesion blocked effective use of its ability to grasp, the cat developed a new behavior in just 4 or 5 trials,[5] namely to bat the food out of the tube (no grasp required) so it fell on the floor, and then grasped the food with its jaws and ate it (Alstermark, Lundberg, Norrsell, & Sybirska, 1981). Fig. 2 shows ethograms for before and after. Learning to go from one action to the other seems like a daunting task for the cat's brain, but the Opportunistic Scheduling model offers a plausible mechanism – so let's look at how it operates.

According to the MNS model, once the mirror neurons have been trained, the mirror neurons for recognizing *self-execution* of an action can be activated via *two* pathways: both via an efferent copy of the canonical code for the movement and via visual observation of the trajectory of the effector toward the affordance. The key insight is this: In general, these two pathways will activate the same mirror neurons – *but only if the action is successful.* Our Opportunistic Scheduling model describes what may be learned when these two inputs "disagree." In our cat example, the failed attempt to grasp an object may rake it onto the floor, in which case the apparent action, raking, comes to seem desirable for the present task.

In Opportunistic Scheduling, competition to determine the next action is based on a priority measure, but now this measure is updated each time an action is executed. Specifically, separate subsystems in Opportunistic Scheduling make two evaluations of each action:

Desirability depends on the current task or goal. Each time the action is performed, a measure of "expected reinforcement" is updated. This will be positive if the action leads "soon enough" to achievement of the goal but will be greater the shorter the time required to reach that goal.[6]

Executability depends on the availability of affordances (can the action be carried out now?) and the probability of the action's success.

At each time step, the priority of available actions is set by combining executability and desirability – the highest priority action will then be executed (or, since failure is possible, its execution will be attempted). Each time an action succeeds, its desirability is updated while executability may be left as is or increased. However, our model hypothesizes that when the action fails, *executability of the intended action* is reduced while *desirability of the apparent action* is adjusted. This process continues until the distal goal is attained. Note the importance of assessing how each action changes the internal and external environments.

Now the dramatic change in ethograms of the cat can be easily explained according to Opportunistic Scheduling: In just a few trials, the executability of grasping declines while the desirability of the apparent action of raking goes up. The rest follows.

7. Construction: Additive and subtractive

We now expand the P}A}E account to include notions of *additive and subtractive construction.* In §8, we extend these insights from hands to beaks to consider the building of bird nests as a prime example of additive construction. §9 then focuses on stone-tool making in the Oldowan and Acheulean traditions to provide a bridge between capuchin nut cracking (which is not a form of construction) and subtractive construction in the hominid line; with hafting then exemplifying the transition from subtractive to additive construction. Construction is distinguished from simple tool use in that it is aimed at the creation of a durable material configuration (a construct) through additive, subtractive, and/or transformative processes.

7.1. Construction: Additive, subtractive, and transformative

While the primary meaning of construction is *additive construction,* to produce a material thing by combination of parts, construction may require shaping of objects and other processes besides assembly, licensing the notion of *subtractive construction.*

Stone knapping is an example of subtractive construction, subtracting something from an object to bring it closer in form to a target object:

Oldowan manufacture (§9.1)*:* The success criterion is removing a "satisfactory" flake from the core, where "satisfactory" rests on the utility of the flake to serve as a tool. The flake is the construct.

Acheulean manufacture (§9.2)*:* The success criterion is sculpting the core into a tool, with repeated removal of a "satisfactory" flake from the core being a repeated subgoal, but where "satisfactory" now alludes to change in the core resulting from removal of the flake. The shaped core is the construct.

We turn to (proto)human additive construction in our discussion of hafting in §9.3, complementing the examples of bird nest construction in §8.

Transformative construction refers to processes, such as preparing food, or chemistry more generally (consider the use of glue), in which components may combine in ways such that they lose their discrete identities and/or assume radically different properties. Here, the transformation of the components is such that their original form is no longer discernible, and *disassembly* of the construct into its original pieces may no longer be feasible. Reversing the subtractive construction in stone knapping (reconstructing the original core from the debris) is also infeasible.

Weaving or sewing exemplify a method of construction that causes one object to follow a desired trajectory to result in its successful and stable embedding in the other. Similarly, in nest building, the goal of actions performed with the grasped material is not only to produce a mechanical effect on a target object/surface (it takes force to carry the twig, insert it, and to place it securely) but also to contribute to an enduring structure, of which the grasped object becomes a part.

The above trichotomy is not exhaustive: additive construction may include transformative processes like making glue and using it as an adhesive, or bending objects before they can become part of an assembly. Swallows add saliva to mud to glue mud bricks together (Jung, Jung, Lee, Kim, & Kim, 2021). It seems useful to call knapping a stone tool an example of subtractive construction but misleading to call nut-cracking an act of construction. Thus, in lieu of an exhaustive definition, we reiterate that "construction" is the process of operating upon an initial set of materials to create a novel object.

In addition to objects, the concept of construction can also be applied

---

[5] Contrast this rapid rescheduling of a few well-practiced individual actions with the extensive practice that may be required in learning to tune application of forces in relation to current affordances for a novel action.

[6] This involves the method of reinforcement learning known as temporal difference learning (Sutton, 1988).

to the combination of actions or mental processes. In particular, it may refer to the action of combining words according to the grammar of a language to convey an intended sense. §11 and §12 will offer a brief look at human languages from the perspective of *construction transferred from praxic to communicative actions*.

7.2. Working memory

To hold a distal goal in mind (whether in tooling, construction, or language) while working on a task requires some aspect of *working memory*. Such memory concerns, in part, various "items" relevant to the ongoing task, and thus may extend beyond the current focus of attention, with internal memory coordinated with the external memory provided by observable features of the current state of play. S*tigmergy*, whereby the trace left in the environment by each action stimulates the performance of a succeeding action by the same or different agent, may well describe the mound construction of termites, but while each termite's actions may create a new affordance for itself and others, but in mammalian or avian acts of construction there may be multiple affordances available so that the choice between them requires some form of "access" to the state of execution of the overall task.

For example, our opportunistic scheduling approach examines cases where the priority of the next action depends on desirability as well as executability (affordances), but that desirability is relative to current goals related to differing stages of an overall task While an affordance may attract attention as an "interrupt," more often it attracts attention only as relevant during a subtask. However, in overlearned behavior, each step may trigger a change in internal state that in turn triggers execution of a specific action without (even nonconscious) recomputation of desirability. Thus, a full analysis may demand that some ordering of subtasks must be learned rather than handled opportunistically. For example, for a certain kind of nest, the bird must first lay the foundations, weaving a strong connection onto a branch before weaving the egg chamber.[11]

Working memory need not be short-term memory, and one may speak of *long-term working memory* (Ericsson & Kintsch, 1995) which "holds" items for "easy retrieval" when an appropriate stage of a task is reached. In a complex task, we need to keep track of what subtasks have already been completed, and what may remain to be done. In humans, this may extend to a symbolic form of external memory, like a checklist. An important difference between humans and nonhuman animals is that humans can develop sustained attention and long-term working memory in ways that can be transferred to a wide range of novel skilled activities. Coolidge and Wynn (2005; Welshon, 2010; Wynn & Coolidge, 2010) are among those who have related working memory to human evolution.

Certainly, working memory and sustained engagement over the course of hours or days are essential for, e.g., nest building but (looking ahead to our later discussion of human evolution) we suggest that nonhuman animals (e.g., birds and monkeys) have little or no capacity – or need – to apply them in novel tasks in the way that humans can. We can exploit such memory in a great diversity of culturally determined situations, including the ongoing use of any tool once we have mastered it.

7.3. The question of the image in construction

As we analyze construction tasks, we become particularly concerned with tracking the object being constructed, not just the tool(s). We have spoken much about affordances – but we need to be more explicit about how each action changes the "target." We thus need to assess the possible role of "imagining the final form," whether for the overall process or a substage, returning to the issue in the explicit examples of subsequent sections.

As stressed in §2, the effect E of an action or behavior may include the "goal" of reaching a state in which the pattern of (multi-modal) perceptual schema activation meet certain criteria. We will use the term *image* for such a multi-modal schema expectation-structure, stressing that this is far more general than a particular visual image. The question for the section is this: at some stages of an activity, is an "image" held in working memory to explicitly guide activity, or is a course of behavior simply "ongoing" until it reaches a state that satisfied the generic conditions that constitute the image. This multisensory image of the expected result of the current stage of construction may elicit a sense of the transformation required to go from the current state to meet that motor expectation, and that "motor image" (Jeannerod & Decety, 1995; Johnson, 2000) may constrain the choice of motor schemas to effect that transformation, and their subsequent tuning to match the affordances that arise.

Consider Acheulean making of a hand axe. The assessment of what part of the core may "contain" the intended hand axe sets the goal for choosing a "proximal" affordance, a target for removal of the next flake. But what of bird nest construction? Which features of building a nest require sustained attention, and how much/how long? Building a particular nest may take hours or days. When a bird is distracted during nest building (as birds forage for food, conduct sexual displays, seek further material), how does it resume? One may hypothesize that it can recall the "stage" (in some sense) of construction, and on which part of the nest it was working – but beyond this general awareness and motivation, is the animal guided by the affordances of the partially constructed nest without recourse to working memory of the state of construction and the current "guiding image"?

8. Bird nest construction

Birds do not use tools to build nests (Collias, 1997; Hansell, 2000). Nonetheless, bird nest construction shares behavioral characteristics with tooling. Building a nest involves perceiving affordances of materials and substrates, working with a body-plus-object system, and managing dynamic spatial relations between a grasped object (with becculation replacing use of the hand) and the nest under construction. To our knowledge, the spatial relation between the grasped object that is put into the nest and the nest under construction has not been investigated.

8.1. The "image" in bird nest construction

In building a nest, the bird appears to have a general schematic for deciding (a) where to construct the nest, which, *depending on the species,* may be in a hole, on branches, on the ground, or hanging from a branch and (b) what to construct next. The suggestion is that the general form of each stage is species-specific but that the outcome at each stage will in part reflect the result of what has come before and the "imaged" result. That "image" may be very flexible, varying both with nature of the site and the available materials. Intraspecific variation in nests demonstrates that any innate "rough program" leaves many details open within the species-typical behavior. For example, it appears that the variation in the three warbler nests in Fig. 3 is due to variation in which materials are those most commonly available to the builder.

The techniques with which birds build their nests are various (see Fig. 4): they range from the sculpting of burrows, through the molding of mud or salivary mucus by vibrating head and/or shaping breast and feet movements or the piling up of materials, to the weaving of hanging nest baskets using intricate bill-made knots to fasten and secure grassy materials (Collias & Collias, 1964; Hansell, 2000).

The building process begins with nest-site selection and the appropriate choice of available materials from the environment, which individuals then manipulate and/or modify into the structure we call a nest. The builders may subsequently continue to modify that structure even when it contains eggs, chicks, or an incubating parent. Learning, defined here as a change in an individual's behavior in response to previous experience, plays an important role in birds' nest building.

For many types of nest, birds take materials of varied shape to be matched with various gaps in the nest, inserting each piece of material and weaving it into place in a way that contributes to the "desired form" of the nest. For this action A, we might say that the effect E is "move this

Fig. 3. These three warbler nests from Dartford exemplify the great variation possible within a species-typical structure. (Courtesy of Mike Hansell and the Hunterian collection, Glasgow.)

Fig. 4. A range of bird nests, from left to right: the leftmost is built in a hole, which may have been excavated; the next is a mud nest of a swallow for which the bird brings damp mud, adds saliva and then pastes this material onto the wall/growing structure; a grass cup nest where the grasses are bent and poked/pushed into the structure; a Cape weaver nest, which the bird makes by weaving fresh grasses into a variety of knots; and a sociable weaver nest, which may contain several hundred separate chambers, each built by a different male, again by weaving grasses. (Photos: Susan Healy).

stage of nest construction toward completion," and the P is then to find both a suitable insertion point and a piece of material whose insertion there would satisfy E.

Here the "unskilled" version of A might be "insert a twig securely into a (partially completed) nest" – but the skilled version has the more refined assessment of selecting an affordance-effectivity pair relevant to the appropriate stage of nest construction. This determines the set of positions that might work for the next twig. And so it goes until the bird decides that the current stage of nest construction is complete.

8.2. In search of species-specific images

Rather than positing that nest building is guided throughout by an image of the final nest, it may make more sense to divide construction into stages, with a separate "image" guiding each stage. What follows is a (perhaps too anthropomorphic) scenario designed to elicit new studies:

Hypothesis 1. There are stages in constructing a bird nest each with an innate "program sketch" and "completion criterion," but subject to learning to improve skill and change certain parameters. Crucially, and in terms of opportunistic scheduling, desirability is assessed relative to the current stage of a task. This points to the subtlety of the notion of *executability* for an action: it involves not only recognizing what objects offer affordances, but being able to assess the "quality" of an affordance learned on the basis of factors like the expected effort involved and quality of the result when one is chosen over another.

Hypothesis 2. The bird may not have an image for the overall nest when it starts, but each stage comes with a malleable (not necessarily visual) image that both guides activity during that stage and provides a criterion for transition from one stage to the next. For example, zebra finches switch from widening the nest as grass is added (the "cup") to narrowing it (the "roof").

The catch, of course, is that human observers may identify a visible structure and define an image but that does not mean that the bird itself employs it. How might we test the alternatives? There is mixed evidence that experience of the nest in which it was fledged could provide a bird with such an image. While Muth and Healy (2011)'s zebra finches did not prefer the color of the nest from which they had hatched when they chose material to build their first nest, Sargent (1965)'s birds did, if the material from the natal nest was not red– but this is a long way from producing the apparently species-specific differences exemplified in Fig. 4, without denying the variability-on-a-theme exemplified in Fig. 3. Similarly, while natal nest experience did not appear to affect nest morphology, all zebra finch males chose to build open nest-cups, while imprinting may impact decisions as to where birds build.

Perhaps more relevant is the case of white-browed sparrow weavers. Family groups build structures (mostly roosts, not nests) that differ between different groups even though new unrelated individuals join the group, suggesting that these new members must learn a new "image" from observing prior structures of the group they join. Nonetheless, this "image" seems to be a variation on the species-specific nest type.

How do birds select material? Fig. 3 shows that a bird will select materials that are in some sense "satisfactory" rather than matching a standard visual image. Can we extract from this some sense of a general notion of "image" that escapes the implication of visual particularity? Are other senses like touch also involved? This leads into a general notion of "material selection" that complements the notion of "image of stage completion."

We speculate that if an action is sufficiently desirable at a *particular* stage of a task, then material with a "poor" affordance may be chosen to execute the action when no "better" affordance is available. Experience helps. Multiple species are increasingly including human-made materials (possibly because their structural and/or their functional attributes replicate those of natural materials). Another subtlety is that while the state of construction may set desirability for the choice of a twig, the locus and method of inserting the twig may depend on properties of the twig.

8.3. Aspects of learning in bird nest construction

Birds in many species will build a nest that is at least moderately successful (in that they fledge their young) in their first year. The few data available show that becculating material as a juvenile increases the speed at which males use material as a nest-building adult (Breen et al., 2020), while adult weavers drop fewer pieces of grass as they build more nests through a season (Walsh, Hansell, Borello, & Healy, 2011). Breen, Sugasawa, and Healy (2021) discuss data on bird nest building and rodent food handling that suggest that expert motor skills can be acquired much sooner than the years it takes nonhuman primates. Birds reared in almost any environment will build a nest. This is in contrast to capuchins cracking nuts, humans making stone tools, and the mastery of a particular human language: these emerge only over a period of years and in a limited range of "cultural" milieux.

Breen, Guillette, and Healy (2016) show that learning occurs *within* diverse stages of nest construction, but there is little to contradict the view that the overall stages of construction are species-specific and so the overall "program" may be (in some sense) innate, even though the bird can become more skillful with repeated nest building. However, their article says very little about the detailed actions within each stage that provide the target for most of the learning that it describes.

When provided with access to nest material, juvenile weaverbirds improved at collecting material from plants (e.g., where to tear the leaf, in what direction, and how to perch in order to do so), and they become more proficient at building (by increasing the number of pieces woven within a 3.5 hour observation period by 26%). The nests of these builders, however, were characterized as "crude" in comparison to those built by experienced adults – thus, motor learning at the trajectory level, at least for weaverbirds, is important to nest building.

Selecting a nest site may involve general criteria that *appear* to be "innate," but that choice may be affected by observation of breeding success by other birds at other sites (Loukola, Seppänen, Krams, Torvinen, & Forsman., 2013; Seppänen & Forsman, 2007; Seppänen, Forsman, Mönkkönen, Krams, & Salmi, 2011). Note that this complements the details of nest construction with *a cognitive map of the territory* which includes actual and potential nest sites and sources of materials. The birds must have a cognitive map of the locations where these materials are located (as well as where to forage for food). Presumably, this does not specify a single cache but, rather, frames foraging for the next batch of materials – while also supporting repeated return to the nest site. What triggers leaving the site to forage; what triggers return to the nest?

Birds may even learn about nest building from watching the choices of other individuals, suggesting that the appearance of new nesting "traditions" (e.g., building in shrubs rather than on dry land) might emerge through social learning. Birds do learn from others what material to use (Guillette, Scott, & Healy, 2016) but only if the demonstrators are familiar. They can learn by watching videos of familiar birds building (Guillette & Healy, 2019) and they will even modify their material choice from just observing a completed nest (Breen, Bonneaud, Healy, & Guillette, 2019).

8.4. The Learning and construction of birdsong

As counterpoint for our concern below with the evolution of human language, we offer a brief section on birdsong. Intriguingly, various species of birds are more gifted than nonhuman primates in their ability in construction and/or their flexible and adaptable vocal control. The study of birdsong, both behavioral and in terms of brain mechanisms, has been pursued far more extensively than has nest construction. Like nest building with its variation from no nest building through to 'complex,' bird species can vary greatly from those that have no songs (although almost all have some kind of repertoire of calls with associated "meanings," as do chickens, some of which may be learned (Ten Cate, 2021)), via those with a limited repertoire of songs learned from the father, to species that are capable of learning novel songs via imitation (see Marler & Slabbekoorn, 2004, for an excellent albeit somewhat dated collection of articles).

Much is known about the structural differences in the brain, and even genetic correlates, that distinguish the brains of birds that are vocal learners from those that are not. Indeed, Cahill et al. (2021) offer data supporting the view that positive selection in noncoding genomic regions of vocal learning birds is associated with genes implicated in vocal learning and speech functions in humans. However, we stress that "speech" here is used in the sense of flexible vocal control that is open to learning, and must thus be distinguished from "spoken language." Petkov and Jarvis (2012) compared behavioral phenotypes and neurobiological substrates in birds and primates and relate their findings to spoken language origins. In addition, song learning has long been used as the model for language learning in humans. Nonetheless, birdsong cannot be broken down into *meaningful* "word-like" sequences and so, a fortiori, is subject to no grammar that can build up new meanings for sentences from the meanings of the words they combine (Ten Cate & Petkov, 2021). Indeed, there seem to be (pending future research) only two messages in birdsong: one is for males courting females with songs that affect their reward centers, and the other is for males defending their territory. Perhaps the best (though still partial) match is between birdsong and the *phonology* of human language (Yip, 2010), rather than with syntax or semantics.

In §§10-12, we distinguish language from speech. The evolution of human language has involved two complementary processes: the evolution of flexible vocal control and learning (something absent in extant nonhuman primates); and the evolution of the ability to create meaningful words (whether signed or spoken) and employ grammar to construct utterances with novel meanings. Note how different is the ability of birds to make small variations on a species-specific nest bauplan and the human ability to design and construct diverse novel objects, including the variety of structures in the architecture of the built environment (Arbib, 2021, §8.6). We will argue that it is manual skill shared with monkeys and apes – rather than vocal control *per se* – that lies at the heart of the emergence of protolanguages and then languages as forms of communication very different from the calls and gestures of other primates. We thus consider the neurobiology of sequential behavior in monkeys to be relevant to the search for an evolutionary basis for language in the last common ancestor (e.g., Wilson & Petkov, 2018) with new research focused on possibly stage-dependent emergence of hierarchical structure. Meanwhile, study of Fos immunoreactivity expression (Edwards et al., 2020) has confirmed a functional role for areas of the anterior motor pathway, social behavior network, and the cerebellum in nest material collection and manipulation by birds.

9. The Oldowan-Acheulean transition and on to assembly

We now step back from the full-fledged [sic] construction of bird nests to focus on the Oldowan and Acheulean protohuman traditions of stone tool making. As in capuchin nut cracking, the key tooling operation is the hammer blow. Let's see what we can learn from the differences:

Capuchin: The aim is to open the nut with one or a few blows, but there are no parametric conditions on the final form.

Oldowan (dating as far back as 2.6 Mya)**:** The aim of the hammer blow is to detach a sharp stone flake, suitable for use as a simple knife, from the core.

Acheulean (1.7-0.3 Mya)**:** Flake removal remains the basic operation, but now the focus is on the cumulative effect of flake removal in shaping variable cores into recurring tool forms characterized by a consistent set of desired properties.

On to Assembly: Only in the transition from Late Acheulean to Middle Stone Age technologies (0.5-0.3 Mya) has evidence been seen of additive construction in the form of compound tools in which, e.g., the shaped stone is hafted to a handle.

The Oldowan-Acheulean transition increases the length and complexity of the chain of actions in manufacture. What are the subgoals and completion criteria? In §10, we suggest that pedagogy may involve teaching skills relevant to mastering parts of the subchain, even without the motivation that only completing the chain could provide.

9.1. Oldowan flake production

Oldowan action sequencing is relatively simple and invariant: first rotate and inspect the core (repeating until a viable target is identified), then strike the target repeatedly with adjusted kinematics until a flake is detached or the target is abandoned as unsuitable. This "flake production chunk" (Stout, Chaminade, Apel, Shafti, & Faisal, 2021) is repeated until sufficient flakes have been produced or the raw material is exhausted. Crucially, this involves little contingency between successive flake removals and exerts minimal demands for explicit planning (Stout, Hecht, Khreisheh, Bradley, & Chaminade, 2015).

One may compare this with the capuchin nut cracking where the chunk is "strike the target repeatedly with adjusted kinematics until a nut is broken to make the meat accessible or the target is abandoned as unsuitable," with positioning the nut on the anvil replacing rotating the core. What is different is the increased Oldowan subtlety of recognizing and choosing an affordance and the related sophistication of mapping the kinematics of the blow to position, size, and shape of the affordance, as well as the incorporation into a longer behavioral chain of anticipatory tool production, transport, and subsequent use.

9.2. Acheulean shaping

By ~1.75 Mya, new "Early Acheulean" tool forms began to appear (Beyene et al., 2013; Lepre et al., 2011). Although there is debate over the biological, behavioral, and economic nature of this transition (Sánchez-Yustos, 2021), it is marked technologically by the invention and spread of shaped tools that archaeologists refer to as "hand axes," "picks," and "cleavers" (Stout, 2011). Here, unlike the Oldowan technology, the manufactured tool is what remains after flakes are removed. The core, rather than the flake, thus becomes the focus of construction. Removing flakes to shape the core requires greater perceptual-motor skill to precisely control stone fracture patterns *and* more complex action plans that relate individual flake removals to larger design goals such as shaping a pointed tip or continuous cutting edge. A key innovation of the Early Acheulean is the production of very large (>10cm) "flakes" from boulder cores, with this becoming the new core, the "blank," from which tools are fashioned (Semaw, Rogers, & Stout, 2009).

Production of consistent tool forms from variable raw materials in this fashion has long been held to indicate the presence of explicit design targets – the images that we have suggested may guide construction – and procedures in the minds of the makers (Gowlett, 1996). However, the degree of "imposed form" actually present in the early Acheulean remains controversial. A conservative interpretation is that early handaxe production was guided by a recurring set of functional, ergonomic, and possibly aesthetic design preferences (Wynn & Gowlett, 2018) with other elements free to vary in response to raw materials, use life, and random population (Kuhn, 2020; Lycett, Schillinger, Eren, von Cramon-Taubadel, & Mesoudi, 2016). This suggests a guiding image but not a very detailed one in the early Acheulean, with much of the form emerging from procedural, raw material, and functional constraints.

Later Acheulean technology, after about 700 Kya, is defined by the appearance of smaller, thinner, more regular, and symmetrical forms. Here, fairly specific imposed "images" guide handaxe shaping (García-Medrano, Ollé, Ashton, & Roberts, 2018; Shipton & White, 2020). Greater control over artifact form is achieved through more forceful and precise percussion (Pargeter, Kreisheh, Shea, & Stout, 2020). Preparation of core edges and surfaces (Schick & Toth, 1993) enables the prospective manipulation of core geometry in order to influence the size, shape, and location of subsequently detached flakes, and thus achieve challenging design goals such as thinning the cross section of the finished piece (Stout, Apel, Commander, & Roberts, 2014). What is distinctive in this period is the emergence of a *tool kit* that includes "soft" hammers (aka *billets*) made from bone or antler and the making of these tools for their role in making other tools.

By 500,000 years ago, these preparatory techniques were deployed to enable the systematic production of stone blades and points suitable for hafting (Wilkins & Chazan, 2012; Wilkins, Schoville, Brown, & Chazan, 2012) – more on hafting in the next subsection.

In hand axe production, these innovations would have: 1) increased learning demands for perceptual-motor skill acquisition (Pargeter et al., 2020), possibly including social support and teaching (Pargeter, Khreisheh, & Stout, 2019); 2) further extended operational chains to include the manufacture (cutting and shaping) and curation of bone or antler hammers; and 3) elaborated the basic Oldowan-style flake production chunk by allowing the simple striking action to be expanded to a complex preparation and percussion chunk in which the current choice of affordance and consequent flake removal is not an end in itself but is dependent on a more distal goal. The goal may to some extent be stage-dependent, with the distal goal (the final shaping of the tool) itself emerging with increasing precision as the work nears completion. This has important implications for the technological pedagogy hypothesis of §10.

Fig. 6 shows 3 stages: Stage 1 involves obtaining the material. By the end of Stage 3 the final form for an Acheulean hand axe form has been achieved. By way of comparison, we suggest:

1 When capuchin monkeys crack nuts, they have no detailed image of how the nut will look or feel after it is broken, only the goal that the meat will be accessible.
2 Oldowan knapping requires forceful and accurate percussion combined with a few simple geometric criteria (e.g., strike near acute edges) and/or procedural habits (e.g., rotate core between blows) to produce numerous useful flakes to choose from while maintaining viable core geometry for further flaking. Criteria for flake utility depend on experience with additional skills for which flakes were used, such as butchery.

The above are both ILGA-like (integrated learning of grasps and affordances, §5) in that the parametrization of the blow in relation to the matching affordance could be gained by trial and practice with different reinforcement schedules provided by a brain that has learned what (1) a satisfactorily cracked nut looks like (capuchin) or (2) what a satisfactory flake can look like (Oldowan). However, the Oldowan skill may not involve an image of the precise form of the intended flake because the reinforcement is based solely on learning what striking position (affordance) and velocity to choose to produce larger more useful flakes, rather than guiding actions on the basis of intended (even though adjustable) expectations on size and shape of what will be produced.

3 Making an Acheulean hand axe required some representation of the desired effects of knapping the core. The nature and detail of such representations would likely vary over the production process, extending from fairly detailed and specific images of immediate subgoals such as establishing an acute angle or removing a surface convexity to loose estimates of the size and orientation of the intended piece and general standards of regularity or symmetry against which emerging products might be compared. Knapping actions must be organized relative to these goals and the choice of affordances for each blow based in part on this *imagined shape*, rather than solely on the judgment that a place on the cobble offers a suitable affordance for striking off a single flake.

Recall our P}A}E notation. Here, P may include the availability of a tool and suitable material as well as the affordances of the acted-upon object, while E may omit "irrelevant" information, such as the disposition of fragments around the knapping site. In the Acheulean case, P may at some stages be responsive to an E that reflects new goal-parameters focused on the *imagined* form of the core. This seems to require holding a plan in working memory that includes visual constraints (perhaps initially vague). Caution: Speaking of imagination and memory here does not imply that the capabilities involved in early Acheulean manufacture are qualitatively the same as those of modern humans. (Similarly, we will distinguish various forms of imitation in §10.)

9.3. On to assembly: Hafting as additive construction

We have contrasted the "Oldowan act" with the Acheulean case where the distal target is, e.g., the imagined hand axe rather than the current flake. We now briefly consider the further skills required to transform multiple objects into a new construct. A spear, for example, may be assembled from "functional components" (e.g., spearhead and shaft) and "affixing components" such as a strap or glue (Wadley, Hodgskiss, & Grant, 2009) to hold elements together.

Details will vary across particular technologies, but in general the move to additive construction of multi-component artifacts will further extend the complexity and temporal duration of production sequences and expand the breadth of material properties and affordances that must be mastered. The neural representation of technologically relevant characteristics such as sharpness, malleability, or durability may (but might not) implicate certain (perhaps limited) semantic and analogical processes (Brand, Mesoudi, & Smaldino, 2021).

By 200 Kya, with the emergence of *Homo sapiens*, the technology was in place to create a variety of tools, since working the rough blank could follow an "image" to ultimately become either a cutting tool, a serrated tool, a flake blade, or a scraper, etc. Once we come to the Upper Paleolithic (50 to 10 Kya), there are beads, tooth necklaces, cave paintings, stone carvings, and figurines. Each innovation opened the way for further innovation in concert with those available before.

Additive construction emerged by the late Acheulean, but whenever such additive construction arose (probably in multiple places), it required a new understanding of components that serve for assembly rather than having a direct functional use in the end product – whether these remain visible (e.g., leather thongs binding other pieces together) or are subtler, like glue. Prepared core technology (making one thing as a resource for making something else), the controlled use of fire, and the making of blades all played a crucial role. Note that assembly of components need not involve tools, even though tool use may have been involved in forming the components.

Cutting, scraping, sewing, adding, piercing and digging actions are all found among recent hunter gatherers (Oswalt, 1976) and we assume that Middle Pleistocene humans structured many of their technologies around these basic activities. Barham (2013, p.154) outlines the types of action the artisan will need to consider in making a hafted tool: the haft design, the properties of the insert and its edge angles in relation to the stresses on the tool as a whole; and the use of a binding or an adhesive to hold the pieces together. The use of binders and adhesives obliges the artisan to know how the components work together as a whole, and to be able to adjust them in response to the realities of daily tool use (Barham, pp.193-4). Barham suggests that as different combinations of handles, hafts, bits, and bindings become available, they could support a *combinatorial principle* for developing new tools to meet specific needs – but we stress that having skills for forming particular combinations does *not* guarantee having a general concept of forming combinations. Nonetheless, experience with these new tools could support a cognitive innovation – the conscious understanding that combining not only parts but even subassemblies of current *and future* tools into new additive constructs could yield new tools. This cognitive innovation of being able to plan for new combinations rather than "discovering" each particular combination by happenstance stands in contrast with the very slow accumulation of tool innovations during the early Acheulean.

A single individual can make a wooden spear and gather materials accordingly, but gathering and processing all the materials required for a stone-tipped wooden spear with adhesive mastic might involve multiple individuals and occur over months of time during logistic procurement trips. This takes us beyond the realm of immediate action planning to long-term prospection and the distributed scaffolding provided by social organization and patterned practices. We see that extended cooperation among those with diverse skills may have been a crucial prerequisite in laying the basis for the pedagogy going beyond the learning by observation based on trial-and-error without caregiver guidance. Indeed, the teacher-apprentice relationship requires a broader context in which people have learned to cooperate over an extended period of time to achieve shared goals. However, some birds do build nests together in a process which may take days or weeks. The comparative study of different species reveals a diversity of body shapes, neural architectures and behaviors – and thus what is an innovation in the hominin line may have been mastered by other means (and in varied forms) in the evolution of other species.

10. Skill acquisition and pedagogy

Until now, we have focused on practical skills that directly change physical structures in the world. Here, we begin the swerve to *communicative* actions. All animals communicate, but humans have distinctive communicative *systems* called *languages.* As human praxic skills became more complex, they became harder and harder to learn through trial-and-error, even if shaped by observation of practitioners. This suggests that the increasing richness of (proto)human skill went hand in hand (literally) with increasing subtlety of demonstration and pantomime to support teaching those skills, and that such increases in pedagogical skill provided a key support (but not the only support) for the eventual emergence of languages.

We suggest that language was not needed for the transfer of Oldowan skills, but that over the long course of the Acheulean there was a virtuous circle (Morgan et al., 2015) between increasing complexity of technology and the emergence of forms of communication intermediate between the calls and gestures of apes and the subtleties of language – we call these forms *protolanguages.*[7] In §11, we introduce the notions of grammar and lexicon in the structure of modern languages and look at ways in which one might characterize a "grammar" of action in the absence of language. This sets the stage for our analysis of the development of modern languages as flexible systems for the *additive construction* and *communication* of meanings. As a key to the transition to the evolution of language, we explore how pedagogy may enrich praxic skills (§10.1), and (§12) how protolanguage and then language could extend the range of pedagogy.

10.1. The technological pedagogy hypothesis

Stone tools provide our earliest and highest resolution evidence for the evolution of the human technological niche. We thus focus discussion here on stone-tool making ("knapping") while noting that study of other Paleolithic skills – such as pyrotechnology, hunting and butchery, tool-making in non-lithic materials, and making pigments and beads (e.g. d'Errico & Stringer, 2011) – could enrich our discussion. If all humans did was to make stone tools, that might add little demand to what can be achieved by a protohuman brain. What is distinctive is the ability to

---

[7] In historical linguistics, the term *protolanguage* refers instead to a full language that is ancestral to a range of later languages, as in "Latin is a protolanguage for the Romance languages" or in the attempt to reconstruct a protolanguage for a family of more-or-less related languages, like the Indo-European languages, for which no shared ancestral language is known.

master multiple skills within a social group – so that as specific social roles emerge, they require extended apprenticeship without losing the ability to master a range of skills, practical and social, shared by the community at large.

The present focus on stone tools is linked to an actual archaeological record, but our speculations on the evolution of language engage with the broader range of (proto)human activities.

Learning to knap can take hundreds of hours for modern humans, even for relatively simple Paleolithic technologies and with supportive teaching conditions (Pargeter et al., 2019; Suddendorf, Brinums, & Imuta, 2016). Modern humans support practice through active teaching and motivation in structured learning environments (Stout, 2002; Stout & Hecht, 2017) but it is controversial when such supports first emerged (Gärdenfors & Högberg, 2017; Stout, Rogers, Jaeggi, & Semaw, 2019; Tennie, Premo, Braun, & McPherron, 2017). There is some experimental evidence that verbal instruction enhances knapping skill acquisition in modern humans and thus *might* have provided selective pressure for language evolution . However, this does not indicate that language *in any extended sense* was required in Acheulean times. Rather, this would fit in with the notion that simple protolanguages (see §12) provided essential stepping-stones for the *gradual* emergence of extended languages. In transferring the skills of Acheulean manufacture, only a limited repertoire of vocal and manual gestures may have been required.

These observations set the stage for the *technological pedagogy hypothesis* (Stout & Chaminade, 2012): namely, that a protohuman niche of increasingly subtle tool use and manufacture (but, we argue, not just these) might have supported the evolution of communication systems more flexible than those of other primates, exerting interacting pressures on individual skill, social learning, and intentional communication that drove the biocultural evolution of language. Stout (Stout & Hecht, 2017; Stout & Khreisheh, 2015) argues that the evolutionary context for this process was an emerging human *technological niche* in which a focus on high-value, difficult-to-acquire food resources provided the surplus nutrition needed to support extended growth, reduced mortality, and accelerated reproduction (Kaplan, Gurven, Winking, Hooper, & Stieglitz, 2010). Such life history effects in turn allowed the protracted learning, extended lifespan, and increased brain size that enabled further surplus production through the discovery and inter-generational reproduction of increasingly effective technological skills, knowledge, and equipment. This virtuous feedback cycle (Isler & Van Schaik, 2014) would have strongly *favored and been favored by* the evolution of enhanced action recognition, control, and imitation as well as the evolutionary, developmental, and/or behavioral repurposing of their neural substrates to support intentional communication for cooperation and teaching (Stout & Hecht, 2017), with broad evolutionary-developmental effects on social cognition.

One hypothesis is that the emergence of *technological pedagogy* might have coopted the brain's structured action sequencing capacities to yield linguistic syntax (Kolodny & Edelman, 2018; Morgan et al., 2015; Stout, 2018; Stout & Chaminade, 2012). The *mirror system hypothesis* (§12) suggests that the path from manual skill to language was less direct, requiring the emergence of new capacities beyond action sequencing alone, spelling out how distinctive skills in imitation might have supported the emergence of *protolanguages* which served as stepping stones to languages even though, lacking a "modern" syntax, each was not yet a language. Such an extension of neural capacities is consistent with a view of the technological pedagogy hypothesis that addresses these new capacities, while suggesting that protolanguage would serve many aspects of social coordination in addition to pedagogy.

10.2. *Imitation, pantomime, and pedagogy*

As tool complexity increases, teaching by demonstration and correcting mistakes offers a more effective way of learning than observation alone, while having the facility for, at least, protolanguage would be particularly useful for explaining the actions, affordances and goals involved in making a novel object.

While Oldowan technology could be mastered by trial and error and learning by imitation from observation, attaining this skill would take perhaps years (compare the timetable for young capuchin monkeys to attain the less-demanding skill of nut-cracking). There has been no systematic study of time to mastery in modern humans in the absence of instruction. The longest study to date found no evidence of improvement in the first two hours of practice. However, intentional instruction (not present in capuchin groups) has been shown to be beneficial for present-day learning of Oldowan-like flake production (Morgan et al., 2015; Pargeter, Liu, Kilgore, Majoe, & Stout, 2023). Acheulean technology involves quite complex and demanding techniques (Stout et al., 2014), some of which modern humans have difficulty conveying without explicit verbal instruction (Lombao, Guardiola, & Mosquera, 2017). However, our concern here is not that language is currently helpful, but rather to trace how that help might have emerged. The observation that language can aid teaching of ancient skills provides support for hypothesized selective pressures (Morgan et al. 2015) and invites us to consider a manual-action-friendly account of that emergence. We do this by first looking at how pantomime may bridge the relation between imitation and pedagogy.

We have distinguished two levels of skill acquisition: learning the high-level "program" and learning individual actions (which may involve "getting a first approximation" and then "fine-tuning"). Both may involve pedagogy scaffolded by demonstration, but the former may have benefited from increasing use of (proto)language. Nonetheless, fine tuning still requires extended practice.

We now consider a variety of forms of imitation and briefly relate them to pedagogy. Rather than "on-line imitation" in which one immediately passes from the recognition of a familiar action to its execution ("automatic imitation"), we focus on *learning by imitation* which comes in different flavors that share the property that the observer comes to add a new skill to their repertoire through (possibly/probably) repeated observation of the execution of that skill by others. These processes involve working memory, but then – somehow – the overall "skill-program" must be transferred to long term procedural memory. Further observation and practice lead both to refinement of the overall program and increasing tuning/automatization of the constituent actions, as well as possible generalization of the applicability of the behavior. Irrespective of any role for a teacher, organizing a novel action or subroutine in relation to a specific goal offers a crucial distinction between:

- Observing that an action already in one's own repertoire achieves a goal for another, and then using that action to achieve such a goal for oneself – this is like having P,A,E in one's repertoire and extending it to P',A,E' for the same A, learning that a new goal (part of E') can be achieved with A if the preconditions (within P') are right.
- Acquiring a novel action or behavior. This involves acquiring sensorimotor coordinations that can be tuned only through repeated practice.

Imitative matching of *shared action repertoires* may constitute a basic unit for the social transmission of behavior (Miller & Dollard, 1941). We hypothesize that our evolving skill in imitation rested in part on the ability not only to acquire novel actions through trial-and-error and through the assemblage of known actions, but in addition the ability to recognize that the observed action was "somewhat like" an action already in the observer's repertoire.

We saw that Byrne (2003) argued that apes acquire new skills through "imitation as behavior parsing." Here, as it were, inferring the Es (the various subgoals) guides the subsequent development of an action to get from one to the next through trial and error. We call this *simple imitation*. Having mirror neurons (or recognition of actions *already in one's repertoire*) can speed this process, but further mechanisms are required if one is to learn to recognize an action *outside* one's own repertoire and then use that recognition to drive adding similar actions to that repertoire.

A crucial expansion of imitative ability yields a blend of *complex action recognition* as well as *complex imitation* – this skill "parses" the behavior into constituent actions, but adds attention to the motion as well as the goals of subactions, with the consequent ability to achieve a first approximation to that motion with little or no trial and error.

- *Complex action recognition:* The ability to recognize when another's performance combines actions which are, or can be approximated by (i.e., more or less crudely imitated by) variants of actions already in the repertoire. Moreover, this now includes the "how," i.e., relating some parametric details of the observed P and E for each (or, initially, some of) the actions.
- *Complex imitation:* The ability to use this analysis to imitate the other's performance with more or less accuracy, though much practice is still required for mastery.

More subtly, complex imitation must extend beyond rote imitation of a fixed sequence if it is to be generally useful. As noted earlier, a skill generally requires the flexibility to adapt to available and changing affordances. Hence, as was apparent even for capuchin nut cracking, mastery interleaves chunks whose order is relatively automatized with portions that (§6.1) are scheduled opportunistically. We thus see something of the necessity for hierarchical structure in behavior and its successful imitation, but the range of behaviors that can be mastered at the event level will depend on the depth of hierarchical structure and the range of sub-behaviors that can be captured in working memory.

There are key differences between performing a skill (but allowing others to observe it), demonstrating a skill, and pantomiming that skill:

In *demonstrating* a skill, one is still performing the skill but now with three key differences: (1) one positions oneself to make it easier for the observer to observe the performance; (2) one slows down the performance while exaggerating movements for each action to emphasize the transitions that make behavior parsing possible and, in particular, aids the accuracy of complex action recognition; and (3) draws attention to the affordances for each action. In demonstrating the skill, the performer is still acting upon the object, so that different actions are shaped to conform with the physical affordances.

In *pantomime*, by contrast, the movements are based on those of actual actions but are no longer performed upon actual objects and thus at best approximate the shaping that conforms object-directed actions to affordances. A pantomime can thus remind students about the general motion to perform next and may be augmented by interleaved pointing to specific affordances, but must build upon prior demonstration to instill which affordances each action must align with in practice.

In either case, successful pedagogy aids the student's talent for complex imitation by the skill of the teacher in slowing down the actions and emphasizing the transitions between them. This helps the student to break a performance into components even when these components are themselves still unfamiliar. These complementary abilities of student and teacher can dramatically increase the diversity of skills that an individual can master, and thus support a form of cultural ratcheting in which the chunks that can provide units with a particular process of complex imitation can become more and more complex.

Moreover, a pantomimed movement may indicate more than just an action. For example, if I pantomime drinking from a mug, the initial shape of my hand may indicate the mug while the final twist of the hand near the mouth may indicate the intention to drink, but this is achieved without the actual motion within the pantomime having separate components for "mug," "mouth," or "drink." This "communicative paradox" will play a key role in shaping our §12 discussion of the transition from protolanguage to language. Here, though, we must note that this same pantomime could be used both in training a child to use a mug and in suggesting the child drink from his mug, or in requesting that someone give one something to drink – a divergence between cases in pedagogy where it is essential that a pantomime gives a good sense of the shape and extent of a movement, and cases where it is used simply to make a request or some other communicative act.

What of the P}A}E framework in the transition from action to pantomime? We lose the careful matching of the kinematics to the current task and affordances, but we gain the use of the movement as a symbolic, communicative tool. As the examples suggest, the same pantomime can have different communicative goals (and thus Es) whereas the physical constraints on emitting the message are few – the hands must be free, and the recipient's attention must be engaged. However, with increasing social complexity would come increasing complexification of P to include a sense of social propriety as well as estimation of the state of mind of the other.

In some sense, action is *synthetic* – we put together diverse actions into an overall behavior. However, as we gain a skill, the separation between actions begins to disappear, as we master the smooth transition from one action to the next. This is what challenges complex action recognition, which has to be *analytic* – trying to recognize the constituent actions, even though they are not clearly separated. This is hard enough when the observed behavior combines known and mastered actions. However, it becomes much harder when trying to learn a new skill through observation, where the constituent actions may themselves be novel and thus harder to recognize in a form that can guide mastering that action for oneself.

Zukow-Goldring (2012) has shown that (modern) mothers can teach diverse skills to children too young to understand spoken instructions by using *assisted* imitation, based on demonstration and then pantomime of the constituent actions plus means for drawing attention to affordances of the objects salient for the skill. Even when the mother assists imitation before the child has language, she uses various words of encouragement and various descriptions as teaching proceeds. We may say that she exhibits "co-gesture speech," complementing the normal use of cospeech gestures in speaking children and adults.

Recalling our discussion of "two pathways for affordances" (Fig. 1), we suggest that the Oldowan-Acheulean transition involves complementary changes in both skill and pedagogy: a "dorsal" increase in mastering the parametric adjustment of motor schemas to affordance details (dexterity) and an increasing "ventral" ability to create novel assemblages that "put the actions together." This ventral path may be the key to the entwinement of action and language in pedagogy. The transmission of complexly organized technologies typically requires fidelity at both lower (e.g., the sensorimotor coordination for successful execution of embodied skills) and higher (e.g., overall "program" linking actions and goals) levels of action organization. The latter is likely to implicate the evolution of mechanisms in prefrontal cortex that support more abstract goal representation; and note, too, the possible importance of tying action elements to abstracted communicative signs that then become stable units for mental manipulation (Brand et al., 2021; Stout, 2018). We would add signs for objects, affordances and goals, complementing mechanisms for more precisely matching the details of affordances to observed movements in actions that can be linked to certain (sub)goals.

Exploring the evolutionary implications of this, we may, extending the technological pedagogy hypothesis, hypothesize that not just tool making but diverse other activities of early human groups such as making a fire, child care, foraging, hunting, and much more may have provided many protolanguage fragments. We must nonetheless avoid over-emphasizing the role of pedagogy in the evolution of language because humans are (and protohumans, presumably, were) social creatures who exhibit a variety of behaviors in which coordination is important. In Acheulean toolmaking, the work might extend over hours or even days, and involve more than one person. In coordinating their behavior, the tool maker might request an assistant to go and find a new core or produce a new antler billet. Here, the request cannot refer to actions or affordances for objects that are present; their very nature requires specifying the object as well as the action. We see the beginnings of *displacement* – communication beyond the here-and-now and here-and-next.

11. Grammars for language and action

We need to briefly review how grammar enables modern languages to construct new meanings and assess what implications this might have for the idea of a "grammar" of action – positioning us to better assess the gap that had to be bridged by biocultural evolution.

11.1. Introducing construction grammar

In linguistics, a *grammar* offers an approach to describing human languages, not a grammar for describing physical acts of construction:

Each human language (whether English or Hindi or Warlpiri) combines an open lexicon (set of words) with a grammar comprised of diverse *constructions* (in the linguists' sense) that combine words hierarchically to generate an endless array of novel, shareable meanings. The earliest protolanguages, we presume, had small lexicons (whose elements may be called *protowords*) and none or very few constructions; but, as tens of millennia went by, not only did lexicons increase in size, but so did the number and generality of the constructions (with consequent restructuring of the lexicon). There is no hard and fast boundary here between complex protolanguages and simple languages.

Language production integrates two processes of construction: constructing a sequence of word-forms (constrained by hierarchical application of constructions), and constructing a meaning, in general with the intention that uttering those words to another person will convey (more or less accurately) that meaning to them. That is, language use must satisfy the *parity principle* that the meaning understood by the hearer resembles that intended by the speaker. (This may involve shared knowledge and shared context.)

In much of so-called generative linguistics (e.g., Chomsky, 1995), the emphasis has been on generating sequences of words that constitute a "well-formed" sentence and on testing whether a sequence of words is syntactically well-formed, without consideration of meaning. By contrast, we hold that language evolved to convey meaning, and thus we locate ourselves in the context of *construction grammar* (Croft, 2001; Goldberg, 1995).[8] Here rules combine syntax *and semantics* so that a construction may look like: "Given a phrase A with meaning m(A) and a phrase B with meaning m(B), then the new combination f(A,B) will have a meaning m(f,A,B) that depends on m(A), m(B) and the way, f, by which they were combined."

Unlike a rule defined in terms of syntactic categories like noun phrase, noun, and adjective, a construction may impose semantic restrictions on the slot fillers. For example, one construction whose form has f(A,B) = "A in B" requires that A describe a person while B names a color to yield a phrase meaning that the person is wearing clothes of that color, as in "the woman in red." §12 explores the notion that, as languages evolved, some "slot fillers" remained semantically defined as in this example, whereas others went from having highly limited semantic variation to becoming abstract enough to form syntactic categories.

We must also address the earlier observation that for the monkey and protohuman skills discussed earlier it requires immense effort to master each of a limited set of skills, whereas language users can readily construct hundreds of utterances a day, many of them novel. Of course, acquiring such flexibility in using language requires years of practice and social interaction. (In contrast to the above "culturally acquired" skills, nest building doesn't seem to require extended learning by birds.)

11.2. The syntax and semantics of action?

A key tenet of the technological pedagogy hypothesis is that communicative acts are required for technological systems of increasing complexity. Of course, tool-making actions are meaningful in the sense of being goal-directed but are not symbolically referential in the way that words are. Instrumental and communicative goals may require organization and manipulation of very different representations (e.g. simulation-based physical inferences vs. semantic associations and even theory of mind). We are more likely to find overlap between technological skill and language in the basic cognitive processes involved in learning, rapidly recognizing, and skillfully executing complex sequential structures across domains. Capacities such as statistical learning, predictive processing, and inhibition could provide an evolutionary and developmental foundation for the acquisition of a broad range of culturally evolved instrumental and communicative skills (Stout & Hecht, 2017). Nonetheless, whatever mechanisms are shared, multiple, modality- or task-specific systems are also engaged.

To establish a more direct connection with the archaeological record in relating skill and language, Stout et al. (2021) conducted 17 naturalistic stone-tool making experiments (9 Oldowan and 8 Acheulean). In each instance, a skilled modern human worked a piece of flint until it was either completely exhausted (Oldowan) or successfully shaped into a refined hand axe (Acheulean). They coded the behavior sequences using a simple ethogram of 7 event types encompassing the elementary body movements and object transformations of stone knapping. To analyze these sequences they identified 8 actions, and formalized each of the 17 instances as a string of letters corresponding to the actions in the order in which they were performed.

As, in part, a bridge between stone tool manufacture and language evolution, they developed different *purely syntactic* "grammars" for the two data sets. They applied two established sequence learning algorithms, k-Sequitur context-free grammar inference and Hidden Markov Models, to the coded event sequences. Without going into details, let's note that the former exploits rules for organizing strings of letters in the style of Chomsky's earliest semantics-free formalization of grammar (Chomsky, 1956) while the latter generates strings by stochastic transitions between a finite set of states. Each provides a formal and objective quantification of sequential structure irrespective of actual content. Both succeeded in revealing information about the intentional structure of stone tool-making by identifying the characteristic sequence of actions used for platform preparation of an Acheulean core for thinning (an operation absent in Oldowan flake removal) and, in agreement with qualitative archaeological assessments (Stout et al., 2014), identifying this operation as central to the greater complexity of Acheulean action sequences.

Action "grammars" extracted in this way describe observable regularities of behavior rather than the processes that generate those regularities – just as Chomskian linguistics is different from psycholinguistics or neurolinguistics. Yet, while the symbol sequences submitted to the above algorithms contain no information about goals or affordances, the grammars they define do encode recurring situations that led the knapper to select a particular action. They thus provide a data-driven way to study the patterns emerging from these processes that: 1) derive structure rather than imposing it, 2) respect the real variability underlying ideal characterizations, and 3) enable objective quantification of complexity.

This quantification of complexity can thus provide a foundation for testing the neural demands of action understanding, much as complexity metrics have been used to test explicit neurocognitive hypotheses of language processing (Henderson, Choi, Lowder, & Ferreira, 2016; Nelson et al., 2017) even if the metrics themselves to not contain the semantic information that is of central importance to actual comprehension. By applying grammar extraction methods to fMRI data from a previous tool-making observation study (Stout, Passingham, Frith, Apel, & Chaminade, 2011), Stout et al. (2018) were able to identify regions whose activation correlated with sequence complexity. However, this correlational approach does not demonstrate what aspect of behavior complexity is proximally responsible for the increased brain response. It could be that "working toward an image" is the crucial cognitive factor

---

[8] There are many different approaches to formalizing construction grammar (Arbib, Gasser, & Barrès, 2014, §4, briefly surveys a few) but their differences are irrelevant here.

that distinguishes the Acheulean from the Oldowan brain data.

12. Language emerging: The Mirror System Hypothesis

We now introduce the *mirror system hypothesis* (MSH) for the emergence of language, hypothesizing how increasing skill in imitation might have combined with pantomime to support processes of conventionalization that yielded *protolanguages* that served as stepping-stones to languages even though, lacking many constructions, each was not yet a language.

MSH hypothesizes changes from LCA-m (last common ancestor with monkeys) via LCA-c (last common ancestor with chimpanzees) to Homo sapiens that may have initiated the path via pantomime and protolanguages to languages that made modern human pedagogy possible. Data supporting the hypothesis (see, e.g., Arbib, 2012, 2016; Arbib & Rizzolatti, 1997; Rizzolatti & Arbib, 1998) include analysis of vocal and manual gesture in apes, monkeys and humans, analysis of sign languages of the Deaf, and data on pidgins and creoles. Here, though, we focus on those features of MSH most relevant to our concern with linking manipulation and additive and subtractive construction with emergence of language. Here are the first two stages posited in MSH:

MSH1. LCA-m had manual dexterity, an ability to recognize the actions of conspecifics (supported in part by mirror neurons), and a limited ability to learn new manual skills, but little in the way of imitation. Communication involved a limited set of species-specific calls and facial gestures.

MSH2. LCA-c preserved these LCA-m abilities, but also had a *simple imitation system* for manual actions. A small innate repertoire of manual gestures could be augmented by novel gestures developed by dyads or within a group to mediate communication.

Building on these capacities of LCA-m and LCA-c, MSH posits that biological and cultural evolution yielded the following in the transition from LCA-c to early humans:

MSH3. *Complex action recognition,* recognition of the actions employed within a behavior and the way they fit together in achieving a desired result, emerged to support coordinating behavior generally and to support *complex imitation* of observed skills, based on trying to replicate not only achievement of a goal but also some details of the movements used by another to achieve it. (Recall §10.2.)

MSH4. *"Ad hoc" Pantomime* emerged to convey to others the need for an object or some associated behavior on objects (whether present or absent). This involved a crucial change from *transitive* actions (actions upon objects, guided by the object's affordances) to *intransitive* actions (actions conducted in the absence of objects).

With this, we come to *a crucial bifurcation* (Arbib, 2023):

- *As an extension of demonstration in pedagogy*, the pantomime is an intransitive capture of the motions of a skill, while the affordances for the corresponding transitive action (upon an object) is available to the apprentice. Here, the pantomimed motions can convey to the student the trajectory and force that is to be applied in acting upon the available object(s). For pantomime-as-demonstration to be effective, all that is required is that the pantomime be understood by the student within the current context.
- By contrast, in a *pantomime-for-request*, such as "get me another core," there are two notable differences – the details of motion during the pantomime are not important so long as they are distinctive; and the pantomime may now be indicating an object rather than an action or sequence of actions. This underwrites the role of *pantomime as an increasingly powerful means of communication*. Indeed, if *conventionalized* pantomimes are to serve as precursors to a group's emerging system of communication, then the pantomime or its successors must eventually have the parity (symmetry) property –when emitted by a member of the group, its intended meaning must be recognizable by many others in the group, at least within context.

MSH then views conventionalized protosigns as the first of several bridges:

MSH5. *Protosign*: A manual-based communication system, in which pantomimes are conventionalized within a community to allow meanings to be conveyed more economically and less ambiguously.

A related innovation to support communication was that the same pantomime might become conventionalized in different ways to convey different ideas – it is easy to pantomime a flying bird, but it requires innovation to derive different protosigns from it to signify "bird" and "flying," and then concatenate another protosign for "dead" with that for "bird" to draw attention to a dead bird, one that is certainly not flying.

While other primates have manual skill, they do not have *flexible* vocal skill and learning. MSH thus offers:

MSH6. *Protospeech.* Other primates do not have flexible vocal learning. MSH posits that *protospeech* rests on the "invasion" of the vocal apparatus by collaterals from the protosign system based on manual control and recognition.[9]

MSH6 does not claim that humans had fully expressive sign languages before they developed vocal control (as seems to be argued by Stokoe, 2001). Rather, it hypothesizes that *even limited protosign* could open the way for arbitrary gestures to express meaning (for the relevant debate, see Arbib, 2005; Fogassi & Ferrari, 2004; MacNeilage & Davis, 2005). Eventually, this provided the selective advantage for biological evolution of vocal control, expanding the range of available gestures while adding sound symbolism to the communicative repertoire.[10]

However, MSH asserts early humans did *not* use language as distinct from protolanguage, but, rather, that MSH5 and MSH6 yielded a brain that was *language-ready* in the sense that it could support the necessary cognitive processes for the *eventual* invention, use and social transmission of languages:

MSH7. With the emergence of *Homo sapiens*, cultural rather than biological evolution dominated as *true languages* (with extended lexicons and grammars) emerged from primitive *protolanguages* (limited vocabulary, little or no grammar), with not only a widening of the lexicons but also the emergence of *compositional semantics* supported by more and more constructions for combination of words to express and comprehend an expanded range of novel meanings.

Brain lesions can impair the use of sign language without impairing the use of pantomime (Corina et al., 1992; Marshall, Atkinson, Smulovitch, Thacker, & Woll, 2004). Thus, ad hoc use of pantomime is neurally different from access to a symbol within a language. This provides important support for the view that *the use of language cannot depend only on the brain mechanisms supporting manual skills* even though MSH specifies a crucial scaffolding role for manual skill in the emergence of language. A *protolanguage* can convey novel meanings in a limited way by stringing together two or more *protowords* (whether manual or vocal) without engaging constructions in the linguist's sense: the neural processes required to support such a protolanguage are little different from those required to support pantomime – they do not implicate those that are distinctive for language. The crucial difference between a *protoword* and a *word* is that the former is an isolated utterance associated with a context-dependent range of meanings; the latter is part of a grammatical system that specifies what constructions are most associated with the use of the word, and how those constructions build on the form and meaning of the word.

We now chart how, according to MSH, constructions in the linguists'

---

[9] Here as elsewhere, evidence for these hypotheses is carefully weighed against alternative claims in (Arbib, 2012) and later publications. The edited book (Arbib, 2020) assesses strengths and weaknesses of MSH from the viewpoints of diverse disciplines, and offers a "new road map" for future research.

[10] And recall "The Learning and Construction of Birdsong" from §8.

[11] The investigation of spiders constructing their webs (Eberhard, 2020) would provide a powerful complement to the study of bird nest construction in §8.

sense came to be added to such protolanguages, so that – perhaps over many tens of millennia – increasing the range and subtlety of constructions yielded complex protolanguages that were rich enough to provide early examples of languages.

We saw that *complex action recognition,* the ability to recognize when another's performance combines known actions must, when acquiring novel skills*,* support segmenting the performance into pieces that are unknown but are candidates for mastery as separate actions, with the teacher modifying pantomime to provide segmentation clues. MSH posits that this capability eventually came to support *fractionation* of protowords – breaking them into pieces (subtractive mental construction) that came to be associated with their own meanings. The complementary process (additive mental construction) yielded constructions as a way to "put the pieces back together" – but now with the advantage that they could also combine pieces that had never been combined before. Note the vital distinction: arranging objects in a row is a very limited form of physical construction; uttering protowords in arbitrary sequences is a very limited form of mental construction. In each case, "interesting" construction requires operating on objects to form a new object that has uses (physical construct) or meanings (mental construct) that could not be obtained without employing some subtle method of construction.

To use an anachronistic example, consider pantomimes for *open-a-door* and *close-a-door*. The pantomimes contain no explicit component for *door*. However, the pantomime of turning the handle, as the one common component, could become fractionated out as the protosign for *door* while the two directions of moving the hand yield protosigns for *open* and *close*. Note that in this example, the slot-filler *X* for, e.g., the construction *open-X* meets a *semantic* criterion of being something-that-opens. We are a long way from generalized constructions whose slot-fillers are defined by syntactic categories like "noun" or "verb," but even at this early stage, we see that protowords are beginning to become words since they are now enriched by their linkage to constructions.

Over time, MSH suggests, constructions became merged or generalized, and complemented by new ones, and so the "slot fillers" went from having highly limited semantic variation to becoming abstract enough to form syntactic categories. Thus lexicon and grammar evolved (culturally) together (Heine & Kuteva, 2007).

For pedagogy, pantomimes may be needed to indicate preconditions and effects (postconditions) for an action, and not just the movement involved in that action. However, a particular challenge remains, namely the indication of non-pantomimable properties – for example, developing a protosign for a color. Here, the almost uniquely human capability for deixis (pointing) would come into play to support protosigns as gestures that need not be conventionalized or fractionated from a pantomime. For example, the use of a novel gesture to mean "red" could be coupled with repeated pointing to, or manipulation of, red objects that share that property. This raises Quine's (1960) "gavagai" problem: if only a few objects are pointed at while saying "gavagai," the observer might associate "gavagai" with some other property shared by the objects – a misunderstanding that may be corrected with sufficient experiences of others using the term (Steels & Belpaeme, 2005).

13. The way forward

We offer an overall framework for ethology as well as primatology to assess how the active organism engages with the world beyond its body in diverse ways that include tool use, construction and communication. A key goal is to move construction into the mainstream of ethology, with bird nest construction as a central example, and initiate investigation of the apparent separation of nest construction and birdsong. The way forward must continue to complement focused studies of animal and human behavior with continued development of the general perspectives that can frame analysis of their convergences and divergences.

"Distalization of the end-effector" plays a key role in the use of tools and the related displacement in a bird of the current end-effector from beak to the tip of a twig being inserted into the nest it is constructing. This is complemented by the ability to observe and perhaps imitate the successes of others, as well as engage in social cooperation. As a link to cultural evolution, we have exemplified the impact of the open-ended accumulation of components and procedures in the "lithic landscape" of nut-cracking capuchin monkeys, and the way in which birds can learn from observing the building behavior of others. When we come to humans, we saw the interplay between the availability of diverse tools and other objects together with a rich verbal and social environment that supports and constrains their use (§§9-12).

Exploring this interplay, Stout (2021) has enunciated a form of *perceptual-motor hypothesis* that builds on the construction of internal models and intuitive physics required for material production, and the sensory predictions and extended forms of action made possible by these models. This part is consistent with our general framework, but Stout further charts how fronto-parietal brain systems supporting action execution and observation have undergone major changes in the human-related evolution of primates (Stout & Hecht, 2017). Changes in these systems are implicated in the Paleolithic technologies that the technological hypothesis views as helping shape the course of human neuro-cultural evolution. However, to constrain the length of an already long paper, we have spent little space on studies of brain anatomy, function, genetics, and plasticity, studies that are crucial to the background and future explorations of our framework.

We address different kinds of "technological" action and insist that the examples must include, although not be limited to, construction. While objects may in general contrast with actions in their temporal persistence, a crucial feature of (even subtractive) construction is that objects may play evanescent roles (to be distinguished from the dynamics of an enduring object, or even the changing states of an agent as it learns). The issue of *images* and *goals* motivates the further investigation about "affordances" and how goal-related and perceptually accessible they need to be. Are the insulating properties of different bird nest materials an affordance for action? Are the chemical properties of ingredients in a glue affordances? Addressing such questions requires widespread comparative studies as we compare/contrast concrete simulation-based action goals versus more abstract construction "images." Indeed, analysis of how birds construct their nests has highlighted two key questions of importance far beyond this domain: to what extent does an "image" set the postconditions for the various substages of construction, and what is the tradeoff between "innate" skills that can be tuned through experience and the mastery of truly novel skills that can be described as the fruit of cultural evolution? This discussion required us to try to distinguish species-general development from development that depends on social traditions.

Birds may exhibit strong species-typical preferences for the kind of nest they build and yet will vary the construction in light of the available materials. Again, there is a bias for certain kinds of sites for nest building, but observation of reproductive success by others can bias that selection. Moreover, many birds build in places that bear little resemblance to "species-specific" locations when building in human-centric environments. All this poses challenges for future research on development in birds where experience is manipulated and detailed observations are made. Lewarch and Hoekstra (2018) have shown that nest building by deer mice has a strong genetic component affecting the shape of the burrow they build, and this seems a promising avenue to consider for birds too.

Young capuchin monkeys do not coordinate with adult monkeys in acquiring nut-cracking skill, but there is an indirect connection in that the young monkeys are in a constructed niche in which there are not only nut-bearing trees but also hammerstones and anvils accumulated through cultural traditions that likely go back centuries. Our model of opportunistic scheduling of actions (as they compete for execution based on their current priority) demonstrated that ethograms as explicit descriptions of behaviors are (most likely) not what is stored in the head so much as a context-dependent way of scheduling priorities based on learned desirability for different tasks.

Capuchin monkeys exhibit the typical developmental property that actions appearing early on in one setting (percussion that is intrinsically rewarding – "joy of percussion") provide the building blocks for later mastery of nut cracking. However, nonhumans seem not to exhibit the open ended ratcheting (adding new features across generations, Tennie, Call, & Tomasello, 2009) seen in human cultures. We hypothesized that what made this possible was the development of pedagogy and language. Tools/technologies interact with each other in a potentially combinatorial way to produce novelty, a crucial point in the study of cultural evolution (Kolodny, Creanza, & Feldman, 2015).

We see that monkey and protohumans required immense effort to master each of a limited set of skills (including skills for constructing a particular type of object), but human language transformed the primate behavioral landscape. Although mastering a language requires years of practice and social interaction, once this is achieved a human can readily construct hundreds or thousands of utterances a day, and many of these may be novel. Language supports rapid acquisition of novel behaviors, even though practice may be required to smooth and speed up their performance. Recalling the distinction between event-level and trajectory-level processes, we stress that language is for the most part related to the event level. Consider following a recipe for food preparation: With sufficiently detailed instructions in which each named skill is familiar, we may be able to perform a novel and highly complex behavior successfully on the first attempt – even though we may become more skillful if we repeat the behavior.

Biryukova, Bril, Frolov, and Koulikov (2015) examined stone knapping by Indian craftsmen of different levels of skill and found that, the higher the level of motor skill (requiring years of practice), the more stable are the functional and the more variable the nonfunctional joint loadings. This suggests that to acquire naturalistically challenging motor expertise, years of practice are needed even though the overall structure of the task has long been mastered. Thus, tuning the actions that link preconditions and effects, P}A}E, by tuning the parametric adjustment of motor schemas to incorporate affordance details (trajectory-level) of the "building blocks" of action may be more challenging than "putting the actions together." The human uniqueness lies in the event-level ability to learn recipes for successfully combining familiar actions to construct new results. The hypothesis, for the human-centric aspect of our work, is that the explosion of tool making in humans co-evolved with language-related declarative knowledge that in particular can scaffold the mastery of new action combinations. And not just tool making. It is only because stone tools provide our major fossil record of distant human prehistory that making such tools is especially privileged in the technological pedagogy hypothesis. Long, long before writing, (proto)human groups – women, men, and children – had diverse needs for communication in many contexts beyond pedagogy and unrelated to knapping. We thus add here the notion of *micro-protolanguages* (Arbib, 2023). Early humans with different responsibilities within social groups may have developed limited protolanguages specific to communicating about their particular activities (while still relying greatly on context, gesture, and pantomime) long before the emergence of larger protolanguages integrating all of them and shared by the whole community.

If we consider a primary aim of language to be to influence others then each pantomime or utterance may have its own meaning, yet its overall aim is to effect a certain state in the mind of the other. It has been debated whether the capacity for skill acquisition was central to the emergence of distinctly human technological capacities (Stout, 2021) or whether human semantic and communicative capacities such as the use of analogy in reasoning and teaching (Brand et al., 2021) were more decisive. The MSH framework (§12) suggests (as it did for the relation of protosign and protospeech, Arbib, 2005) an "expanding spiral" as new technologies and styles of habitation supported further development of (micro-proto)language and as the development of constructions in language supported ever more complex planning of novel technologies and constructions exploiting them. However, it must be stressed (perhaps recalling the gossip theory of Dunbar, 1996) that telling stories has long been a crucial function of language (Barnard, 2013) but has been overlooked in both MSH and the technological pedagogy hypothesis, and thus sets a target for future research.

Thus, with our account of the technological pedagogy hypothesis and the praxis-based biocultural evolution of language, we have implicitly addressed the suggestion by Osiurak and colleagues (Osiurak & Badets, 2016; Osiurak & Danel, 2018) that modern human tool use might be based on an ability to reason about physical properties of tools and objects. By linking human tool use (and motor skills more generally) to language, we open the discussion (but only for humans) to a role for reasoning skills in creating programs for developing new tools, adapting elements of prior behaviors to new purposes, and imagining how to construct novel objects to suit our purposes. In this paper, we have not gone beyond the basics of language to offer an account of reasoning, but our account does underwrite how, once language is available, such reasoning naturally extends our evolutionarily prior ability to master some new skills through demonstration or observation. We suggest that research on how humans reason about tools and objects will be enriched by building on the approach we offer to the linkage of "what" and "how" in motor programs that address affordances and goals.

To close, though, we reiterate that the work presented here is not defined primarily by its relevance to understanding human evolution. Our framework is intended to invigorate the comparative study of cognition in diverse species by directing increased attention to construction tasks, with and without the use of tools or tooling, for which the construction of bird nests provides a powerful, and too long neglected, model system.

13.1. Data and Code Availability Statement

The relevant data and code are associated with the individual research efforts of the four authors and are linked where appropriate to our separate papers referred to in this submitted paper. The synthesis offered here is based on the presentation of key findings and ideas, rather than the specific data details.

Fig. 5

Fig. 5. A male zebra finch building a nest in the Healy lab using material that he can only handle with a beak, while the female sits by his side.

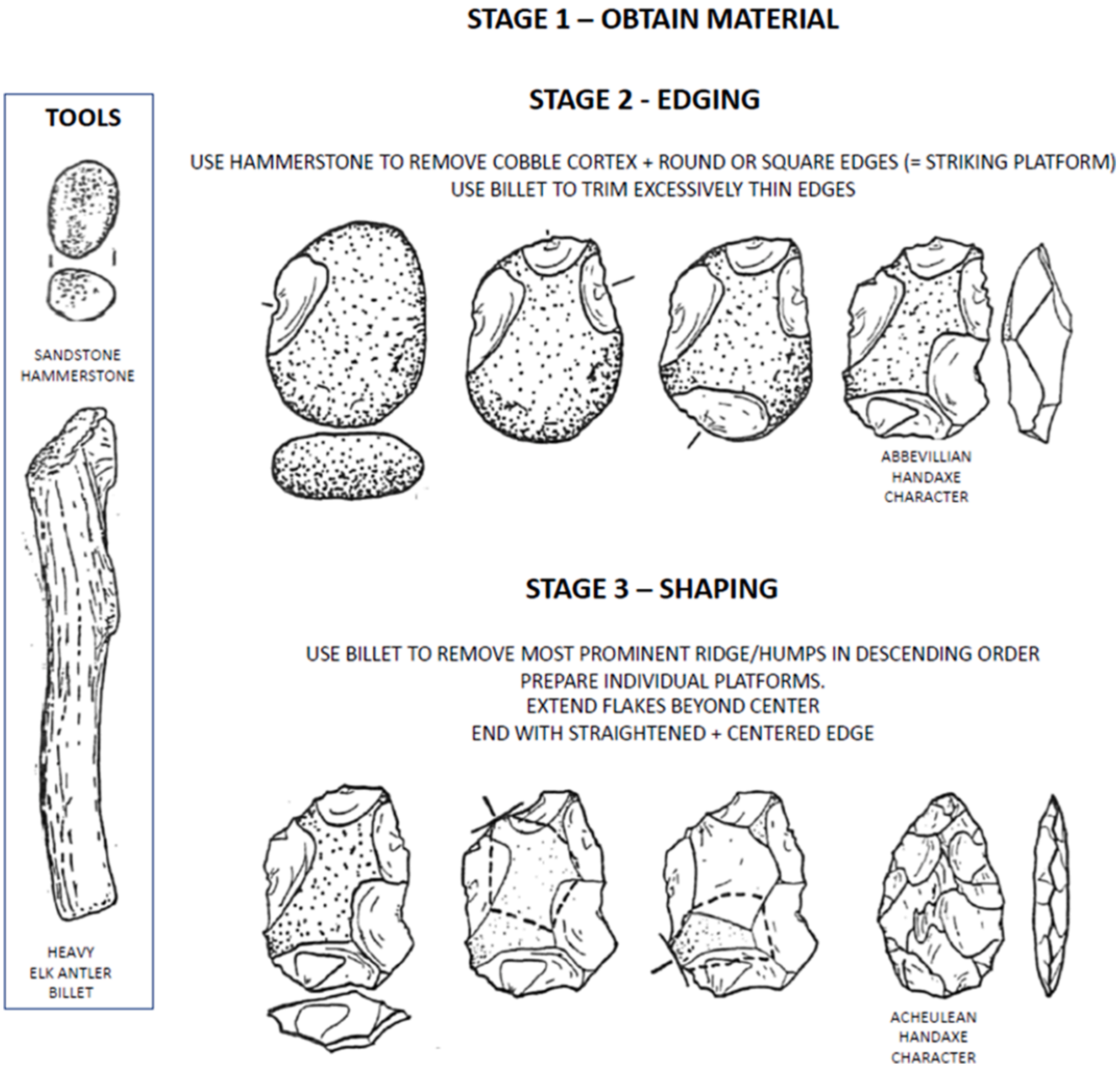

Fig. 6. Stages in Acheulean manufacture of a hand axe. “Abbevillian” is a disused culture-historical term traditionally used for “crude” early European handaxes. Comparison of stages 2 and 3 illustrates the additional technical operations and intentions required to produce refined later Acheulean forms. Adapted from (Callahan, 1987) by permission of Melody Callahan and the estate of Errett Callahan.

Declaration of Competing Interest

The authors declare that they have no Conflicts of Interest.

Data availability

No new data were used for the research described in the article.

Acknowledgements

This article is based in part on the ABLE (Action, Brain, Language, Evolution) mini-workshop on “Construction, Tools and Language” held at Emory University, Atlanta, November 2019, which was supported in part by the Center for Mind, Brain, and Culture at Emory University and the National Science Foundation under Grant No. BCS-1343544 “INSPIRE Track 1: Action, Vision and Language, and their Brain Mechanisms in Evolutionary Relationship” (Michael A. Arbib, Principal Investigator). Arbib, Fragaszy, and Healy are grateful for Dietrich Stout’s hospitality in Atlanta. We thank the anonymous reviewers for their thoughtful critiques of an earlier draft.

References

Alstermark, B., Lundberg, A., Norrsell, U., Sybirska, E., 1981. Integration in descending motor pathways controlling the forelimb in the cat: 9. Differential behavioural defects after spinal cord lesions interrupting defined pathways from higher centres to motoneurones. Experimental Brain Research 42, 299–318.

Arbib, M.A., 1981. Perceptual structures and distributed motor control. In: Brooks, V.B. (Ed.), Handbook of Physiology — The Nervous System II. Motor Control, Ed. American Physiological Society, Bethesda, MD, pp. 1449–1480.

Arbib, M.A., 2005. Interweaving Protosign and Protospeech: Further Developments Beyond the Mirror. Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems (6), 145–171.

Arbib, M.A., 2012. How the Brain Got Language: The Mirror System Hypothesis. Oxford University Press, New York & Oxford.

Arbib, M.A., 2016. Towards a Computational Comparative Neuroprimatology: Framing the Language-Ready Brain. Physics of Life Reviews 16, 1–54.

Arbib, M.A., 2021. When Brains Meet Buildings: A Conversation between Neuroscience and Architecture. Oxford University Press, New York.

Arbib, M.A., 2023. Pantomime within and beyond the evolution of language. In: Zywiczynski, P., Wacewicz, S., Boruta-Żywiczyńska, M., Blomberg, J. (Eds.), Perspectives on pantomime: evolution, development, interaction, Eds. John Benjamins, Philadelphia/Amsterdam.

Arbib, M.A., 2020. How the Brain Got Language: Towards a New Road Map. Johns Benjamins, Amsterdam/Philadelphia.

Arbib, M.A., Bonaiuto, J.J., Jacobs, S., Frey, S.H., 2009. Tool use and the distalization of the end-effector. Psychol Res 73 (4), 441–462. https://doi.org/10.1007/s00426-009-0242-2.

Arbib, M.A., Gasser, B., Barrès, V., 2014. Language is handy but is it embodied? Neuropsychologia 55, 57–70.

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

Barham, L., 2013. From Hand to Handle: The first industrial revolution. Oxford University Press.

Barnard, A., 2013. Cognitive and social aspects of language origins. In: Lefebvre, C., Comrie, B., Cohen, H. (Eds.), New perspectives on the origins of language, Eds. Amesterdam/Philadelphia: John Benjamins, pp. 53–71.

Beyene, Y., Katoh, S., WoldeGabriel, G., Hart, W.K., Uto, K., Sudo, M., Asfaw, B., 2013. The characteristics and chronology of the earliest Acheulean at Konso, Ethiopia. Proceedings of the National Academy of Sciences 110 (5), 1584–1591. https://doi.org/10.1073/pnas.1221285110.

Biryukova, E. V., Bril, B., Frolov, A. A., & Koulikov, M. A. (2015). Movement Kinematics as an Index of the Level of Motor Skill: The Case of Indian Craftsmen Stone Knapping. 19(1), 34. doi:10.1123/mc.2013-0042.

Bonaiuto, J.J., Arbib, M.A., 2010. Extending the mirror neuron system model, II: what did I just do? A new role for mirror neurons. Biol Cybern 102 (4), 341–359. https://doi.org/10.1007/s00422-010-0371-0.

Bonaiuto, J.J., Arbib, M.A., 2015. Learning to grasp and extract affordances: the Integrated Learning of Grasps and Affordances (ILGA) model. Biological cybernetics 109 (6), 639–669 doi:610.1007/s00422-00015-00666-00422.

Brand, C.O., Mesoudi, A., Smaldino, P.E., 2021. Analogy as a Catalyst for Cumulative Cultural Evolution. TRENDS in Cognitive Sciences 25 (6), 450–461. https://doi.org/10.1016/j.tics.2021.03.002.

Breen, A.J., Bonneaud, C.C., Healy, S.D., Guillette, L.M., 2019. Social learning about construction behaviour via an artefact. Animal cognition 22 (3), 305–315. https://doi.org/10.1007/s10071-019-01240-x.

Breen, A.J., Guillette, L.M., Healy, S.D., 2016. What can nest-building birds teach us? Comparative Cognition & Behavior Reviews 11, 83–102.

Breen, A.J., Lovie, K.E., Guerard, C., Edwards, S.C., Cooper, J., Healy, S.D., Guillette, L. M., 2020. Juvenile socio-ecological environment shapes material technology in nest-building birds. Behavioral Ecology 31 (4), 892–901. https://doi.org/10.1093/beheco/araa027.

Breen, A.J., Sugasawa, S., Healy, S.D., 2021. Manipulative and technological skills do not require a slow life history. Frontiers in Ecology and Evolution, Behavioral and Evolutionary Ecology, 635802.

Byrne, R.W., 2003. Imitation as behaviour parsing. Philos Trans R Soc Lond B Biol Sci 358 (1431), 529–536. https://doi.org/10.1098/rstb.2002.1219.

Byrne, R.W., Russon, A.E., 1998. Learning by imitation: a hierarchical approach. Behav Brain Sci 21 (5), 667–684 discussion 684-721. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10097023.

Cahill, J.A., Armstrong, J., Deran, A., Khoury, C.J., Paten, B., Haussler, D., Jarvis, E.D., 2021. Positive selection in noncoding genomic regions of vocal learning birds is associated with genes implicated in vocal learning and speech functions in humans. Genome Research 31 (11), 2035–2049.

Callahan, E., 1987. Primitive Technology: Practical Guidelines for Making Stone Tools, Pottery, Basketry, etc., the Aboriginal Way. Piltdown Productions, Lynchburg, VA.

Chomsky, N., 1956. Three models for the description of language. IEEE Transactions on Information Theory 2 (3), 113–124.

Chomsky, N., 1995. The Minimalist Program. The MIT Press, Cambridge, MA.

Coelho, C., Falótico, T., Izar, P., Mannu, M., Resende, B., Siqueira, J., Ottoni, E., 2015. Social learning strategies for nut-cracking by tufted capuchin monkeys (Sapajus spp.). Animal cognition 18 (4), 911–919.

Colbourne, J.A.D., Auersperg, A.M.I., Lambert, M.L., Huber, L., Völter, C.J., 2021. Extending the Reach of Tooling Theory: A Neurocognitive and Phylogenetic Perspective. Topics in Cognitive Science. https://doi.org/10.1111/tops.12554. *n/a* (n/a).

Collias, N.E., 1997. The origin and evolution of nest building by passerine birds. Condor 99, 253–269.

Collias, E.C., Collias, N.E., 1964. The development of nest-building behavior in a weaverbird. The Auk 81 (1), 42–52.

Coolidge, F.L., Wynn, T., 2005. Working Memory, its Executive Functions, and the Emergence of Modern Thinking. Cambridge Archaeological Journal 15 (01), 5–26. https://doi.org/10.1017/S0959774305000016.

Corina, D.P., Poizner, H., Bellugi, U., Feinberg, T., Dowd, D., O'Grady-Batch, L, 1992. Dissociation between linguistic and nonlinguistic gestural systems: a case for compositionality. Brain Lang 43 (3), 414–447.

Croft, W., 2001. Radical construction grammar: syntactic theory in typological perspective. Oxford University Press, Oxford.

d'Errico, F., Stringer, C.B., 2011. Evolution, revolution or saltation scenario for the emergence of modern cultures? Philosophical Transactions of the Royal Society B: Biological Sciences 366 (1567), 1060.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. Exp Brain Res 91 (1), 176–180. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1301372.

Dunbar, R., 1996. Grooming, Gossip and the Evolution of Language. Faber and Faber Ltd, London.

Eberhard, W., 2020. Spider Webs: Behavior, Function, and Evolution. University of Chicago Press, Chicago.

Edwards, S.C., Hall, Z.J., Ihalainen, E., Bishop, V.R., Nicklas, E.T., Healy, S.D., Meddle, S. L., 2020. Neural circuits underlying nest building in male zebra finches. Integrative and Comparative Biology 60 (4), 943–954.

Ericsson, K.A., Kintsch, W., 1995. Long-term working memory. Psychological Review 102, 211–245.

Eshchar, Y., Izar, P., Visalberghi, E., Resende, B., Fragaszy, D.M., 2016. When and where to practice: social influences on the development of nut-cracking in bearded capuchins (Sapajus libidinosus). Animal cognition 19 (3), 605–618.

Fagg, A.H., Arbib, M.A., 1998. Modeling parietal-premotor interactions in primate control of grasping. Neural Netw 11 (7-8), 1277–1303. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12662750.

Flynn, E.G., Laland, K.N., Kendal, R.L., Kendal, J.R., 2013. Developmental niche construction (Target Article with Commentaries). Developmental Science 16 (2), 296–313. https://doi.org/10.1111/desc.12030.

Fogassi, L., Ferrari, P.F., 2004. Mirror neurons, gestures and language evolution. Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems 5, 345–363.

Fragaszy, D.M., 2011. Community resources for learning: how capuchin monkeys construct technical traditions. Biological Theory 6 (3), 231–240.

Fragaszy, D.M., Aiempichitkijkarn, N., Eshchar, Y., Mangalam, M., Izar, P., Resende, B., Visalberghi, E., 2023. The development of expertise at cracking palm nuts by wild bearded capuchin monkeys, *Sapajus libidinosus*. Animal Behaviour 197, 1–14.

Fragaszy, D.M., Eshchar, Y., Visalberghi, E., Resende, B., Laity, K., Izar, P., 2017. Synchronized practice helps bearded capuchin monkeys learn to extend attention while learning a tradition. Proceedings of the National Academy of Sciences 114 (30), 7798–7805.

Fragaszy, D.M., Izar, P., Liu, Q., Eshchar, Y., Young, L.A., Visalberghi, E., 2016. Body mass in wild bearded capuchins,(Sapajus libidinosus): Ontogeny and sexual dimorphism. American Journal of Primatology 78 (4), 473–484.

Fragaszy, D.M., Mangalam, M., 2018. Tooling. In: Advances in the Study of Behavior, 50. Elsevier, pp. 177–241.

Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. Brain 119, 593–609.

García-Medrano, P., Ollé, A., Ashton, N., Roberts, M.B., 2018. The mental template in handaxe manufacture: new insights into Acheulean lithic technological behavior at Boxgrove, Sussex, UK. Journal of Archaeological Method and Theory 26 (1), 1–27.

Gärdenfors, P., Högberg, A., 2017. The Archaeology of Teaching and the Evolution of Homo docens. Current Anthropology 58 (2), 188–208. https://doi.org/10.1086/691178.

Gibson, J.J., 1977. The ecological approach to visual perception. Boston: Houghton Mifflin.

Goldberg, A.E., 1995. Constructions: a Construction Grammar approach to argument structure. Chicago: The University of Chicago Press.

Goodale, M.A., Milner, A.D., Jakobson, L.S., Carey, D.P., 1991. A neurological dissociation between perceiving objects and grasping them. Nature 349 (6305), 154–156. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1986306.

Gowlett, J.A.J, 1996. Mental abilities of early Homo: elements of constraint and choice in rule systems. In: Mellars, P., Gibson, K.R. (Eds.), Modelling the early human mind, Eds. Oxbow Books, Oxford.

Guillette, L.M., Healy, S.D., 2019. Social learning in nest-building birds watching live-streaming video demonstrators. Integrative zoology 14 (2), 204–213.

Guillette, L.M., Scott, A.C.Y., Healy, S.D., 2016. Social learning in nest-building birds: a role for familiarity. Proceedings of the Royal Society B: Biological Sciences 283 (1827), 20152685.

Gwan, J.G.L., McShea, D.W., 2020. Operationalizing goal directedness: An empirical route to advancing a philosophical discussion. *Philosophy*. Theory, and Practice in Biology 12 (5). https://doi.org/10.3998/ptpbio.16039257.0012.005.

Hansell, M., 2000. Bird Nests and Construction Behaviour. Cambridge University Press.

Heine, B., Kuteva, T., 2007. The genesis of grammar: A reconstruction. Oxford University Press, Oxford.

Henderson, J.M., Choi, W., Lowder, M.W., Ferreira, F., 2016. Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. Neuroimage 132, 293–300.

Iberall, T., Bingham, G., Arbib, M.A., 1986. Opposition Space as a Structuring Concept for the Analysis of Skilled Hand Movements. In: Heuer, H., Fromm, C. (Eds.), Generation and Modulation of Action Patterns, Eds. Springer-Verlag, Berlin, pp. 158–173.

Isler, K., Van Schaik, C.P., 2014. How humans evolved large brains: Comparative evidence. Evolutionary Anthropology: Issues, News, and Reviews 23 (2), 65–75. https://doi.org/10.1002/evan.21403.

Jeannerod, M., Biguer, B., 1982. Visuomotor mechanisms in reaching within extra-personal space. In: Ingle, D.J., Mansfield, R.J.W., Goodale, M.A. (Eds.), Advances in the Analysis of Visual Behavior, Eds. The MIT Press, Cambridge, MA, pp. 387–409.

Jeannerod, M., Decety, J., 1995. Mental motor imagery: a window into the representational stages of action. Curr Opin Neurobiol 5 (6), 727–732. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8805419.

Jeannerod, M., Decety, J., Michel, F., 1994. Impairment of grasping movements following a bilateral posterior parietal lesion. Neuropsychologia 32 (4), 369–380.

Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8047246.

Johnson, S.H., 2000. Thinking ahead: the case for motor imagery in prospective judgements of prehension. Cognition 74 (1), 33–70. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10594309.

Jung, Y., Jung, S., Lee, S.-i., Kim, W., Kim, H.-Y, 2021. Avian mud nest architecture by self-secreted saliva. Proceedings of the National Academy of Sciences 118 (3), e2018509118. https://doi.org/10.1073/pnas.2018509118.

Kabadayi, C., Osvath, M., 2017. Ravens parallel great apes in flexible planning for tool-use and bartering. Science 357 (6347), 202–204. https://doi.org/10.1126/science.aam8138.

Kaplan, H., Gurven, M., Winking, J., Hooper, P.L., Stieglitz, J., 2010. Learning, menopause, and the human adaptive complex. Annals of the New York Academy of Sciences 1204 (1), 30–42.

Keysers, C., Kohler, E., Umiltà, M.A., Nanetti, L., Fogassi, L., Gallese, V., 2003. Audiovisual mirror neurons and action recognition. Experimental Brain Research 153 (4), 628–636. https://doi.org/10.1007/s00221-003-1603-5. Retrieved from.

Kolodny, O., Creanza, N., Feldman, M.W., 2015. Evolution in leaps: The punctuated accumulation and loss of cultural innovations. Proceedings of the National Academy of Sciences 112 (49), E6762–E6769. https://doi.org/10.1073/pnas.1520492112.

Kolodny, O., Edelman, S., 2018. The evolution of the capacity for language: the ecological context and adaptive value of a process of cognitive hijacking. Philosophical Transactions of the Royal Society B: Biological Sciences 373 (1743), 20170052. https://doi.org/10.1098/rstb.2017.0052.

Kuhn, S.L., 2020. The evolution of Paleolithic technologies. New York: Routledge.

Kuperstein, M., 1988. Neural model of adaptive hand-eye coordination for single postures. Science 239 (4845), 1308–1311. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3344437.

Lepre, C.J., Roche, H., Kent, D.V., Harmand, S., Quinn, R.L., Brugal, J.-P., Feibel, C.S., 2011. An earlier origin for the Acheulian. Nature 477 (7362), 82–85.

Lewarch, C.L., Hoekstra, H.E., 2018. The evolution of nesting behaviour in Peromyscus mice. Animal Behaviour 139, 103–115. https://doi.org/10.1016/j.anbehav.2018.03.008.

Lind, J., 2018. What can associative learning do for planning? Royal Society Open Science 5 (11), 180778. https://doi.org/10.1098/rsos.180778.

Liu, Q., Simpson, K., Izar, P., Ottoni, E., Visalberghi, E., Fragaszy, D.M., 2009. Kinematics and energetics of nut-cracking in wild capuchin monkeys (Cebus libidinosus) in Piauí, Brazil. American Journal of Physical Anthropology 138 (2), 210–220. https://doi.org/10.1002/ajpa.20920.

Lombao, D., Guardiola, M., Mosquera, M., 2017. Teaching to make stone tools: new experimental evidence supporting a technological hypothesis for the origins of language. Scientific Reports 7 (1), 14394. https://doi.org/10.1038/s41598-017-14322-y.

Loukola, O.J., Seppänen, J.-T., Krams, I., Torvinen, S.S., Forsman, J.T, 2013. Observed fitness may affect niche overlap in competing species via selective social information use. The American Naturalist 182 (4), 474–483.

Lycett, S.J., Schillinger, K., Eren, M.I., von Cramon-Taubadel, N., Mesoudi, A., 2016. Factors affecting Acheulean handaxe variation: Experimental insights, microevolutionary processes, and macroevolutionary outcomes. Quaternary International 411, 386–401. https://doi.org/10.1016/j.quaint.2015.08.021.

MacNeilage, P.F., Davis, B.L., 2005. The frame/content theory of evolution of speech: A comparison with a gestural-origins alternative. Interaction Studies 6 (2), 173–199. Retrieved from. http://www.sign-lang.uni-hamburg.de/BibWeb/LiDat.acgi?ID=63566.

Marler, P., Slabbekoorn, H., 2004. Nature's Music: The Science of Birdsong. Elsevier Academic Press, Amsterdam.

Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A., Woll, B., 2004. Aphasia in a user of British Sign Language: Dissociation between sign and gesture. Cognitive Neuropsychology 21, 537–554.

Miller, N.E., Dollard, J., 1941. Social learning and imitation. Yale University Press, New Haven, CT.

Morgan, T.J.H., Uomini, N.T., Rendell, L.E., Chouinard-Thuly, L., Street, S.E., Lewis, H.M., Laland, K.N., 2015. Experimental evidence for the co-evolution of hominin tool-making teaching and language. Nature Communications 6 (1), 6029. https://doi.org/10.1038/ncomms7029.

Mulcahy, N.J., Call, J., 2006. Apes save tools for future use. Science 312 (5776), 1038–1040. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16709782.

Muth, F., Healy, S.D., 2011. The role of adult experience in nest building in the zebra finch, Taeniopygia guttata. Animal Behaviour 82 (2), 185–189. https://doi.org/10.1016/j.anbehav.2011.04.021.

Nelson, M.J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Pallier, C., 2017. Neurophysiological dynamics of phrase-structure building during sentence processing. Proceedings of the National Academy of Sciences, 201701590.

Osiurak, F., Badets, A., 2016. Tool use and affordance: Manipulation-based versus reasoning-based approaches. Psychological Review 123 (5), 534–568. https://doi.org/10.1037/rev0000027.

Osiurak, F., Danel, S., 2018. Tool use and dexterity: beyond the embodied theory. Animal Behaviour 139, e1–e4. https://doi.org/10.1016/j.anbehav.2018.03.016.

Oswalt, W.H., 1976. An anthropological analysis of food-getting technology. John Wiley & Sons, New York.

Oztop, E., Arbib, M.A., 2002. Schema design and implementation of the grasp-related mirror neuron system. Biol Cybern 87 (2), 116–140. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12181587.

Oztop, E., Arbib, M.A., Bradley, N, 2006. The development of grasping and the Mirror System. In: Arbib, M.A. (Ed.), Action to language via the mirror neuron system, Ed. Cambridge University Press, Cambridge, pp. 397–423.

Oztop, E., Bradley, N.S., Arbib, M.A., 2004. Infant grasp learning: a computational model. Exp Brain Res 158 (4), 480–503. Retrieved from. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15221160.

Pargeter, J., Khreisheh, N., Stout, D., 2019. Understanding stone tool-making skill acquisition: Experimental methods and evolutionary implications. Journal of Human Evolution 133, 146–166. https://doi.org/10.1016/j.jhevol.2019.05.010.

Pargeter, J., Kreisheh, N., Shea, J.J., Stout, D., 2020. Knowledge vs. know-how? Dissecting the foundations of stone knapping skill. Journal of Human Evolution 145, 102807. https://doi.org/10.1016/j.jhevol.2020.102807.

Pargeter, J., Liu, C., Kilgore, M.B., Majoe, A., Stout, D., 2023. Testing the Effect of Learning Conditions and Individual Motor/Cognitive Differences on Knapping Skill Acquisition. Journal of Archaeological Method and Theory 30 (1), 127–171. https://doi.org/10.1007/s10816-022-09592-4.

Petkov, C.I., Jarvis, E.D., 2012. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Frontiers in Evolutionary Neuroscience 4, 12. https://doi.org/10.3389/fnevo.2012.00012.

Quine, W.V.O., 1960. Word and Object. MIT Press, Cambridge, MA.

Resende, B.D., Ottoni, E.B., Fragaszy, D.M., 2008. Ontogeny of manipulative behavior and nut-cracking in young tufted capuchin monkeys (Cebus apella): a Perception–action perspective. Developmental Science 11 (6), 828–840.

Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. Trends in Neurosciences 21 (5), 188–194. Retrieved from. http://www.ncbi.nlm.nih.gov/pubmed/9610880.

Rogoff, B., 1991. Social interaction as apprenticeship in thinking: Guided participation in spatial planning. In: Resnick, L.B., Levine, J.M., Teasley, S.D. (Eds.), Perspectives on socially shared cognition, Eds. American Psychological Association, Washington, DC, pp. 349–364.

Sánchez-Yustos, P., 2021. Knocking on Acheulean's door. DK revisited (Bed I, Olduvai, Tanzania). Journal of Archaeological Science: Reports 35, 102763. https://doi.org/10.1016/j.jasrep.2020.102763.

Sargent, T.D., 1965. The role of experience in the nest building of the zebra finch. The Auk 82 (1), 48–61.

Schick, K.D., Toth, N., 1993. *Making Silent Stones Speak. Human Evolution and the Dawn of Technology*. Orion Books, London: Phoenix.

Semaw, S., Rogers, M., Stout, D., 2009. The Oldowan-Acheulian Transition: Is there a "Developed Oldowan" Artifact Tradition? In: Camps, M., Chauhan, P. (Eds.), Sourcebook of Paleolithic Transitions: Methods, Theories, and Interpretations, Eds. Springer, New York, pp. 173–193.

Seppänen, J.-T., Forsman, J.T., 2007. Interspecific social learning: novel preference can be acquired from a competing species. Current Biology 17 (14), 1248–1252.

Seppänen, J.-T., Forsman, J.T., Mönkkönen, M., Krams, I., Salmi, T., 2011. New behavioural trait adopted or rejected by observing heterospecific tutor fitness. Proceedings of the Royal Society B: Biological Sciences 278 (1712), 1736–1741.

Shipton, C., White, M., 2020. Handaxe types, colonization waves, and social norms in the British Acheulean. Journal of Archaeological Science: Reports 31, 102352.

Steels, L., Belpaeme, T., 2005. Coordinating perceptually grounded categories through language: A case study for colour. Behavioral and Brain Sciences 28, 469–529.

Stokoe, W.C., 2001. Language in Hand: Why Sign Came Before Speech. Gallaudet University Press, Washington, DC.

Stout, D., 2002. Skill and cognition in stone tool production: An ethnographic case study from Irian Jaya. Current Anthropology 45 (3), 693–722.

Stout, D., 2011. Stone toolmaking and the evolution of human culture and cognition. Philosophical Transactions of the Royal Society B: Biological Sciences 366, 1050–1059.

Stout, D., 2018. Archaeology and the evolutionary neuroscience of language: the technological pedagogy hypothesis. Interaction Studies 19 (1-2), 256–271.

Stout, D., 2021. The Cognitive Science of Technology. Trends Cogn Sci 25 (11), 964–977. https://doi.org/10.1016/j.tics.2021.07.005.

Stout, D., Apel, J., Commander, J., Roberts, M., 2014. Late Acheulean technology and cognition at Boxgrove, UK. Journal of Archaeological Science 41, 576–590.

Stout, D., Chaminade, T., 2012. Stone tools, language and the brain in human evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 367 (1585), 75–87. https://doi.org/10.1098/rstb.2011.0099.

Stout, D., Chaminade, T., Apel, J., Shafti, A., Faisal, A.A., 2021. The measurement, evolution, and neural representation of action grammars of human behavior. Scientific Reports 11 (1), 13720. https://doi.org/10.1038/s41598-021-92992-5.

Stout, D., Hecht, E., 2017. Evolutionary neuroscience of cumulative culture. Proceedings of the National Academy of Sciences 114 (30), 7861–7868. https://doi.org/10.1073/pnas.1620738114.

Stout, D., Hecht, E., Khreisheh, N., Bradley, B., Chaminade, T., 2015. Cognitive Demands of Lower Paleolithic Toolmaking. PLoS One 10 (4), e0121804. https://doi.org/10.1371/journal.pone.0121804.

Stout, D., Khreisheh, N., 2015. Skill Learning and Human Brain Evolution: An Experimental Approach. Cambridge Archaeological Journal 25 (04), 867–875. https://doi.org/10.1017/S0959774315000359.

Stout, D., Passingham, R., Frith, C., Apel, J., Chaminade, T., 2011. Technology, expertise and social cognition in human evolution. European Journal of Neuroscience 33 (7), 1328–1338. https://doi.org/10.1111/j.1460-9568.2011.07619.x.

Stout, D., Rogers, M.J., Jaeggi, A.V., Semaw, S., 2019. Archaeology and the origins of human cumulative culture: a case study from the earliest Oldowan at Gona, Ethiopia. Current Anthropology (3), 60. https://doi.org/10.17605/OSF.IO/UYBVW.

Suddendorf, T., Brinums, M., Imuta, K., 2016. Shaping one's future self–The development of deliberate practice. In: Michaelian, K., Klein, S.B., Szpunar, K.K. (Eds.), Seeing the future: Theoretical perspectives on future-oriented mental time travel, Eds. Oxford Univeristy Press, New York, pp. 343–366.

Sutton, R.S., 1988. Learning to predict by the methods of temporal differences. Machine Learning 3, 9–44.

Ten Cate, C., 2021. Re-evaluating vocal production learning in non-oscine birds. Philosophical Transactions of the Royal Society B-Biological Sciences 376, 20200249.

Ten Cate, C., Petkov, C.I., 2021. The grammatical abilities of animals: a comparative overview. In: Hagoort, P. (Ed.), Human Language: From Genes and Brains to Behavior, Ed. The MIT Press, Cambridge, MA, pp. 687–700.

Tennie, C., Call, J., Tomasello, M., 2009. Ratcheting up the ratchet: on the evolution of cumulative culture. Philos Trans R Soc Lond B Biol Sci 364 (1528), 2405–2415.

Tennie, C., Premo, L.S., Braun, D.R., McPherron, S.P., 2017. Early stone tools and cultural transmission: Resetting the Null hypothesis. Current Anthropology 58 (5), 652–654.

Trestman, M.A., 2012. Implicit and Explicit Goal-Directedness. Erkenntnis 77 (2), 207–236. https://doi.org/10.1007/s10670-012-9379-2.

Visalberghi, E., Barca, V., Izar, P., Fragaszy, D.M., Truppa, V., 2021. Optional tool use: The case of wild bearded capuchins (Sapajus libidinosus) cracking cashew nuts by biting or by using percussors. American Journal of Primatology 83 (1), e23221.

Visalberghi, E., Fragaszy, D.M., 2013. The EthoCebus project. Stone tool use by wild capuchin monkeys. In: Sanz, C., Call, J., Boesch, C. (Eds.), Tool use in animals. Cognition and ecology, Eds. Cambridge University Press, Cambridge, pp. 203–222.

Wadley, L., Hodgskiss, T., Grant, M., 2009. Implications for complex cognition from the hafting of tools with compound adhesives in the Middle Stone Age, South Africa. Proceedings of the National Academy of Sciences 106 (24), 9590–9594. https://doi.org/10.1073/pnas.0900957106.

Walsh, P.T., Hansell, M., Borello, W.D., Healy, S.D., 2011. Individuality in nest building: Do Southern Masked weaver (Ploceus velatus) males vary in their nest-building behaviour? Behavioural Processes 88 (1), 1–6. https://doi.org/10.1016/j.beproc.2011.06.011.

Welshon, R., 2010. Working Memory, Neuroanatomy, and Archaeology. Current Anthropology 51 (S1), S191–S199. https://doi.org/10.1086/650480.

Wilkins, J., Chazan, M., 2012. Blade production∼ 500 thousand years ago at Kathu Pan 1, South Africa: support for a multiple origins hypothesis for early Middle Pleistocene blade technologies. Journal of Archaeological Science 39 (6), 1883–1900.

Wilkins, J., Schoville, B.J., Brown, K.S., Chazan, M., 2012. Evidence for Early Hafted Hunting Technology. Science 338 (6109), 942–946. https://doi.org/10.1126/science.1227608.

Wilson, B., Petkov, C.I., 2018. From evolutionarily conserved frontal regions for sequence processing to human innovations for syntax. Interaction Studies 19 (1-2).

Wynn, T., Coolidge, F.L., 2010. Beyond Symbolism and Language: An Introduction to Supplement 1, Working Memory. Current Anthropology 51 (S1), S5–S16. https://doi.org/10.1086/650526.

Wynn, T., Gowlett, J., 2018. The handaxe reconsidered. Evolutionary Anthropology: Issues, News, and Reviews 27 (1), 21–29. https://doi.org/10.1002/evan.21552.

Yip, M.J., 2010. Structure in human phonology and in birdsong: a phonologists's perspective. In: Bolhuis, J.J., Everaert, M. (Eds.), Birdsong, Speech and Language. Converging Mechanisms, Eds. The MIT Press, Cambridge, MA.

Yu, C., Smith, Linda B., 2016. The Social Origins of Sustained Attention in One-Year-Old Human Infants. Current Biology 26 (9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026.

Zukow-Goldring, P., 2012. Assisted imitation: first steps in the seed model of language development. Language Sciences 34 (5), 569–582. https://doi.org/10.1016/j.langsci.2012.03.012.