Received: 7 September 2021 | Revised: 17 December 2021 | Accepted: 30 January 2022

DOI: 10.1111/mila.12419

SUBMITTED ARTICLE

WILEY

Stone tools, predictive processing and the evolution of language

Ross Pain

School of Philosophy, Research School of Social Sciences, Australian National University, Canberra, Australian Capital Territory, Australia

Correspondence
Ross Pain, School of Philosophy, Research School of Social Sciences, Australian National University, Canberra, ACT 2600, Australia.
Email: ross.pain@anu.edu.au

Funding information
ANU University Research Scholarship and the ANU Futures Scheme

Recent work by Stout and colleagues indicates that the neural correlates of language and Early Stone Age toolmaking overlap significantly. The aim of this paper is to add computational detail to their findings. I use an error minimisation model to outline where the information processing overlap between toolmaking and language lies. I argue that the Early Stone Age signals the emergence of complex structured representations. I then highlight a feature of my account: It allows us to understand the early evolution of syntax in terms of an increase in the number and complexity of models in a cognitive system, rather than the development of new types of processing.

KEYWORDS

evolution of language, evolutionary cognitive archaeology, gradualism, predictive processing, tool-language co-evolution

1 | INTRODUCTION

There is renewed interest in developing an old hypothesis regarding the evolution of language.[1] This hypothesis begins with the observation that there are important similarities between the production of speech and the production of stone tools. In particular, both are hierarchically organised, goal-oriented tasks that require fine-grained motor-control regulated by top-down

[1]As Stout and Chaminade note (2012, p. 75), historical proponents include both Darwin (2004/1871) and Engels (2003/1876).

processing. Consequently—so the idea goes—selective pressure for one behavioural phenotype may have aided the development of the neural substrates required to produce the other, and vice versa. This observation has produced a range of co-evolutionary hypotheses regarding the emergence of language and toolmaking, and of communication and skilled behaviour more broadly.[2]

Recent neuroimaging work by Stout and colleagues lends weight to tool-language co-evolutionary hypotheses.[3] This work indicates that the areas of the brain used to produce language and those used to produce the tool industries of the Early Stone Age overlap significantly. My goal in this article is to add computational detail to Stout and colleagues' account by integrating their results with the predictive processing framework. According to proponents of predictive processing, the brain uses hierarchical generative models to produce predictions of its future sensory states, and updates those models based on any difference between predictions and actual sensory states. Over the long-term, this results in more accurate predictions. Cognitive systems are thus characterised in terms of their ability to *minimise prediction error* (Clark, 2016; Hohwy, 2013). Adding this computational detail to tool-language co-evolutionary accounts is important. As Stout and Chaminade note, any behaviour can be analysed using a hierarchically structured sequence of units; consequently, "a *special* evolutionary relationship between toolmaking and language predicts more particular overlap in information processing demands and/or neuroanatomical substrates between these two behaviours." (Stout & Chaminade, 2012, p. 76). In other words, if co-evolutionary hypotheses are to be convincing, the similarities between toolmaking and language must extend beyond the mere fact that they can be modelled using hierarchies. Ideally, there would be similarities between: (i) the areas of the brain the two behaviours co-opt; and (ii) the information processing required to produce them. Stout and colleagues' work has focused on (i); in this article I focus on (ii).

Predictive processing has features that make it well-suited to this task.[4] At a minimum, proponents of the theory claim to account for perception and action using prediction error minimisation. Predictive processing thus offers an account of the sensorimotor skills that unite toolmaking and language production using mechanisms that operate according to a common computational principle. In addition, the notion that cognition is hierarchical is at the core of the theory. This suggests the possibility of linking the structure of cognition with the hierarchical structure we find in action sequences and language.[5] However, as we shall see, this requires clarifying precisely what it means to describe a system as "hierarchical" across these three domains.

I have two main goals in this article. The first is to use a prediction error minimisation model to identify where the information processing overlap between toolmaking and language production might lie. The second is to situate the resulting account within the literature on language evolution, and highlight some of its features.

Achieving the first of these goals requires getting clear about what it means to say that there is an information processing overlap between toolmaking and language production. I previously noted that both are hierarchically organised, goal-oriented tasks that require fine-grained motor

[2]See, for instance, Reynolds (1976, 1993), Montagu (1976), Isaac (1976), Greenfield (1991), Kimura (1993), Planer (2017), and Planer and Sterelny (2021).

[3]My focus here will be Stout et al. (2008), but see also Stout and Chaminade (2012), and Stout et al. (2021).

[4]Indeed, Stout and Chaminade themselves note that predictive processing might be used to account for evidence that the inferior parietal cortex is a supramodal processing region (Stout & Chaminade, 2012, p. 78).

[5]Put together, these two features make predictive processing a tantalising fit with Stout's (2021) *perceptual motor hypothesis* (thanks to an anonymous reviewer for this point).

control regulated by top-down processing. As such, any account of the cognitive capacities underwriting the two behaviours must explain these features. This challenge can be made more precise in the following way. It is typically thought that both complex intentional action and language display *nested part-whole structure*. Consider the action of brushing your teeth. When you reach out and apply pressure to the tap, these actions are nested within—or are a *constituent of*—the broader goal of wetting the end of your toothbrush. Likewise, both applying pressure to the tap and wetting the end of your toothbrush are nested within the broader goal of brushing your teeth.[6] This observation suggests that we can *represent* nested part-whole structure. More specifically, the thought is that the representations controlling such action must themselves be structured such that lower-level motor commands are nested within—or are a constituent of—higher-level goal representations. It follows that complex intentional action requires representations with nested part-whole structure; and, as we shall see, producing the tools of the Early Stone Age required complex intentional action.

Sentences also display nested part–whole structure. For instance, consider the sentence, "The woman brushed her teeth vigorously". "The" is a constituent of "The woman", insofar as the article is nested within the broader noun phrase. Likewise, the noun phrase "The woman" is nested within the broader sentence. And again, explaining this feature of language plausibly requires positing representations with nested part-whole structure.[7] The upshot is this: If we treat the production of language as a complex intentional action, then we can understand the information processing overlap between toolmaking and language as an attempt to explain the nested part–whole structure of the representations required for complex intentional action.[8] Of course, there are significant differences between action and syntax. The claim is *not* that the toolmakers of the Early Stone Age had full-blown syntactical capacities. Rather, the claim is that one feature of the capacities underwriting the nested part–whole structure found in toolmaking—namely, structured representations—might plausibly have been co-opted and evolved to produce the nested part–whole structure found in syntax.

The above considerations highlight an important terminological point. In the predictive processing literature, the layered sequence of models thought to be embodied in the brain are described as "hierarchical". In the linguistics literature, the ability to combine meaningful units into strings of meaningful units (or, the ability to combine words into sentences) is described as "hierarchical". These are very different. In many cases, context will be sufficient to determine which meaning is in use. However, in some cases the meaning will need to be specified. I will hence distinguish *stratified hierarchies* from *tree-like hierarchies*. "Stratified hierarchies" corresponds to the way "hierarchies" is used in the predictive processing context; that is, to refer to a series of levels within a cognitive system. "Tree-like hierarchies" refers to the branching structures used by generative linguists to model syntactic properties and cognitive scientists to model action; for instance, Stout (2011), and Fitch (2014). Finally, I will use "structured representations" to refer to representations with nested part–whole structure, which plausibly underwrite our toolmaking and syntactical abilities. I am particularly concerned with the way lower-level sensorimotor representations are influenced, or modulated, by higher-level goal representations. The former are nested within, and hence constituents of, the latter. For instance, the

[6]See Pulvermüller (2014) for a hierarchical action schema based on the tooth-brushing example; see Moro (2014) for further discussion on the apparent similarities between syntax and action (thanks to an anonymous reviewer for pointing out these references).

[7]For instance, those outlined in the generative grammars produced by linguists in the Chomskyan tradition.

[8]Importantly, there are some who dispute the claim that language is composed of nested part–whole relations—for instance, Frank et al. (2012), or Christiansen and Chater (2016). I will look in more detail at these issues in Section 4.

representation that controls the action of applying pressure to the tap and wetting my toothbrush is a constituent of the goal representation of brushing my teeth, because the latter modulates the former. As a result, these are "structured" representations. Motivating this claim, using the resources of predictive processing, will be my task in Section 3.

In addressing my second goal, I outline some features of the error minimisation account that distinguish it from others in the literature. A key dispute concerns whether language evolved suddenly (*saltationism*) or whether it evolved slowly via incremental stages (*gradualism*). An important issue in this dispute concerns the evolution of structured representations. On some accounts, structured representations are treated as a necessary condition for language production. Consequently, explaining the evolution of language requires positing an evolutionary transition from cognitive systems without structured representations to cognitive systems with structured representations. This transition might be sudden (Berwick & Chomsky, 2016) or gradual (Planer & Sterelny, 2021). Typically, this is achieved by positing an associated transition from sequential information processing to tree-like hierarchical information processing. Others emphasise evolutionary continuity, and want to avoid a commitment to transitions; either between non-structured and structured representations or between sequential and tree-like hierarchical processing. They thus deny that structured representations are required for language production (Frank et al., 2012). The error minimisation account occupies a unique place within this dialectic, because the transition between non-structured representations and structured representations can occur *without* positing a transition between sequential and tree-like hierarchical processing. The reason for this is that the former transition is made possible solely by increasing the number and sophistication of layered models in a cognitive system, which nonetheless remains governed by the principle of error minimisation. I argue that this feature makes the account an attractive choice for those attempting to produce gradualist explanations of language evolution.

In sum, I aim to motivate two claims. First, the error minimisation approach demonstrates that producing the tools of the Early Stone Age required sophisticated structured representations. Second, the error minimisation approach can account for the evolution of structured representations without needing to posit a corresponding transition in types of processing.

I proceed as follows. Section 2 outlines the results of Stout and colleagues' neuroimaging work on Early Stone Age toolmaking. Section 3 provides an overview of predictive processing and outlines a minimal model for understanding the information processing overlap between toolmaking and language. Section 4 situates this account in the literature and highlights its advantages. Section 5 concludes.

2 | STOUT AND COLLEAGUES ON THE NEUROANATOMY OF TOOLMAKING

Despite the long history of tool-language co-evolutionary hypotheses, attempts to confirm neural overlap between toolmaking and language production are a relatively recent addition to the literature.[9] In their 2008 study, Stout et al. took three right-handed, expert Early Stone Age toolmakers and used fluorodeoxyglucose positron emission tomography to identify the areas of the brain recruited by three separate tasks: (i) a control task, which involved striking cobbles

[9]See, for instance, Higuchi et al. (2009), Putt et al. (2017), Putt (2019), Stout and Chaminade (2007), Stout et al. (2008), Stout (2011), Stout and Chaminade (2012), and Stout et al. (2021).

together without attempting to produce tools; (ii) Oldowan toolmaking and (iii) Late Acheulean toolmaking (Stout et al., 2008).[10] In this section I provide a summary of the results and implications of this work, as well as a more general overview of Early Stone Age toolmaking and its cognitive demands.

2.1 | Oldowan and Late Acheulean toolmaking

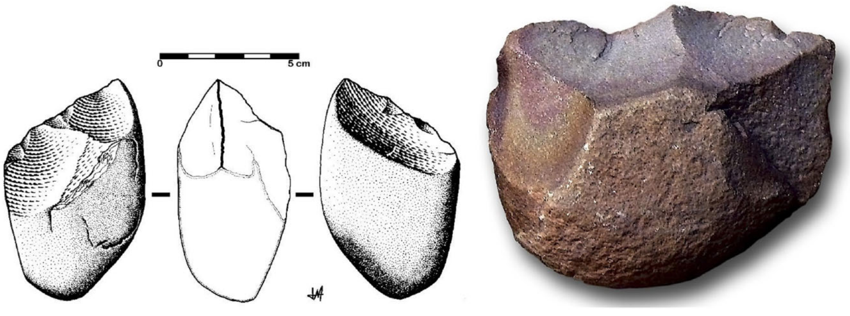
The Early Stone Age runs from roughly 3.3 million years ago (mya) to around 300 thousand years ago (kya). It is typified by three stone tool industries: the *Lomekwian*, the *Oldowan* and the *Acheulean*. The Lomekwian is the earliest known industry, is associated with *Australopithecus*, and predates the first *Homo* fossils by around 500,000 years. The Oldowan spans 2.6–1.7 mya, and is associated with late *Australopithecus* and early *Homo* variants (particularly *Homo habilis*). The Acheulean spans 1.7 mya–300 kya, and is associated with *Homo erectus* and *Homo heidelbergensis*. Toolmaking processes across the three industries become more complex and refined over time. In particular, a distinction between Early and Late Acheulean production is often made. Late Acheulean tools are thought to emerge between 600 and 500 kya, and are smaller, thinner and more symmetrical than tools of the Early Acheulean (Stout, 2018, p. 263), (Stout et al., 2014, p. 577). The Early Stone Age also sees a significant increase in hominin brain size, from 450cm^3 to 1200cm^3 (Klein, 2009). There is thus a general trajectory of increasingly sophisticated toolmaking and expanding brain size across this period. Stout and colleagues' work focuses on the Oldowan and Late Acheulean industries, and I will do likewise.

Oldowan toolmaking involves using a *hammer-stone* to strike a *core*. This process produces *flakes*, which typically have a very sharp cutting-edge. It also shapes the core into a *chopper* (see Figure 1). Typically, reduction of the core via flaking is not comprehensive, and so large parts of the original surface of the core are preserved. At a minimum, Oldowan toolmaking requires refined visuomotor coordination and the ability to evaluate the physical properties of stone. The latter might be thought of in terms of *causal*, or perhaps *technical* reasoning.[11] More refined Oldowan toolmaking plausibly utilises strategic planning. For instance, repeated striking leaves "scars" on the core, which produce better conditions (a *striking platform*) for future strikes. Flake removal can thus be performed in a manner that aids future flake removal (Pargeter et al., 2020; Stout et al., 2019).

Late Acheulean toolmaking requires a significant increase in strategic planning capacities. The archetypal artefact of this industry is the *hand-axe*; a pear-shaped tool characterised by its bifacial, symmetrical structure (see Figure 2). Creating this tool requires using a series of overlapping and increasingly refined procedures to impose a pre-determined form onto a core. Reduction via flaking is more comprehensive as compared with the Oldowan, and typically none of the original surfaces of the core will be preserved. Stout et al. (2006) divide Late Acheulean toolmaking into three phases: *roughing-out*; *primary thinning and shaping*; and *secondary*

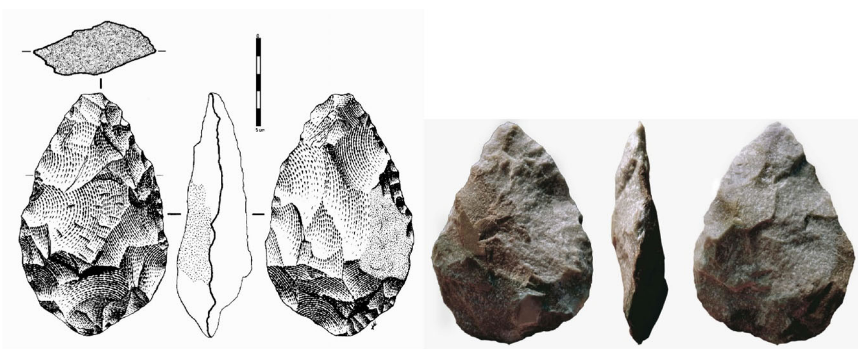
[10]There are complicated questions that arise when the neuroscience of modern humans is used to support inferences about the cognitive capacities of ancient hominins. Addressing such questions is beyond the scope of this paper; but, at the very least, Stout and colleagues' results offer some general proof of principle for co-evolutionary theories. If, for instance, no neural overlap was found, this would not be the case.

[11]See Khemlani et al. (2014), Spellman and Mandel (2006) for accounts of causal reasoning; see De Oliveira et al. (2019), Osiurak and Reynaud (2020) for accounts of technical reasoning; see Moore (2011) for an account of the specific causal relations Early Stone Age knappers may have understood.



(a) (b)

FIGURE 1 Oldowan tools. Oldowan toolmaking involves striking a *core* with a *hammer-stone* in order remove *flakes*. Both the resulting flakes and modified core (a *chopper*—(a) and (b) above) can then be used. Core reduction is usually partial, so large parts of the original surface are preserved. *Source*: José-Manuel Benito Álvarez/Wikimedia commons. Image (a) is reproducible under the terms of the Creative Commons Attribution-Share Alike 2.5 Generic licence; image (b) has been released into the public domain. URL (a) https://en.wikipedia.org/wiki/Stone_tool#/media/File:Chopping_tool.gif. URL (b) https://en.wikipedia.org/wiki/Oldowan#/media/File:Canto_tallado_2-Guelmim-Es_Semara.jpg



(a) (b)

FIGURE 2 Acheulean tools. Acheulean hand-axes are characterised by their bifacial structure and symmetry. Producing them requires performing a sequence of overlapping procedures in order to impose a predetermined form onto a core. Core reduction is usually comprehensive, and typically none of the original surfaces of the core will be preserved. *Source*: José-Manuel Benito Álvarez/Wikimedia commons. Image (a) has been released into the public domain; image (b) is reproducible under the terms of the Creative Commons Attribution-Share Alike 2.5 Generic licence. URL (a) https://en.wikipedia.org/wiki/Stone_tool#/media/File:Hand_axe_spanish.gif. URL (b) https://en.wikipedia.org/wiki/Acheulean#/media/File:Bifaz_cordiforme.jpg

thinning and shaping. During the roughing-out phase, a hammer-stone is used to impose the general form of the hand-axe onto the core. This involves creating a sharp edge around the perimeter of the core, centred between the two faces, by removing large, long flakes. The primary thinning and shaping phase is characterised by the removal of long, thin flakes which thins the biface and starts to impose the desired symmetry on the tool. This requires the creation of striking platforms through intensive light flaking along the edge, and also the use of a *soft-hammer*; that is, a hammer made of antler, bone or wood. In the secondary thinning and shaping stage, a soft-hammer, along with abrasion and grinding, is used more intensively to create a sharp, regular edge around the perimeter of the tool. In addition to those required for Oldowan toolmaking, the cognitive demands of producing Late Acheulean tools include: (i) more fine-grained motor control; (ii) better understanding of the physical properties of stone, particularly with respect to the effects of different striking techniques and (iii) greater strategic planning, particularly with respect to understanding the link between individual actions (striking, grinding, etc.), short-term goals (creating a striking platform) and long-term goals (imposing symmetrical form on to the core).

2.2 | The neural correlates of Early Stone Age toolmaking and language

We would predict Late Acheulean production to have different neural correlates as compared to Oldowan production (and the control task). Stout and colleagues found significant overlap in early visual processing in the posterior occipital cortices across Oldowan and Late Acheulean tasks (Stout et al., 2008).[12] This is what would be expected given the base-level complexity of core and hammerstone manipulation required for each task. In addition, both toolmaking procedures placed increased demands on an inferior parietal-ventral premotor circuit that is anatomically and computationally similar to circuits involved in phonological processing (Stout & Chaminade, 2012, p. 81).

However, Late Acheulean toolmaking produced a marked increase in right hemisphere activity, particularly in supramarginal, ventral precentral and inferior prefrontal gyri. This is likely a product of the increasing length and complexity of the motor actions involved, and of the more technical role played by the left hand in manipulating the core. In addition, hand-axe production saw an increase in activity in the right inferior and ventrolateral prefrontal cortex. The former is thought to be at the apex of sensory and motor processing, and hence play a central role in producing goal-directed and flexible behaviour. The latter is thought to play a role in matching perceptual cues with action sequences (Stout et al., 2008, p. 1946). More broadly, the activation of inferior frontal gyrus in Late Acheulean toolmaking is of interest. The region forms part of Broca's area, and so is associated with language production and, increasingly, the general computation of supramodal and tree-like hierarchical information (Fadiga et al., 2009; Fedorenko et al., 2012; Koechlin & Jubault, 2006; Poldrack, 2006).[13] Also of interest is the activation of the right and left dorsal portions of inferior frontal gyrus *pars triangularis*. The former is thought to play a role in syntactic/semantic integration, and in processing high-level motor representation and complex action-plans. The latter is associated with working memory and sentence processing (Elmer, 2016; Makuuchi et al., 2009; Matchin, 2018).

2.3 | Summary

Stout and colleagues' produce two findings important for my goals. First, the increasing complexity of toolmaking from Oldowan to Late Acheulean triggers activity in areas commonly

[12]See Stout et al. (2021) for further corroboration of these results.

[13]An important issue here is lateralisation, as language processing appears to be more left lateralised than manual praxis tasks. Stout (2018, p. 264) provides some discussion of why this might be the case, including that it represents increasing developmental accommodation to linguistic and technological environments; a theme I will return to (thanks to an anonymous reviewer for raising this issue).

associated with language production. Second, activity in inferior frontal gyrus in both Late Acheulean toolmaking and language production suggests that it plays a central role in processing information associated with behaviour displaying nested part-whole structure (I will go into further detail on this in Sections 3 and 4). This corroborates claims that Broca's area more broadly is important in the production of complex intentional action (Fedorenko et al., 2012; Koechlin & Jubault, 2006). These results lend weight to tool-language co-evolutionary hypotheses, insofar as the emergence of a sophisticated new goal-oriented behavioural phenotype—language—could be supported by a pre-existing domain-general mechanism that had evolved to support a pre-existing complex, goal-oriented behavioural phenotype—toolmaking.

3 | AN ERROR MINIMISATION MODEL FOR LATE ACHEULEAN TOOLMAKING

With the details of Stout and colleagues' work in mind, we can now move to the task of integrating their results with the framework of predictive processing. I do this in two moves. First, I outline the basic computational details of predictive processing. Second, I propose a minimal model for understanding the cognitive demands of Late Acheulean toolmaking based on those details.

My general strategy here was laid out in the introduction: Both complex intentional action and language production are thought to require goal representations with nested part-whole structure. As a result, any information processing overlap between toolmaking and language production may be thought of in terms of the computational mechanisms required to produce structured goal representations. The idea, then, is that we can explain the early evolution of syntax by appeal to more domain general capacities—namely, those required for the broader class of complex intentional action.[14]

Predictive processing offers an account of perception and action using a single computational principle. Moreover, as we shall see, the stratified hierarchical nature of the theory provides a way of understanding how lower-level sensorimotor representations are modulated by higher-level goal representations. For these reasons, it provides an attractive framework for understanding the cognition underlying complex intentional action.

3.1 | The predictive processing framework

There are plenty of excellent overviews of predictive processing (Clark, 2013, 2016; Hohwy, 2013; Wiese & Metzinger, 2017). Here I focus on two aspects of the theory that are important for the explanation of complex, goal-directed action trajectories: (1) hierarchical prediction and prediction error; and (2) prediction error minimisation. But before I begin, an important caveat. Predictive processing, in some of its formulations, has been claimed to offer the cognitive sciences a unifying principle governing brain function.[15] None of my conclusions here are predicated on such claims. All I borrow from the framework is the notion that perception and action are governed by prediction error minimisation in generative hierarchical models.

[14]In adopting this general line I borrow from Planer and Sterelny (2021). I will say more about the differences between our respective approaches in Section 4.

[15]See Friston (2010) and Hohwy (2013) for strong versions of this claim. Clark (2013, Section 5.2) is more cautious.

3.1.1 | Stratified hierarchical prediction and prediction error

Predictive processing understands top-down processing as the transfer of *predictions*, and bottom-up processing as the transfer of *prediction errors*.

Predictions are a cognitive systems' attempts to predict future sensory input. They are generated according to encoded models of the world, which in turn are the product of evolution, experience and learning. These models allow the brain to generate a hypothesis about the source of sensory input, which can then be used to produce predictions about future sensory input. The models are hierarchically organised according to the spatiotemporal scale of the causal regularities they address. At lower levels of the hierarchy there are models generating predictions at finer-grained spatiotemporal scales; for instance, regarding the sensory input associated with adjusting the force of a hammer-stone strike in response to scars on a core. At higher levels of the hierarchy there are models generating predictions at broader spatiotemporal scales; for instance, regarding the sensory input associated with assessing the properties of a potential core. In this way, cognitive systems work to produce accurate predictions of future sensory input over a hierarchy of spatiotemporal grain.

Bottom-up processing also plays an important role in the predictive processing framework. But rather than being the construction of perceptual experience from the raw data of sensory input, bottom-up processing is instead the transfer of any difference between top-down predictions of sensory input and actual sensory input. In other words, bottom-up processing involves the transfer of *prediction error*. At any given layer in the hierarchy, a model will receive prediction error signals from the model below it, attempt to explain away this error, and forward any error that it cannot explain to the model above it. Linked probabilistic models of this sort are called *generative models* due to their ability to recreate incoming sensory states via top-down processing (Hinton, 2007).

In sum, predictive processing envisages global interaction between top-down and bottom-up processing. Top-down processing involves the flow of increasingly spatiotemporally constrained predictions from more abstract to less abstract representational layers; bottom-up processing involves the transfer of prediction error signals—mismatch between the predicted and actual signal—from less abstract to more abstract representational layers.

3.1.2 | Perception, action and prediction error minimisation

According to predictive processing, cognitive systems attempt to minimise the difference between predicted sensory input and actual sensory input. In other words, they attempt to *minimise prediction error*. The brain has two strategies for eliminating any prediction error rising through the hierarchy: it can either change the parameters of its models to account for the error, or else it can change the sensory input causing the error such that it aligns with the current parameters of its models. In simpler terms, the system either adjusts the model to fit the data or adjusts the data to fit the model. Perception is the product of adopting the model-to-data direction of fit; action is the product of adopting the data-to-model direction of fit. This yields a unified computational account of perception and action: Although the direction of fit changes, both phenomena are strategies for minimising prediction error.

In the case of perception, models generate a new hypothesis regarding the source of sensory input. This produces new predictions, which, if successful, will have the effect of dampening any prediction error rising through the hierarchy. For example, suppose I catch a glimpse of a figure at the end of my room from the corner of my eye. On closer inspection, however, said figure is revealed to be a pile of clothes on a chair. The best explanation of the initial (quite patchy) sensory input is that there is a person in my room. Yet this explanation does not account for the data produced when I focus my attention on the figure. As a result, prediction error rises through the system. By producing a new explanation of the source of the sensory input—the untidy chair hypothesis—the brain can eliminate this prediction error.

In the case of action, the brain holds its hypotheses about the world fixed. Any prediction error rising through the system is then addressed by moving around in the world such that the incoming sensory data is changed to fit the hypothesis. To continue the previous example, the hypothesis that there is a person in my room generates the prediction that by turning my head I will be able to get a better look at them. That prediction will be inconsistent with the current, patchy sensory input emanating from the corner of my eye. This mismatch between predicted sensory input and actual sensory input generates prediction error, but by moving my head in order to look to the corner of the room, I can minimise this error.

Here the brain actively tests a hypothesis against the world. As Clark notes, this account of action is, in a certain sense, *subjunctive* (Clark, 2016, p. 121). Motor control involves predicting future proprioceptive and exteroceptive input were a certain action to be carried out, and reducing the gap between those predictions and sensory input by performing the modelled action. The mechanism of prediction error minimisation works because prediction errors carry information about the difference between the way the body is currently situated and the way the body would be situated were an action to be carried out. Implementing that action sequence thus eliminates prediction error. Importantly, the predicted proprioceptive and exteroceptive input of an action is generated across all levels of the hierarchy. At lower levels, models will be predicting sensory input over the short-term; for instance, the proprioceptive change associated with the action of turning my head. At higher levels in the hierarchy, models will be predicting sensory input over the long-term; for instance, the exteroceptive input associated with identity of the person.[16]

Moreover, each level of the hierarchy recapitulates this general procedure in the following way. Just as the system as a whole uses its active states to influence the world in ways that alter its sensory states, each model in the hierarchy uses prediction—here equivalent to active states—to influence the layer below it in ways that will alter incoming prediction error, and thus influence the model's sensory states (Kirchhoff et al., 2018, pp. 3–4). This top-down effect on lower levels is also described in terms of "modulation" or "guidance". In Clark's words, higher level predictions "guide and nuance lower level response. This guidance involves expectations concerning the most likely patterns of unfolding activity at the level below it" (Clark, 2016, p. 146). As such, predictive processing architectures display a kind of transitivity. If level A predicts and guides the activity of level B, and level B predicts and guides the activity of level C, then level A has, to some extent, predicted and guided the activity of level C.[17] Consequently, higher-level goal representations are able to predict, and hence guide, the predictive outputs of the layers below them. And reducing error against this cascade of predictions will bring about the original goal. As I will emphasise in the following sub-section, such architectures produce representations with nested part-whole structure.

[16]This general picture of multi-level processing according to spatio-temporal grain integrates well with the notion of hierarchical process memory (Hasson, 2017; Hasson et al., 2015). Thanks to an anonymous reviewer for this pointer.
[17]See Drayson (2017) for more nuanced discussion.

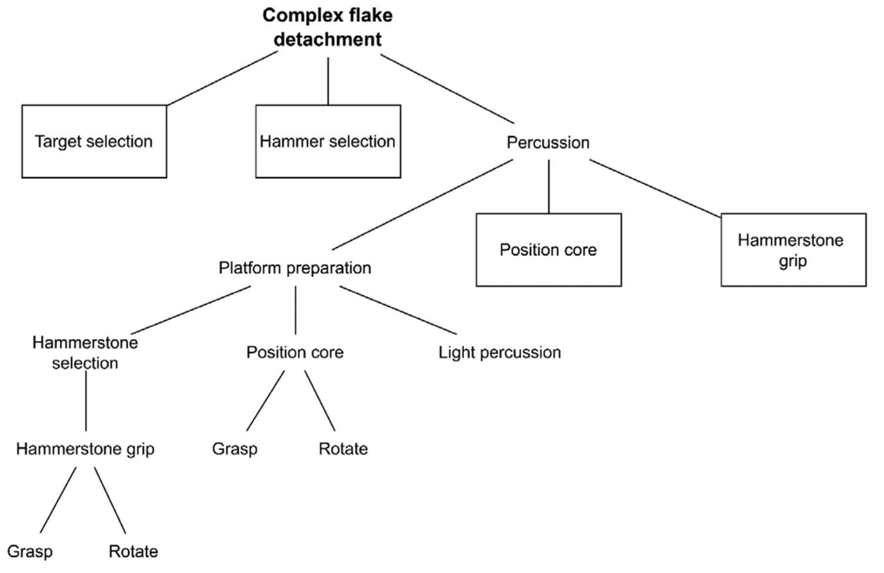


FIGURE 3 An action plan depicting complex flake detachment in Late Acheulean toolmaking. Lines connect subordinate goals/actions with broader goals/actions. Boxes enclose action chunks that have been "collapsed", and whose subordinate elements have been omitted to avoid crowding. Adapted from (Stout, 2011, p. 1052)

3.2 | An error minimisation model for Late Acheulean toolmaking

Stout and colleagues' data suggests that the neural overlap between toolmaking and language (specifically in inferior prefrontal gyrus and Broca's area more broadly) may indicate that the region underwrites the production of complex intentional action. My task now is to outline how these behaviours might be implemented using a prediction error minimisation model.

We have seen that Early Stone Age toolmaking—and, in particular, Late Acheulean toolmaking—is a complex goal-oriented task requiring significant strategic planning and technical skill. The nested part-whole structure of these tasks have been outlined in some detail by Stout, using tree-like hierarchies (see Figure 3). This work is important for my purposes, as it outlines the complexity of the action plans required to produce a tool, rather than the complexity of the tool itself. Indeed, Stout (2011) argues that focusing on the latter has meant we have underestimated the level of cultural variation across the Early Stone Age.[18] Transposing Stout's action plans into the error minimisation framework gives us a potential reading of the cognition of those toolmakers. The results indicate that the Early Stone Age signals the presence of representations with sophisticated nested part-whole structure. I take my proposal to be broadly consistent with a range of more detailed work applying generative hierarchical models of cognition to complex intentional action in general, and to language in particular (Farmer et al., 2013; Koechlin & Jubault, 2006; Lupyan & Clark, 2015; Pickering & Clark, 2014; Pickering & Gambi, 2018; Pickering & Garrod, 2013); though none of these accounts deal specifically with the demands of stone toolmaking.

[18]See also (Stout et al., 2021).

The basic mechanics of the proposal are as follows. Consider the action trajectory required for complex flake detachment, outlined in Figure 3. We can transpose this into the error minimisation framework as follows. First, a goal representation, <complex flake detachment>, becomes salient in the upper levels of the model hierarchy. This produces a prediction of the sensory input expected were this goal to be brought about, which modulates the structure of goal representations in the models below it; to <target selection>, <hammer selection> and <percussion>. These predictions select further specific goal representations further down the hierarchy. For example, from <percussion> to <platform preparation>, and so on. Lower-level goal representations, such as <hammerstone selection>, <position core>, and <light percussion>, are nested within—or constituents of—<platform preparation>. And all of these are constituents of <percussion>, which itself is a constituent of <complex flake detachment>.

Each new goal representation within a model produces predictions, which in turn modulate the goal representations of models below it, right down to predictions of the proprioceptive input expected from fine-grained motor outputs such as <grasp> and <rotate>. The initial goal thus produces a set of structured representations running down the generative hierarchy. If these goal representations are held fixed, the system will encounter a corresponding cascade of prediction errors running back up the hierarchy. However, by moving in such a way as to reduce these errors, the motor system will produce behaviours that achieve the initial goal of <complex flake detachment>.

During the performance of an action plan, new goal representations will be selected at various levels in the hierarchy. For instance, the system might select <position core> and deselect <hammerstone grip>. When this new goal representation becomes salient, a new set of predictions is produced. These then remodulate the goal representations in models beneath it, such that <grasp> and <rotate> become goal representations regarding the core, which in turn modulate further fine-grained proprioceptive predictions further down the hierarchy. This new set of structured representations will produce a new set of corresponding prediction errors. By reducing this error through action, and by repeating this process, the system can reiterate a behaviour.[19]

So the combination of top-down and bottom-up transfer of predictions and prediction errors between hierarchical layers of models, and the presence of transitively linked goal representations, produces representations with nested part-whole structure. And this is a product of the fact that creating a Late Acheulean hand-axe is a particularly complex, goal-oriented activity requiring significant strategic planning abilities.

At this stage it is important to make clear what I am—and what I am not—claiming via this account. First, I am not claiming that structured representations emerged in the Late Acheulean. Goal representations with some level of nested part-whole structure were required to produce both the Early Acheulean and the Oldowan (though with much less sophistication in the case of the latter). Indeed, Planer and Sterelny (2021, pp. 121–126) argue that structured representations are found in the cognitive profiles of the great apes and some primates, and I agree with them. What I am claiming is that the production of stone tools—and in particular, given the significant increase in strategic planning capacities required, the Late Acheulean—drove a

[19]An interesting question following from this is how a new goal representation becomes salient. There are at least two ways an account like mine may go on this issue, depending on how unifying one takes the error minimisation framework to be. On one hand, some theorists think the error minimisation can do the explanatory work—Hohwy and colleagues (Corcoran et al., 2020; Hohwy, 2013) appear to think this. On the other hand, we might posit some other computational mechanism to account for the raising of a new goal representation. I suspect the latter move is the better bet, but elaborating those details is beyond the scope of this paper.

marked increase in the complexity and sophistication of pre-existing representational capacities with nested part-whole structure. This echoes Stout's work emphasising the increasing complexity of action plans across the Early Stone Age (Stout, 2011; Stout et al., 2021). Second, I am not claiming that the above framework provides a computational account of the cognitive capacities required to produce the language abilities we find in modern day humans. Proponents of tool-language co-evolutionary theories are often accused of failing to provide such an account,[20] despite the fact that doing so would actually negate any need for a co-evolutionary approach. If we were to find that all the cognitive capacities required for modern day language abilities were present in the Early Stone Age, then there would be no need to make the claim that a new behaviour was able to co-opt cognitive resources that had evolved for a pre-existing behaviour. The claim is not that the cognitive underpinnings of toolmaking and language are the same. Rather, what is required from a co-evolutionary theory is a stipulated set of capacities, that are importantly *not* equivalent to those required to produce language, but which are nonetheless capable of being acted on by selection to bring about full-blown language abilities. My point is that the above model offers an account of this kind. In particular, the level of planning and motor skill required to produce Late Acheulean tools signals the emergence of representations displaying sophisticated nested part-whole structure, and these are an important ingredient in explaining the evolution of language.

An important related issue concerns a dis-analogy between the way language and toolmaking are learnt. Becoming an expert Early Stone Age toolmaker takes a long period of time and significant focused mental effort.[21] Sequential steps are reinforced over many repetitions, and the fast, fluid, tree-like hierarchical action performed by experts is the end product of a prolonged, iterated learning process. By comparison, it is often claimed that children appear to learn language almost effortlessly.[22] This might suggest that, in modern humans, sequential processing is required to learn how to produce a hand-axe, whereas tree-like hierarchical processing is employed from the outset by children learning to speak.

The above observation can be easily accommodated by the error minimisation account. To reiterate, the claim is not that the cognitive capacities required to produce language and stone tools are the same; rather it is that language evolution involved co-opting and refining some pre-existing cognitive resources that had evolved for toolmaking. Furthermore, there is plenty of scope between the Early Stone Age and today for syntax specific, tree-like hierarchical capacities to have evolved in response to increasingly language rich cultural environments. The differences in the way language and toolmaking are learnt may well be a product of the fact that domain specific mechanisms have evolved with respect to the former, but not the latter. It may even be the case that, in modern-day humans, sequentially learnt action plans exploit linguistic cognitive resources. All this is perfectly consistent with the claim that those syntax specific capacities had an evolutionary precursor in the structured representations required to produce stone tools; which is to say that it is perfectly consistent with a co-evolutionary error minimisation account.[23]

[20]For instance, Berwick and Chomsky (2017, p. 171) and Botha (2020) have criticised Stout and colleagues on such grounds. Gabrić (2021) makes a similar point to mine regarding the inadequacy of such criticisms.

[21]For instance, Stout reports that it took him 300 hours of practice to produce a Late Acheulean hand-axe that equalled the skills of those produced by ancient toolmakers at the famous Boxgrove site in England (Stout, 2016, p. 33).

[22]My own view is that there is considerable evidence that language learning in childhood is far from "effortless" (Chater & Christiansen, 2018; Yang & Piantadosi, 2022). However, I want to show here that the error minimisation view can accommodate both sides of this debate.

[23]Thanks to an anonymous reviewer for pushing me on this issue.

Finally, it is worth saying something about how the predictive processing account relates to traditional, Chomskyan computational models. There are at least two ways this could go. On one hand, some theorists believe the two frameworks can be fruitfully combined. For instance, Fitch's account of tree-like hierarchical cognition builds in aspects of both predictive processing and Chomskyan computation (Fitch, 2014; Fitch & Martins, 2014). On his view, integrating the two frameworks provides a route to a broad theory of neuro-computation; one that does justice to the fact that we are animals, and also to the fact that we are animals with a peculiar set of capacities. Key here is the claim that humans have an advanced (but not species-specific) propensity to infer tree-like hierarchies from sequential data. The programme outlined in this article fits well with Fitch's account. In particular, we can see how a phylogenetically diverse trait—the control of action via error minimisation—might have provided the right kind of machinery that could be evolved, via toolmaking, into the more advanced tree-like hierarchies that are typically thought to underly language production.

On the other hand, some theorists are sceptical that Chomskyan computational models can be neurobiologically realised. Petersson, Folia and Hagoort, for instance, argue that the infinite memory resources required by Chomskyan computation is in contradiction with the fact that the brain is a finite-state system (2012). As a result, they think it "seems natural to try to understand language acquisition and language processing in terms of adaptive stochastic dynamical systems" and "syntax processing in terms of noisy spiking network processors" (2012, p. 93). The error minimisation account is an attractive package for theorists of this stripe for two reasons. First, predictive processing is comparatively well-grounded in empirical neuroscience (Fitch, 2014, p. 343). Second, the probabilistic neural networks of predictive processing lend themselves well to interpretation in terms of "stochastic dynamical systems" and "noisy spiking network processors". As such, the framework might be fruitfully developed by those who reject Chomskyan computation.[24]

3.3 | Summary

In this section, I proposed an error minimisation model for understanding the action sequences required to produce Late Acheulean tools. Stout and colleagues have demonstrated the neuroanatomical overlap between toolmaking and language. The model I propose outlines a plausible computational overlap between the two behaviours: namely, the requirement for complex structured representations. As language production is one instance of the broader category of skilled intentional action, the practice of toolmaking may well have produced the right kind of cognitive platform required to support the production of language.

4 | THE EVOLUTION OF LANGUAGE

In this section, I situate my proposal within broader debates concerning the evolution of language. I begin by showing that the error minimisation account supports gradualism—the view that language evolved via incremental steps over a long period of evolutionary time—and rejects saltationism—the view that language appeared via sudden random mutation. I then

[24]Including Petersson et al. (2012), but also others, for example, Christiansen and Chater (2016), and Rodriguez (2001). Thanks to an anonymous reviewer for helpful comments here.

emphasise an important feature of the account: It can explain the evolution of structured representations without positing a corresponding transition from sequential information processing to tree-like hierarchical information processing. This, I argue, makes it better suited to gradualism than some other gradualist proposals; in particular that offered by Planer and Sterelny (2021).

Saltationism is typified by Berwick and Chomsky (2016, 2017, 2019). In their view, the ability to produce syntax is the result of a random genetic mutation that produced a neural "re-wiring". This event happened comparatively recently—some 80–120 kya. Consequently, among the hominin genus, the cognitive capacities required to produce syntax are restricted to *Homo sapiens*. In contrast, the error minimisation account suggests that the cognitive capacities driving the early evolution of our syntactical abilities were not specific to either language or *H. sapiens*. Rather, those capacities are to be found at low levels in any organism capable of intentional action with nested part–whole structure. They have, however, reached particularly high levels of sophistication in the hominin lineage—notably from *Homo erectus* through *Homo heidelbergensis* to *H. sapiens*—in part through the co-evolutionary effects of toolmaking and language production. This is significant, as the capacity to produce syntax is often thought to be a—if not *the*—distinctive, domain specific feature of the language faculty (Hauser et al., 2002). Language production utilises many capacities that play a broader role within cognitive systems; for instance, memory, executive control and theory of mind. While these are integral to the production of language, they are not language specific. The capacities underwriting the production of syntax, however, are plausibly language specific. The account developed here suggests that evolution of syntax was able to co-opt pre-existing capacities—structured representations—that had been increasingly refined through toolmaking. As such, the error minimisation approach supports gradualism.

Indeed, the gradualist credentials of the error minimisation model run deeper than this. Consider the following three competing views. As we have seen, Berwick and Chomsky (2016) are saltationists: They hold that explaining the complex nested part–whole relations we find in language commit us to structured representations, and that there are no plausible intermediary stages between representations without structure and representations with structure. So language cannot have evolved incrementally, and instead must have appeared suddenly. Frank et al. (2012) believe that parsimony and evolutionary continuity considerations make saltationism unattractive, and hence want an account that avoids sudden transitions. Their solution is to reject the claim that language production requires structured representations—instead, they develop an account of syntax based on sequential processing. Planer and Sterelny (2021) take a different line. On their reading, some basic nested part-whole structure is present in the behavioural and learning profiles of the great apes and some other primates. Thus there is no need to posit any transition from non-structured to structured representations between the last common ancestor with the chimpanzee and modern day humans; all that is required is the incremental sophistication of pre-existing capacities across that time-frame (some 7 million years).

These various accounts are attempting to balance two desiderata. The first is evolutionary continuity, which motivates a reluctance to posit any sudden transition to structured representations. The second is the notion that language production requires structured representations. Both Berwick and Chomsky and Frank and colleagues see these two desiderata as mutually incompatible. The former are committed to both the psychological reality of generative grammars and structured representations, and to the view that a gradualist account of such mechanisms is impossible. A sudden transition is thus an unavoidable aspect of their theory. For the latter, evolutionary continuity concerns outweigh the commitment to language production requiring structured representations. They thus attempt to generate a sequential account of the production of syntax. In developing their gradualist theory, Planer and Sterelny neatly sidestep this dilemma by decoupling the evolution of structured representations from the evolution of language, insofar as the former long predates the latter. However, this does not rid them of the need to posit a transition—albeit gradual—from representations without structure to representations with structure; rather, they have pushed that transition further back in phylogeny. And accounting for *that* transition, on their account, requires positing a corresponding transition from sequential information processing to tree-like hierarchical information processing.

The error minimisation account provides a different story; one that offers a more straightforward fit with the demands of gradualist explanation. This is because predictive processing architectures can account for the transition between non-structured representations and structured representations without needing to posit a corresponding transition in types of information processing. On the error minimisation account, structured representations evolve via a gradual increase in the quantity and complexity of models in a cognitive system; not via the emergence of a new type of processing. In what remains of the article, I will develop this point a bit further.

Planer and Sterelny tie the evolution of behaviour displaying nested part-whole structure to the evolution of Broca's area (2021, p. 141). In their view, the development of this area in the great apes, our own species and some other primates—particularly macaques (Ferrari et al., 2003; Rizzolatti et al., 1996)—represents the emergence of a new type of information processing: tree-like hierarchical processing. Such capacities are capable of producing structured representations, which in turn are required to get complex goal-oriented behaviours like toolmaking and language up and running. However, complex goal-oriented behaviours are not confined to primates. For instance, consider dolphin hunting techniques (Gazda et al., 2005; Krützen et al., 2014), tool use among New Caledonian Crows (Bugnyar, 2019; Miller et al., 2020; Taylor et al., 2009), or the ability of Elephant matriarchs to locate water during a drought (Foley et al., 2008). On Planer and Sterelny's account, these behaviours will need to be explained without recourse to structured representations, or they must appeal to convergent evolution. Now, there is no in-principle reason why these explanations are not possible; but some story must be told.

The error minimisation approach avoids this explanatory burden, and offers a more evolutionarily continuous account. On this view, the stratified cognitive hierarchies of organisms displaying very simple, reactive, inflexible behaviour will be composed of a small number of models characterised by sensorimotor representations that issue simple predictions to the motor plant. Systems of this kind will not produce structured representations, as they lack the higher-level, multimodal goal representations required to modulate representations further downstream in the hierarchy. On the other hand, organisms displaying more complicated behaviour will have extended generative hierarchies, with models in the upper levels employing sophisticated, multi-modal representations capable of producing predictions over broad spatiotemporal scales, and capable of structuring the representations in models below it according to specific goals.[25]

So the predictive processing approach allows us to move from inflexible, reflexive behaviour to more complex intentional action via the gradual increase in number and sophistication of models in a stratified hierarchy. As such, there is no need to posit a transition from sequential information processing to tree-like hierarchical information processing; it is error minimisation all the way down. The key transition to be explained on this approach is the transition from

[25]See Corcoran et al. (2020) for a detailed account of how generative hierarchies might be elaborated in this way.

basic, low-level sensorimotor representations to more abstract, multi-modal goal representations. The role that Broca's area plays in the processing of complex goal-oriented actions suggests that its function may lie in processing representations of this latter kind. Explaining this transition in representational sophistication may be a formidable task; but it is a task that Planer and Sterelny must also address.

To sum up: The work of Frank and colleagues' indicates that, when it comes to cognitive processing, there are those who take evolutionary continuity concerns very seriously; to the extent that they are willing to reject the dominant view that language processing requires structured representations. Planer and Sterelny's account offers a gradualist route to the evolution of structured representations, but one that is committed to a transition between types of information processing. The predictive processing account also provides a gradualist route to structured representations, but one that does not require any transition between types of processing. It is an attractive view, then, for those who emphasise evolutionary continuity.

5 | CONCLUSION

My goals in this paper have been twofold. I aimed first to provide an error minimisation account of the cognitive processes underlying Late Acheulean toolmaking. I argued that doing so reveals that the task requires complex structured representations, and that these represent a plausible pre-cursor to the fully formed syntactic capacities found in modern humans. I then situated my account in the literature, and gave some reasons why it might be a better fit with gradualist theories of language evolution than Planer and Sterelny's account.

More broadly speaking, the error minimisation approach offers a distinctly continuous story of the evolution of complex action; it is thus an attractive option for those who prioritise pragmatics and gradualism in the evolution of language.

ACKNOWLEDGEMENTS

Multiple drafts of this manuscript benefitted from extensive feedback from Ron Planer and Kim Sterelny, for which I am very grateful. I would also like to thank Rachael Brown, Bruno Ippedico, Justin D'Ambrosio, Dan Burnston, Dietrich Stout, and a meeting of Colin Klein's group at ANU for helpful comments and discussion.

Open access publishing facilitated by Australian National University, as part of the Wiley-Australian National University agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

There are no data available.

ORCID

Ross Pain https://orcid.org/0000-0002-5354-6301

REFERENCES

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. Cambridge, MA: MIT Press.

Berwick, R. C., & Chomsky, N. (2017). Why only us: Recent questions and answers. *Journal of Neurolinguistics*, *43*, 166–177. https://doi.org/10.1016/j.jneuroling.2016.12.002

Berwick, R. C., & Chomsky, N. (2019). All or nothing: No half-merge and the evolution of syntax. *Public Library of Science Biology*, *17*(11), e3000539. https://doi.org/10.1371/journal.pbio.3000539

Botha, R. (2020). *Neanderthal language: Demystifying the linguistic powers of our extinct cousins*. Cambridge: Cambridge University Press.

Bugnyar, T. (2019). Tool use: New Caledonian crows engage in mental planning. *Current Biology*, *29*(6), R200–R202. https://doi.org/10.1016/j.cub.2019.01.059

Chater, N., & Christiansen, M. H. (2018). Language acquisition as skill learning. *Current Opinion in Behavioral Sciences*, *21*, 205–208. https://doi.org/10.1016/j.cobeha.2018.04.001

Christiansen, M. H., & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. MIT Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford, NY: Oxford University Press.

Corcoran, A. W., Pezzulo, G., & Hohwy, J. (2020). From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition. *Biology and Philosophy*, *35*(3), 32. https://doi.org/10.1007/s10539-020-09746-2

Darwin, C. (2004/1871). *The descent of man, and selection in relation to sex*. London: Penguin books.

De Oliveira, E., Reynaud, E., & Osiurak, F. (2019). Roles of technical reasoning, theory of mind, creativity, and fluid cognition in cumulative technological culture. *Human Nature*, *30*(3), 326–340. https://doi.org/10.1007/s12110-019-09349-1

Drayson, Z. (2017). Modularity and the predictive mind. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958573130

Elmer, S. (2016). Broca Pars Triangularis constitutes a "hub" of the language-control network during simultaneous language translation. *Frontiers in Human Neuroscience*, *10*, 491. https://doi.org/10.3389/fnhum.2016.00491

Engels, F. (2003/1876). The part played by labour in the transition from ape to man. In R. C. Scharff & V. Dusek (Eds.), *Philosophy of technology* (pp. 71–77). London: Blackwell.

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's Area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*(1), 448–458. https://doi.org/10.1111/j.1749-6632.2009.04582.x

Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences*, *36*(3), 211–212. https://doi.org/10.1017/S0140525X12002312

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's Area. *Current Biology*, *22*(21), 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *The European Journal of Neuroscience*, *17*(8), 1703–1714. https://doi.org/10.1046/j.1460-9568.2003.02601.x

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, *11*(3), 329–364. https://doi.org/10.1016/j.plrev.2014.04.005

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*(1), 87–104. https://doi.org/10.1111/nyas.12406

Foley, C., Pettorelli, N., & Foley, L. (2008). Severe drought and calf survival in elephants. *Biology Letters*, *4*(5), 541–544. https://doi.org/10.1098/rsbl.2008.0370

Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, *279*(1747), 4522–4531. https://doi.org/10.1098/rspb.2012.1741

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gabrić, P. (2021). Book review: Neanderthal language: Demystifying the linguistic powers of our extinct cousins. *Frontiers in Psychology*, *12*, 702361. https://doi.org/10.3389/fpsyg.2021.702361

Gazda, S. K., Connor, R. C., Edgar, R. K., & Cox, F. (2005). A division of labour with role specialization in group–hunting bottlenose dolphins (Tursiops truncatus) off Cedar Key, Florida. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1559), 135–140. https://doi.org/10.1098/rspb.2004.2937

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, *14*(4), 531–551. https://doi.org/10.1017/S0140525X00071235

Hasson, U. (2017). The neurobiology of uncertainty: Implications for statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160048. https://doi.org/10.1098/rstb.2016.0048

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304–313. https://doi.org/10.1016/j.tics.2015.04.006

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Higuchi, S., Chaminade, T., Imamizu, H., & Kawato, M. (2009). Shared neural correlates for language and tool use in Broca's Area. *Neuroreport*, *20*, 1376–1381. https://doi.org/10.1097/WNR.0b013e3283315570

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*(10), 428–434. https://doi.org/10.1016/j.tics.2007.09.004

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Isaac, G. L. (1976). Stages of cultural elaboration in the pleistocene: Possible archaeological indicators of the development of language capabilities. *Annals of the New York Academy of Sciences*, *280*(1), 275–288. https://doi.org/10.1111/j.1749-6632.1976.tb25494.x

Khemlani, S. S., Barbey, A. K., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience*, *8*, 849. https://doi.org/10.3389/fnhum.2014.00849

Kimura, D. (1993). *Neuromotor mechanisms in human communication*. Oxford: Oxford University Press.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, *15*(138), 20170792. https://doi.org/10.1098/rsif.2017.0792

Klein, R. G. (2009). *The human career: Human biological and cultural origins*. Chicago, IL: University of Chicago Press.

Koechlin, E., & Jubault, T. (2006). Broca's Area and the hierarchical organization of human behavior. *Neuron*, *50*(6), 963–974. https://doi.org/10.1016/j.neuron.2006.05.017

Krützen, M., Kreicker, S., MacLeod, C. D., Learmonth, J., Kopps, A. M., Walsham, P., & Allen, S. J. (2014). Cultural transmission of tool use by Indo-Pacific bottlenose dolphins (Tursiops sp.) provides access to a novel foraging niche. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1784), 20140374. https://doi.org/10.1098/rspb.2014.0374

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(20), 8362–8367. https://doi.org/10.1073/pnas.0810928106

Matchin, W. G. (2018). A neuronal retuning hypothesis of sentence-specificity in Broca's Area. *Psychonomic Bulletin & Review*, *25*(5), 1682–1694. https://doi.org/10.3758/s13423-017-1377-6

Miller, R., Gruber, R., Frohnwieser, A., Schiestl, M., Jelbert, S. A., Gray, R. D., Boeckle, M., Taylor, A. H., & Clayton, N. S. (2020). Decision-making flexibility in New Caledonian crows, young children and adult humans in a multi-dimensional tool-use task. *Public Library Of Science One*, *15*(3), e0219874. https://doi.org/10.1371/journal.pone.0219874

Montagu, A. (1976). Toolmaking, hunting, and the origin of language. *Annals of the New York Academy of Sciences*, *280*(1), 266–274. https://doi.org/10.1111/j.1749-6632.1976.tb25493.x

Moore, M. W. (2011). The design space of stone flaking: Implications for cognitive evolution. *World Archaeology*, *43*(4), 702–715. https://doi.org/10.1080/00438243.2011.624778

Moro, A. (2014). On the similarity between syntax and actions. *Trends in Cognitive Sciences*, *18*(3), 109–110. https://doi.org/10.1016/j.tics.2013.11.006

Osiurak, F., & Reynaud, E. (2020). The elephant in the room: What matters cognitively in cumulative technological culture. *Behavioral and Brain Sciences*, *43*, E156. https://doi.org/10.1017/S0140525X19003236

Pargeter, J., Khreisheh, 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

Petersson, K.-M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*(2), 83–95. https://doi.org/10.1016/j.bandl.2010.08.003

Pickering, M., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Clark, A. (2014). Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*, *18*(9), 451–456. https://doi.org/10.1016/j.tics.2014.05.006

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Planer, R. J. (2017). Talking about tools: Did early pleistocene hominins have a protolanguage? *Biological Theory*, *12*(4), 211–221. https://doi.org/10.1007/s13752-017-0279-1

Planer, R. J., & Sterelny, K. (2021). *From signal to symbol: The evolution of language*. Cambridge, MA: MIT Press.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Pulvermüller, F. (2014). The syntax of action. *Trends in Cognitive Sciences*, *18*(5), 219–220. https://doi.org/10.1016/j.tics.2014.01.001

Putt, S. S. (2019). The stories stones tell of language and its evolution. In K. A. Overmann & F. L. Coolidge (Eds.), *Squeezing minds from stones: Cognitive archaeology and the evolution of the human mind* (pp. 304–318). Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/oso/9780190854614.001.0001/oso-9780190854614-chapter-15

Putt, S. S., Wijeakumar, S., Franciscus, R. G., & Spencer, J. P. (2017). The functional brain networks that underlie Early Stone Age tool manufacture. *Nature Human Behaviour*, *1*(6), 1–8. https://doi.org/10.1038/s41562-017-0102

Reynolds, P. (1976). Language and skilled activity. *Annals of the New York Academy of Sciences*, *280*(1), 150–166. https://doi.org/10.1111/j.1749-6632.1976.tb25483.x

Reynolds, P. (1993). The complementation theory of language and tool use. In K. R. Gibson & T. Ingold (Eds.), *Tools, language and cognition in human evolution* (pp. 407–428). Cambridge University Press.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Rodriguez, P. (2001). Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Computation*, *13*(9), 2093–2118. https://doi.org/10.1162/089976601750399326

Spellman, B., & Mandel, D. (2006). Psychology of causal reasoning. In L. Nadel (Ed.), *Encyclopedia of cognitive science*. SAGE. https://doi.org/10.1002/0470018860.s00491

Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions: Biological Sciences*, *366*(1567), 1050–1059.

Stout, D. (2016). Tales of a Stone Age neuroscientist. *Scientific American*, *314*(4), 28–35.

Stout, D. (2018). Archaeology and the evolutionary neuroscience of language: The technological pedagogy hypothesis. *Interaction Studies*, *19*(1–2), 256–271. https://doi.org/10.1075/is.17033.sto

Stout, D., Apel, J., Commander, J., & Roberts, M. (2014). Late Acheulean technology and cognition at Boxgrove, UK. *Journal of Archaeological Science*, *41*, 576–590. https://doi.org/10.1016/j.jas.2013.10.001

Stout, D., & Chaminade, T. (2007). The evolutionary neuroscience of tool making. *Neuropsychologia*, *45*(5), 1091–1100. https://doi.org/10.1016/j.neuropsychologia.2006.09.014

Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philosophical Transactions: Biological Sciences*, *367*(1585), 75–87.

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., 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*, *60*(3), 309–340. https://doi.org/10.1086/703173

Stout, D., Toth, N., & Schick, K. (2006). Comparing the neural foundations of Oldowan and Acheulean Toolmaking: A pilot study using positron emission tomography (PET). In N. Toth & K. Schick (Eds.), *The Oldowan: Case studies into the earliest stone age* (pp. 321–331). Stone Age Institute Press.

Stout, D., Toth, N., Schick, K., & Chaminade, T. (2008). Neural correlates of Early Stone Age toolmaking: Technology, language and cognition in human evolution. *Philosophical Transactions: Biological Sciences*, *363*(1499), 1939–1949.

Taylor, A., Roberts, R., Hunt, G., & Gray, R. (2009). Causal reasoning in New Caledonian crows. *Communicative & Integrative Biology*, *2*(4), 311–312. https://doi.org/10.4161/cib.2.4.8224

Wiese, W., & Metzinger, T. K. (2017). Vanilla PP for philosophers: A primer on predictive processing. In W. Wiese & T. K. Metzinger (Eds.), *PPP—Philosophy and predictive processing*. MIND Group. https://doi.org/10.15502/9783958573024

Yang, Y., & Piantadosi, S. T. (2022). One model for the learning of language. *Proceedings of the National Academy of Sciences*, *119*(5), e2021865119. https://doi.org/10.1073/pnas.2021865119

How to cite this article: Pain, R. (2022). Stone tools, predictive processing and the evolution of language. *Mind & Language*, 1–21. https://doi.org/10.1111/mila.12419