INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Gabora L, Steel M. 2020 A model of the transition to behavioural and cognitive modernity using reflexively autocatalytic networks. *J. R. Soc. Interface* **17**: 20200545.
http://dx.doi.org/10.1098/rsif.2020.0545

Received: 10 July 2020
Accepted: 7 October 2020

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
biomathematics, evolution

Keywords:
autocatalytic network, behavioural modernity, cultural evolution, Hohlenstein–Stadel Löwenmensch figurine, semantic network, Upper Palaeolithic

Author for correspondence:
Liane Gabora
e-mail: liane.gabora@ubc.ca

THE ROYAL SOCIETY PUBLISHING

A model of the transition to behavioural and cognitive modernity using reflexively autocatalytic networks

Liane Gabora[1] and Mike Steel[2]

[1]Department of Psychology, University of British Columbia, Okanagan Campus, Kelowna, BC, Canada
[2]Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand

LG, 0000-0002-0382-7711

This paper proposes a model of the cognitive mechanisms underlying the transition to behavioural and cognitive modernity in the Upper Palaeolithic using autocatalytic networks. These networks have been used to model life's origins. More recently, they have been applied to the emergence of *cognitive* structure capable of undergoing *cultural* evolution. Mental representations of knowledge and experiences play the role of catalytic molecules, the interactions among them (e.g. the forging of new associations or affordances) play the role of reactions, and thought processes are modelled as chains of these interactions. We posit that one or more genetic mutations may have allowed thought to be spontaneously tailored to the situation by modulating the degree of (i) divergence (versus convergence), (ii) abstractness (versus concreteness), and (iii) context specificity. This culminated in persistent, unified autocatalytic semantic networks that bridged previously compartmentalized knowledge and experience. We explain the model using one of the oldest-known uncontested examples of figurative art: the carving of the Hohlenstein–Stadel Löwenmensch, or lion man. The approach keeps track of where in a cultural lineage each innovation appears, and models cumulative change step by step. It paves the way for a broad scientific framework for the origins of both biological and cultural evolutionary processes.

1. Introduction

How did we become distinctively human? What enabled us to develop imagination, ingenuity and complex belief systems? These questions are central to understanding who we are, how we got here and where we are headed. Behavioural and cognitive modernity are thought to have come about between 100 000 and 30 000 years ago, as evidenced by the proliferation in cultural artefacts of both utilitarian and aesthetic value. (Although some researchers argue that the onset of behavioural modernity was less pronounced than once thought [1–3], and the concept of behavioural modernity itself has been called into question [4], this paper does not delve into these discussions so as to focus squarely on the task of modelling the cognitive changes underlying this cultural transition.) Some attribute this transition to an enhanced ability to process social information [5,6]. Cognitive explanations have been proposed; for example, it has been attributed to the onset of conceptual fluidity [7], dual modes of information processing [8,9] or enhanced working memory [10]. We suggest that each of these proposals holds merit and that they are not mutually exclusive, but reflect the onset of a new kind of semantic network structure, which is modelled here.

Although evidence of human culture dates back millions of years, behavioural–cognitive modernity is associated with the transition to cultural change that is not just adaptive (new innovations that yield some benefit for their bearers tend to predominate) but also cumulative (later innovations build on earlier ones) and open-ended (the space of possible innovations is not finite, since each innovation can give rise to spin-offs). In other words, culture became an *evolutionary* process [11–18]. By *culture*, we mean extrasomatic adaptations, including behaviour and artefacts, that are socially rather than genetically transmitted. Although cultural *transmission*—in which one individual acquires elements of culture from another—is observed in many species, cultural *evolution* is much rarer, and perhaps unique to our species.[1]

Critical to cultural evolution is the capacity to combine ideas, adapt existing solutions to new situations and reframe information in one's own terms.[2] This paper uses network theory to address how this capacity arose. Networks allow for a comprehensive understanding of the dynamics of complex entities and their relationships [21]. Network-based approaches to characterizing the kind of cognitive structure that could sustain cultural evolution enable us to address the question of how minds carry out the contextual, combinatorial and hierarchically structured thought processes needed to generate cumulative, adaptive and open-ended cultural novelty [22,23]. Here, to capture the self-organizing, self-maintaining and indeed self-replicating nature of human cognition, rather than a generic semantic or neural network, we use an autocatalytic network. Autocatalytic network theory grew out of studies of the statistical properties of *random graphs* consisting of nodes randomly connected by edges [24]. As the ratio of edges to nodes increases, the size of the largest cluster increases and the probability of a phase transition resulting in a single giant connected cluster also increases. The recognition that connected graphs exhibit phase transitions led to their application to efforts to develop a formal model of the origin of life (OOL); namely, of how abiogenic catalytic molecules crossed the threshold to the kind of collectively self-sustaining, self-replicating structure we call 'alive' [25,26]. In the application of graph theory to the OOL, the nodes represent catalytic molecules and the edges represent reactions. It is exceedingly improbable that any catalytic molecule present in the primordial soup of Earth's early atmosphere catalysed its own formation. However, reactions generate new molecules that catalyse new reactions, and as the variety of molecules increases, the variety of reactions increases faster. As the ratio of reactions to molecules increases, the probability increases that the system will undergo a phase transition. When, for each molecule in a set, there is a catalytic pathway to its formation, the set is said to be collectively *autocatalytic*, and the process by which this state is achieved has been referred to as *autocatalytic closure* [26]. The molecules thereby become a self-sustaining, self-replicating structure (i.e. a living protocell [27]). Thus, the theory of autocatalytic networks has provided a promising avenue for modelling the OOL and thereby understanding how biological evolution began [28]. The approach is consistent with claims for the centrality of transitions across the life sciences [29].

Autocatalytic networks have been developed mathematically and generalized for cross-disciplinary application in other settings in the theory of reflexively autocatalytic food set generated (RAF) networks [30,31]. The term *reflexively* is used in its mathematical sense, meaning that every element is related to the whole. The term *food set* refers to the reactants that are initially present, as opposed to those that are the products of catalytic reactions. RAFs have been used extensively to model the origins of biological evolution [28,30–32]. Thus, one strength of the approach is that by adapting a formalism that has been used successfully to model one evolutionary process to model another, we pave the way for a broad conceptual framework that can shed light on both [22,33]. This is in keeping with the suggestion that autocatalytic networks may hold the key to understanding the origins of *any* evolutionary process, including the origin of culture [15,22,34–36].[3] In application to culture, the products and reactants are not catalytic molecules but culturally transmittable *mental representations*[4] (MRs) of experiences, ideas and chunks of knowledge, as well as more complex mental structures such as schemas and scripts (tables 1 and 2).

Table 1. Application of graph theoretic concepts to the origin of life (OOL) and origin of culture (OOC).

graph theory	origin of life (OOL)	origin of culture (OOC)
node	catalytic molecule	mental representation (MR)
edge	reaction pathway	association
cluster	molecules connected via reactions	MRs connected via associations
connected graph	autocatalytic closure [25,26]	conceptual closure [22]

Another strength of the approach is that because it distinguishes reactants that are external in origin—in our case, MRs that were acquired through social learning or individual learning of *existing* information—from those that are the products of internal reactions—in our case, MRs that come about through the creative generation of *new* information—MRs are tagged with their source. This enables us to model how networks emerge and to trace cumulative change in cultural lineages step by step.

In previous work, we used the RAF framework to model what is arguably the earliest significant transition in the archaeological record: the transition from Oldowan to Acheulean tool technology approximately 1.76 Ma [36,39]. We posited that this was precipitated by the onset of the capacity for *representational redescription* (RR), in which the contents of working memory are recursively restructured by drawing upon similar or related ideas, or through concept combination. This enabled the forging of associations between MRs, and the emergence of hierarchically structured concepts, making it possible to shift between levels of abstraction as needed to carry out tasks composed of multiple subgoals. This culminated in what is referred to as a transient RAF, a critical step towards what has been referred to as *conceptual closure* [34], characterized by the emergence of persistent 'autocatalytic' cognitive structure. The application of RAFs presented in this paper builds on that work to model a pivotal cultural transition in human history that has been referred to as the origins of art, science and religion [40]. We propose that behavioural and cognitive modernity was brought about by the emergence of an autocatalytic semantic network. We first summarize the archaeological evidence for a transition to behavioural modernity in the Upper Palaeolithic. We then present our RAF model of the underlying cognitive transition that brought it about. Finally, we compare and contrast our proposal with existing literature.

Table 2. Abbreviations used throughout this paper.

abbreviation	meaning
OOL	origin of life
OOC	origin of culture
MR	mental representation
RR	representational redescription
RAF	reflexively autocatalytic and food set generated (F-generated)
CCP	cognitive catalytic process

2. Archaeological evidence for behavioural and cognitive modernity

We begin with a brief summary of the evidence for a transition to behavioural and cognitive modernity in the Upper Palaeolithic.[5] Although one can argue that the earliest stone tools marked the onset of a 'proto' form of cultural evolution, following the initial appearance of the Acheulean hand axe, the archaeological record exhibits considerable stasis,[6] and—with the exception of a more sophisticated knapping (Levallois) technique 200 000–400 000 years ago—little in the way of creative embellishment or improvement [43].

This changed in the Aurignacian period of the Upper Palaeolithic, at which point there is evidence of recognizably human ways of living and thinking. The earliest evidence of behavioural and cognitive modernity comes from Africa less than 100 000 years ago, in sub-Himalayan Asia and Australasia more than 50 000 years ago [44] and in Continental Europe until approximately 30 000 years ago [45]. This evidence consists of a proliferation of different complex, task-specific tools including effective cutting blades [2,46]. It also marks the appearance of representational art [47–49], artefacts indicating personal symbolic ornamentation [50], complex living spaces [51], sophisticated ways of obtaining food, including aquatic resources [52], and burial sites indicating ritual [53] and possibly religion [54]. The Upper Palaeolithic is also widely believed to have marked the onset of modern syntactically rich language [55] (though some argue that language arose more gradually; see [56]). In short, this period witnessed an unprecedented dramatic increase in the variety, utility and aesthetic value of cultural outputs.

A celebrated example of Upper Palaeolithic art to which we will devote considerable attention is the Löwenmensch or 'lion man' figurine from the Hohlenstein–Stadel cave in Germany (figure 1). This figurine, carbon dated to the Interpleniglacial period between 35 000 and 40 000 years ago, is one of the oldest-known zoomorphic (animal-shaped) sculptures in the world, and one of the oldest-known examples of figurative art. It measures 31.1 cm, and was carved out of mammoth ivory using a flint stone knife.

Figure 1. Sketch of the Löwenmensch or 'lion man' figurine from the Hohlenstein–Stadel cave in Germany. According to the Ulm Museum, carbon-14 dating puts it at an age of 35 000–40 000 years. The hand indicates its relative size. (Obtained with permission from the artist, Cameron Smith).

3. An underlying cognitive transition

The model of the transition to cognitive and behavioural modernity in the Upper Palaeolithic developed here grew out of the hypothesis that it was due to the onset of *contextual focus:* the capacity to, in a spontaneous and ongoing manner, shift between convergent and divergent modes of thought, thereby tailoring one's mode of thought to one's situation [23,57,58]. Focused attention is conducive to *convergent thought* because the activation of neural cell assemblies is constrained enough to zero in on the most defining properties. In this compact form, the contents of thought are more readily amenable to deliberate executive-level operations. In convergent thought, one can access only *close associates* of the current thought: items that are highly related to it with respect to the most conventional, default context. For example, FIG and PLUM are close associates because they are both fruits, and they are most commonly thought about with respect to their membership in the category FRUIT.

By contrast, defocused attention is conducive to *divergent thought* because it causes diffuse activation of neural cell assemblies in memory, such that obscure (but potentially relevant) properties come into play [34,58–60]. This is useful for creative tasks, and when one is in need of a new approach or innovative solution. Divergent thought may include more details of the current subject of thought, or incorporate related items; one is simply taking in more of the situation and its associations. In divergent thought, one can access *remote associates*: words or concepts that are not related to each other with respect to their most conventional, default context. Highly divergent thought may result in cross-domain thinking, in which ideas from different domains are combined, or a solution to a problem in one domain is borrowed from another domain.

Although divergent thought is useful for escaping local minima, it confers the risk of getting perpetually side-tracked, whereby irrelevant thoughts readily intrude, interfering with survival tasks. Unless the capacity for divergent thought is accompanied by the ability to reign it in, it would be counter-productive. Therefore, it seems likely that in the pre-modern mind, before the advent of the capacity to shift along the spectrum from convergent to divergent, all mental contents were processed convergently, such that each successive thought was a close associate of its predecessor, and remote associates were not accessed.

Contextual focus came about through the onset of the capacity to adjust the focus of attention to current constraints and affordances, making it more focused or diffuse, as needed, thereby stretching or shrinking conceptual space, and tailoring working memory to task demands (or lack thereof, as in mind wandering). Contextual focus made it possible to shift between (i) convergent thought to modify the content of working memory on the basis of close associates when that was sufficient and (ii) divergent thought to usher in remote associates when 'stuck in a rut'. The theory that contextual focus can have a transformative impact on cultural evolution was tested using an agent-based model [23]. Incorporating the ability to shift between convergent and divergent processing modes into neural network-based agents in the agent-based model resulted in an increase in the mean fitness of cultural outputs.

The model that follows builds on the hypothesis that behavioural modernity was due to the onset of contextual focus, but goes further in positing that thought acquired the capacity to shift along a multimodal spectrum through spontaneous tuning of the following three variables.

3.1. Divergence

The first variable is the capacity to shift between divergent and convergent thought, as discussed above.

3.2. Level of abstraction

The second variable is *degree of abstraction*. It has been shown that there is what is called a *basic level* of abstraction (e.g. BIRD, as opposed to ANIMAL or SPARROW) that mirrors the correlational structure of properties in the object's real-world perception and use [61]. Categories form, and are first learned and perceived, at this basic level, before they are further discriminated at the subordinate level (e.g. SPARROW), and abstracted at the superordinate level (e.g. ANIMAL).[7] Since basic-level categories contain the degree of abstraction most useful for carrying out daily activities [61], it seems reasonable that they precede other levels of abstraction not just developmentally but evolutionarily. We posit that the arrival of behavioural/cognitive modernity involved onset of the capacity to shift along the hierarchy from abstract to concrete, thereby identifying relatedness at different hierarchical levels, and incorporating these distinctions into one's mental model of the world. Abstraction provides another means of connecting MRs, but instead of forging a remote association between them, it makes explicit that they are both instances of some more general MR (e.g. LION and MAMMOTH are both instances of ANIMAL). Thus, the second variable involves the capacity to shift from basic-level categories to other levels of abstraction.

3.3. Context specificity

To generate ideas and solutions that are not just new but also task-relevant may require thought that is not just divergent but also context-specific [59]. Thus, the third variable is *context specificity*: the degree to which thought is biased by a specific motivating contextual factor such as a goal or desire [62]. Divergent thought need not *always* be context-specific (e.g. during mind-wandering or writing free verse). However, context-specific divergent thought allows one to access information that is related to the current contents of working memory in ways that may be unconventional yet precisely relevant to the current situation [63]. For example, thinking of lions in the context of wishing for an inspirational reminder of a lion's power might prompt one to modify one's concept of a lion to incorporate the possibility of carving a lion. This unusual context makes this remote yet feasible relationship 'pop out'.

3.4. Multimodality

The spectrum of thought is multimodal, where by a 'mode' we refer to a particular combination of these three variables (e.g. divergent, abstract and context-specific). In short, we posit that, by using this multimodal spectrum to modify how one thought gives way to the next, cognitive processes could be carried out more effectively. Moreover, the fruits of one mode of thought could become ingredients for another mode, thereby facilitating the forging of a richly integrated network of understandings about the world and one's place in it, sometimes referred to as a *worldview* [34,35]. This, we posit, set the stage for behavioural modernity.[8]

Note that although we may be 'wired for culture' [64], and the cognitive changes underlying this cultural transition may have (directly or indirectly; see [42]) involved one or more genetic mutations [10,41,57,65], we are not proposing that control over these variables came online instantaneously, nor that control over each of them arose simultaneously. The challenge may have been not so much to *possess* the capacity to change these variables as to *coordinate* them so as to continuously tune one's mode of thought in response to changing task demands and effectively navigate semantic space. The evolutionary and developmental tinkering required to achieve this could explain the lag between anatomically modern *Homo sapiens* 200 000 to 100 000 years ago and behavioural modernity 100 000 to 30 000 years ago.

4. Autocatalytic networks (RAFs)

The mathematical model we will describe and analyse in this paper is based on the notion of RAFs. Our use of RAFs as an underlying model is based on three considerations: (i) the model has a high degree of generality, which has allowed its application to explain transition events in a variety of fields (as mentioned above), (ii) it has a well-developed mathematical theory, and (iii) in earlier work [36,39] RAFs have provided a way to investigate cognitive processes (and transitions in them in early cultural evolution), which we develop further here in a more complete mathematical model.[9] Thus, in order to explain our approach we first summarize the key concepts of RAF theory.

A *catalytic reaction system* (CRS) is a tuple $\mathcal{Q} = (X, \mathcal{R}, C, F)$ consisting of a set X of molecule types, a set $\mathcal{R}$ of reactions, a catalysis set C indicating which molecule types catalyse which reactions, and a subset F of X called the food set. A *Reflexively Autocatalytic and F-generated* set—i.e. a RAF—is a non-empty subset $\mathcal{R}' \subseteq \mathcal{R}$ of reactions that satisfies the following two properties:

(i) *Reflexively autocatalytic*: each reaction $r \in \mathcal{R}'$ is catalysed by at least one molecule type that is either produced by $\mathcal{R}'$ or is present in the food set F; and
(ii) *F-generated*: all reactants in $\mathcal{R}'$ can be generated from the food set F via a series of reactions only from $\mathcal{R}'$ itself.

A set of reactions that forms a RAF is simultaneously self-sustaining (by the *F*-generated condition) and (collectively) autocatalytic (by the RA condition; as each of its reactions is catalysed by a molecule associated with the RAF). A CRS need not have a RAF, but when it does there is a unique maximal one. Moreover, a CRS, may contain many RAFs, and it is this feature that allows RAFs to evolve, as demonstrated (both in theory and in simulation studies) through selective proliferation and drift acting on possible subRAFs of the maxRAF [30,32].

In the OOL context, a RAF emerges in systems of polymers (molecules consisting of repeated units called monomers) when the complexity of these polymers (as measured by their maximum length) reaches a certain threshold [26,66]. The phase transition from no RAF to a RAF incorporating most or all of the molecules depends on (i) the probability of any one polymer catalysing a given reaction that forms another polymer and (ii) the maximum length (number of monomers) of polymers in the system. This transition has been formalized and analysed (mathematically and via simulations), and applied to real biochemical systems [30,66–69], ecology [70] and cognition [36,39]. RAF theory has proven useful for identifying how phase transitions might occur, and at what parameter values.

4.1. Terminology

We now introduce the mathematical framework and terminology that will be used to model the transition to cognitive modernity. All MRs in a given individual i are denoted X_i, and a particular MR $x = x_i$ in X_i is denoted by writing $x \in X_i$. As in an OOL RAF, we have a *food set*; for individual i, this is denoted F_i. In the OOC context, F_i encompasses MRs for individual i that are either innate or result from direct experience in the world, including natural, artificial and social stimuli. F_i includes everything in the long-term memory of individual i that was not the direct result of individual i engaging in RR. This includes information obtained through social learning from *someone else* who may have obtained it by way of RR. For example, if individual i learns from individual j how to edge a blank flake through percussive action, this is an instance of social learning, and the concept EDGING is therefore a member of F_i.

F_i also includes existing information obtained by i through individual learning (which, as stated earlier, involves learning from the environment by non-social means), so long as this information retains the form in which it was originally perceived (and does not undergo redescription or restructuring through abstract thought). The crucial distinction between food set and non-food set items is not whether another person was involved, nor whether the MR was originally obtained through abstract thought (by *someone*), but whether the abstract thought process originated in the mind of the individual i in question. Thus, F_i has two components:

— $\mathbb{S}_i$ denotes the set of MRs arising through direct stimulus experience that have been encoded in individual i's memory. It includes MRs obtained through social learning from the communication of an MR x_j by another individual j, denoted $\mathbb{S}_i[x_j]$, and MRs obtained through individual learning, denoted $\mathbb{S}_i[l]$, as well as the contents of memory arising through direct perception that does not involve learning, denoted $\mathbb{S}_i[p]$.
— I_i denotes any *innate knowledge* with which individual i is born.

A particular catalytic event (i.e. a single instance of RR) in a stream of abstract thought in individual i is referred to as a *reaction*, and denoted $r \in \mathcal{R}_i$. A stream of abstract thought, involving the generation of representations that go beyond what has been directly observed, is modelled as a sequence of catalytic events. Following [36], we refer to this as a *cognitive catalytic process* (CCP). The set of reactions that can be catalysed by a given MR x in individual i is denoted $C_i[x]$. The entire set of MRs either *undergoing* or *resulting from* r is denoted A or B, respectively, and a member of the set of MRs undergoing or resulting from reaction r is denoted $a \in A$ or $b \in B$.

The term *food set derived*, denoted $\neg F_i$, refers to mental contents that are *not* part of F_i (i.e. $\neg F_i$ consists of all the products $b \in B$ of all reactions $r \in \mathcal{R}_i$). In particular, $\neg F_i$ includes the products of any reactions derived from F_i and encoded in individual i's memory. Its contents come about through mental operations *by the individual in question* on the food set; in other words, food set derived items are the direct product of RR. Thus, $\neg F_i$ includes everything in long-term memory that *was* the result of one's own CCPs. $\neg F_i$ may include an MR in which social learning played a role, so long as the most recent modification to this MR was a catalytic event (i.e. it involved RR).[10]

The set of *all* possible reactions in individual i is denoted $\mathcal{R}_i$. The mental contents of the mind, including all MRs and all RR events, is denoted $X_i \oplus \mathcal{R}_i$. This includes F_i and $\neg F_i$. Recall that the set of all MRs in individual i, including both the food set and elements derived from that food set, is denoted X_i.

$\mathcal{R}_i$ and C_i are not prescribed in advance; because C_i includes remindings and associations on the basis of one or more shared property, different CCPs can occur through interactions among MRs. Nevertheless, it makes perfect mathematical sense to talk about $\mathcal{R}_i$ and C_i as sets. Table 3 summarizes the terminology and correspondences between the OOL and the OOC.

Our model includes elements of cognition that have no obvious parallel in the OOL. We denote the subject of attention at time t as $\mathring{w}_t$. It may be an external stimulus, or an MR

Table 3. Terminology and correspondences between the origin of life (OOL) and the origin of culture (OOC).

term	OOL	OOC
X_i	all molecule types in protocell i	all mental representations (MRs) in individual i
$x \in X_i$	a molecule in X_i	an MR in X_i
F_i	food set for protocell i	innate or directly experienced MRs by i
$r \in \mathcal{R}_i$	a particular reaction in i	a particular representational redescription (RR) in i
$C_i[x]$	reactions catalysed by x in i	RR events 'catalysed' by x in i
$(x, r) \in C$	x catalyses r	x 'catalyses' redescription by r
$a \in A$	member of set of reactants in r	member of set of MRs undergoing r
$b \in B$	member of set of products of r	member of set of MRs resulting from r
$\neg F_i$	non-food set for i (i.e. all B of $\mathcal{R}_i$)	MRs resulting from $\mathcal{R}_i$ (i.e. all B of $\mathcal{R}_i$)

retrieved from memory. Any other contents of $X_i \oplus \mathcal{R}_i$ that are accessible to working memory, such as close associates of $\mathring{w}_t$, or recently attended MRs, are denoted W_t, with W_t constituting a very small subset of $X_i \oplus \mathcal{R}_i$. The focus here is on how non-food set derived MRs (i.e. a non-empty $\neg F$) emerge and connect, giving rise to a semantic network that is reflexively autocatalytic and food set generated.

5. RAF model of the cognitive transition

We now use the RAF formalism to model the transition to behavioural/cognitive modernity in the Upper Palaeolithic. To address how the mind as a whole acquired autocatalytic structure, the model is, by necessity, abstract. It does not distinguish between semantic memory (memory of words, concepts, propositions and world knowledge) and episodic memory (personal experiences); indeed, we are sympathetic to the view that these are not as distinct as once thought [71]. Nor does it address how MRs are obtained (i.e. whether through Hebbian learning versus probabilistic inference). Although MRs are represented simply as points in an N-dimensional space (where N is the number of distinguishable differences, i.e. ways in which MRs could differ), our model is consistent with models that use convolution [72], random indexing [73] or other methods of representing MRs.

We assume that associations form between MRs but do not address whether they are due to similarity or co-occurrence, or whether they are learned through Bayesian inference [74] or other means. We view associations as probabilistic; when we say that an association was forged between two MRs we mean a spike in the probability of one MR evoking another, which we refer to as the 'catalysis' of one MR by the other. We view context as anything external (e.g. an object or person) or internal (e.g. other MRs) that influences the instantiation of an MR in working memory. Although our approach is influenced by how context is modelled in quantum approaches to concepts [75,76], it is not committed to any formal approach to modelling context.

MRs are composed of one or more *concepts:* mental constructs such as CAT or FREEDOM that enable us to interpret new situations in terms of similar previous ones. The rationale for treating MRs as catalysts comes from the literature on concept combination, which provides extensive evidence that, when concepts act as contexts for each other, their meanings change in ways that are often non-trivial and defy classical logic [75–78]. The extent to which the meaning of one MR is modified by another is referred to here as its *reactivity*. A given MR's reactivity varies depending on the other MRs present in working memory.[11] Although we do not explicitly model the dimensionality of semantic space itself (i.e. the features or properties of MRs), we do so indirectly, by representing hierarchical structure in terms of reactivity, as explained below. Our model hinges on the fact that interactions between two or more MRs in working memory alter (however slightly) the network of association strengths [80,81]. Conceptual closure is achieved and a cognitive RAF network emerges when, for each MR, there is an associative pathway to its formation; in other words, any given concept can be explained using other concepts, and new ideas can be re-framed in terms of existing ones.

We now show how the RAF framework is used to model the emergence of a persistent and integrated cognitive RAF, through onset of the capacity to spontaneously control the 'spectrum of thought' variables introduced in §3, and summarized in table 4. In this table, the variables γ_D, γ_A and γ_C quantify the three variables: divergence (D), abstractness (A) and context specificity (C), respectively.

We model the capacity to shift between convergent and divergent thought by introducing a metric geometry. We let d denote the *semantic distance* between an item m in memory M_t and an item in working memory $\mathring{w}$. In convergent thought, the semantic distance d between successive contents of working memory remains small, and only close associates catalyse RR reactions and participate in CCPs. By contrast, by spontaneously engaging in divergent thought when stymied, or as a form of mental exploration or mind-wandering, the modern mind gained access to remote associates (i.e. items for which the semantic distance d to the content of working memory was large). These remote associates catalysed RR reactions, and participated in CCPs. Thus, divergent thought could (in our terminology) bring about reactions among previously unconnected MRs, including MRs from different knowledge domains. The variable γ_D determines how 'remote' an associate can be in order to catalyse an update (i.e. how 'far afield' one looks for ingredients for one's stream of thought). Thus, γ_D provides a threshold on d that increases as one shifts from convergent to divergent thought.

The more abstract a concept, the more associations it can have with other MRs. Therefore, we represent hierarchical semantic structure from concrete instances to increasingly

Table 4. Examples of the three variables of the spectrum of thought.

variable	example	symbol
divergence	LION → KILL → POWER	γ_D
abstractness	LION → ANIMAL → ANIMATE BEING	γ_A
context specificity	LION (context: desire to possess lion's power) → LION FIGURINE	γ_C

abstract concepts in terms of reactivity. Consequently, the more abstract (as opposed to concrete) x in individual i is, the larger the value of $C_i[x]$. Thus, abstract concepts facilitate the navigation of semantic space through CCPs. For example, during the transition from thinking about a particular sharp axe to thinking about the abstract quality of sharpness, abstractness increases, and therefore so does the reactivity, potentially leading the CCP to something quite different from a sharp axe, such as a 'claw'.

As mentioned earlier, the capacity for divergent thought could be made even more useful by broadening the sphere of associates in a context-specific manner, such that one's current needs bias the retrieval of information from memory. This facilitates the forging of new connections between MRs that would be irrelevant in most contexts but are relevant in the current one. We make the notion of context specificity more precise by introducing a context-dependent association structure. As above, we let d denote the *semantic distance* between an item m in memory M_t and an item in working memory $\mathring{w}$. A small value of $d(m, \mathring{w})$ means that, in the current context, m is closely related to $\mathring{w}$ whereas a large value of $d(m, \mathring{w})$ means that, with respect to the current context, they are distantly related.

Let $\mathcal{C}_t$ denote a context at time t (which is determined by the goals and needs of the individual at time t). We can represent the context-dependent associations explicitly by writing $m \sim_{\mathcal{C}_t} m'$ if m and m' are related with respect to context $\mathcal{C}_t$.

For m to catalyse a cognitive updating reaction

$$\mathring{w} \xrightarrow{m} \mathring{w}',$$

m should satisfy at least one of the following two properties, where γ_D is as described above, and γ_C is the extent to which context can facilitate the catalysis of a particular RR reaction:

(i) $d(m, \mathring{w}) \le \gamma_D$, or
(ii) $m \sim_{\mathcal{C}_t} \mathring{w}$ and $d(m, \mathring{w}) \le \gamma_D(1 + \gamma_C)$.

In other words, for m to catalyse an RR reaction involving $\mathring{w}$, either the default semantic distance to m must be sufficiently small that divergent thought makes it accessible or it is pulled within reach because context specificity warps semantic space in such a way as to make this particular association salient. In addition, a particular context $\mathcal{C}_t$ at time t may mean that a stimulus s_t that is relevant to the current contents of working memory catalyses an RR reaction $\mathring{w} \xrightarrow{s} \mathring{w}'$ that would not occur otherwise. For example, seeing an animal puncture a food source with its claw could be a source of ideas for how to make a sharper tool.

5.1. Example: the Hohlenstein–Stadel figurine

We now make the transition from pre-modern to modern mind more concrete using the example of the Löwenmensch or 'lion man' figurine from the Hohlenstein–Stadel cave, discussed in §2. Although we cannot know exactly how the Hohlenstein–Stadel figurine was created, by reverse-engineering the process it is possible to infer what conceptual structure would, at a minimum, have had to be in place [82–84]. We carry this out using available evidence, such as our knowledge that the lion was the largest and most dangerous predator in the ecosystem of the Interpleniglacial [85,86], and likely a source of fear and awe owing to its power and aggression [87]. Since the word 'representation' is often used to refer to an internal, mental construct of something in the world, to avoid confusion, we use the term *iconic* to refer to an object that represents something else in a way that is not merely symbolic but captures its physical attributes.

We now consider the sequence of steps culminating in the creation of the lion man, summarized in table 5 and depicted in figures 2 and 3. Note that the steps culminating in the Hohlenstein–Stadel figurine were preceded by, and dependent upon, the development of lithic reduction (i.e. knapping and carving) techniques. (These are not discussed here, since they are the subject of another paper [39].) Note that the role of catalysts tends to be played by MRs that represent needs or questions, since like catalysts these speed up conceptual change that would otherwise occur very slowly, yet their participation in this process does not fundamentally change them (that is, the degree to which the need is satisfied may change, but the MR of the need does not).

(i) **Form abstract concept, CARVE** (from any suitable material). This consisted of abstracting the general concept of lithic reduction with stone as the source material to lithic reduction using any suitable material (e.g. mammoth ivory). There is evidence that the capacity to abstract a general concept from particular instances dates back to at least 1.76 Ma (well before the Upper Palaeolithic) [88].

(ii) **Social learning of first step.** Creative contributions to culture begin with a preparation stage involving thorough assimilation of relevant background knowledge [89]. Thus, the first step that took place in the Upper Palaeolithic involved the social learning of existing knapping and carving techniques, including the abstract concept CARVE (from any suitable material).

(iii) **Abstraction of CARVE TOOL to CARVE (something)**. The next step was to extricate the concept CARVE TOOL from its conventional function of generating something utilitarian such as a hand axe. This resulted in the abstract concept CARVE, which could now be applied in domains other than technology, such as art. This would have likely involved divergent, abstract thought. The existence of objects in bone, ochre and ostrich eggshell with geometric engravings from southern Africa dates this to at least 77 000 years ago [90], with earlier dates for ivory engraving in China [91].

(iv) **Social learning of the abstract concept CARVE (something)**. The carver of the Hohlenstein–Stadel figurine acquired the abstract concept CARVE (something) through social learning.[12]

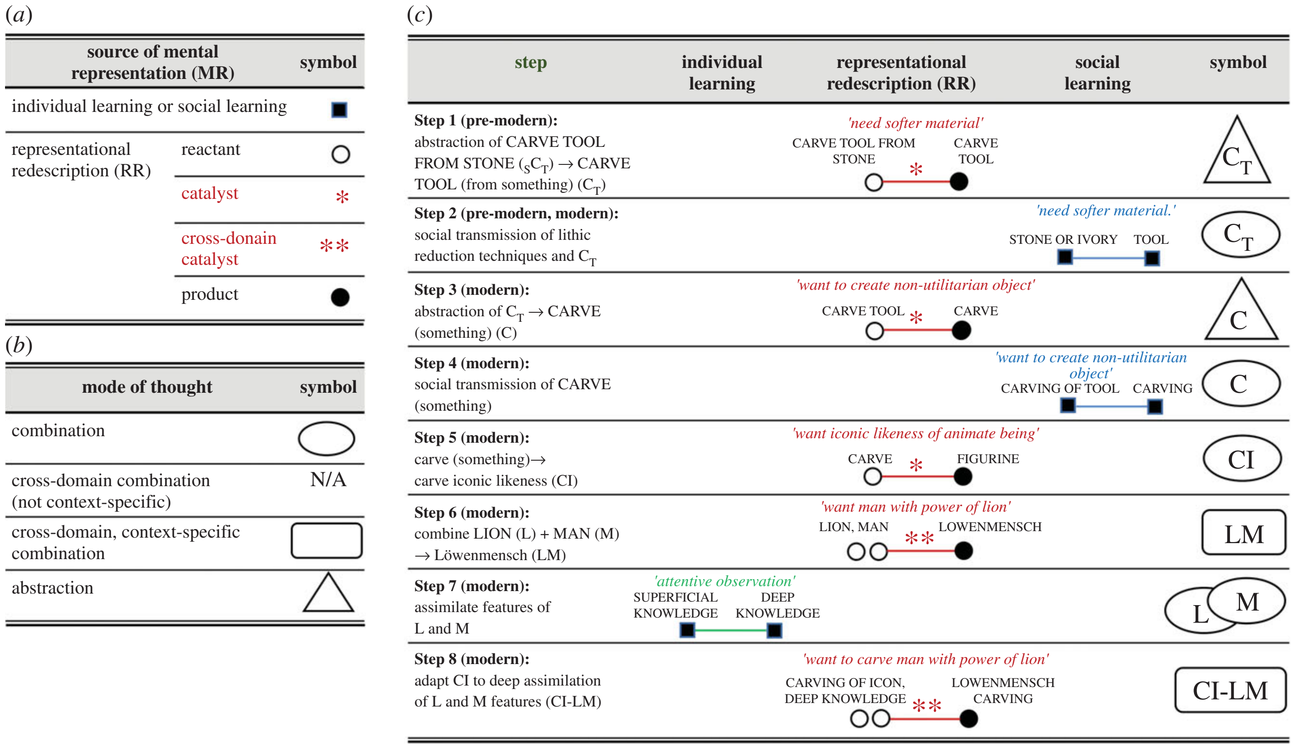

Figure 2. (*a*) Sources of mental representations involved in the creation of the Hohlenstein–Stadel figurine, and the symbols used to depict them. (*b*) Modes of thought and symbols used to depict them. (*c*) Steps involved in the creation of the figurine. Top two rows show the steps that occurred prior to the Upper Palaeolithic; subsequent rows depict steps that took place during the Upper Palaeolithic.

Table 5. The sequence of steps culminating in the creation of the Hohlenstein–Stadel Löwenmensch figurine.

step	description	origin	mode
1	carve from stone → carve (from something)	RR	abstract
2	transmission of lithic reduction techniques	social learning	convergent
3	carve functional tool → carve (something)	RR	divergent, abstract
4	transmission of CARVE (something)	social learning	convergent
5	carve (something) → carve iconic likeness	RR	divergent, concrete
6	combine human form and lion head → internalize lion's power	RR	divergent, cross-domain, context-specific
7	assimilate features of lion and man	individual learning	divergent
8	carve Löwenmensch	individual learning + RR	divergent

(v) **Apply CARVE (something) to the domain of figures (animal and human), yielding CARVE FIGURINE.** We may never know exactly what motivated the first artisan who took the step of carving an iconic likeness, a figurine. It may have been the product of idle mind wandering. An alternative and perhaps more likely explanation is that it was shaped by a goal or desire, such as to (1) know the depicted subject more deeply, or (2) gain a sense of control or mastery over it, or (3) preserve a memory of it, or (4) have constant access to a feeling associated with it, such as the feeling of power associated with a lion. Whatever the motive, it involves taking the concept CARVE and applying it to a new domain, that of ANIMATE BEINGS.

(vi) **Combine LION HEAD with HUMAN BODY.** We also do not know what motivated this step. Like the previous step, it is possible that it was the product of idle mind wandering. It could be that, by endowing a human body like ours with the head of a lion, the artisan hoped that those who held it would internalize the lion's power as their own. An alternative possibility is that it held some religious significance. Again, for the purpose of this model it is not essential to know which of these is correct, for, whatever the underlying motive, this cross-domain combination would have required RR using divergent, context-specific thought.

(vii) **Assimilate features.** To carve an iconic figurine, knowledge of lithic reduction techniques is not sufficient; the artisan would have had to deeply absorb the physical characteristics of lions and humans through individual learning. We characterize this process as divergent because it involves assimilating the details and,

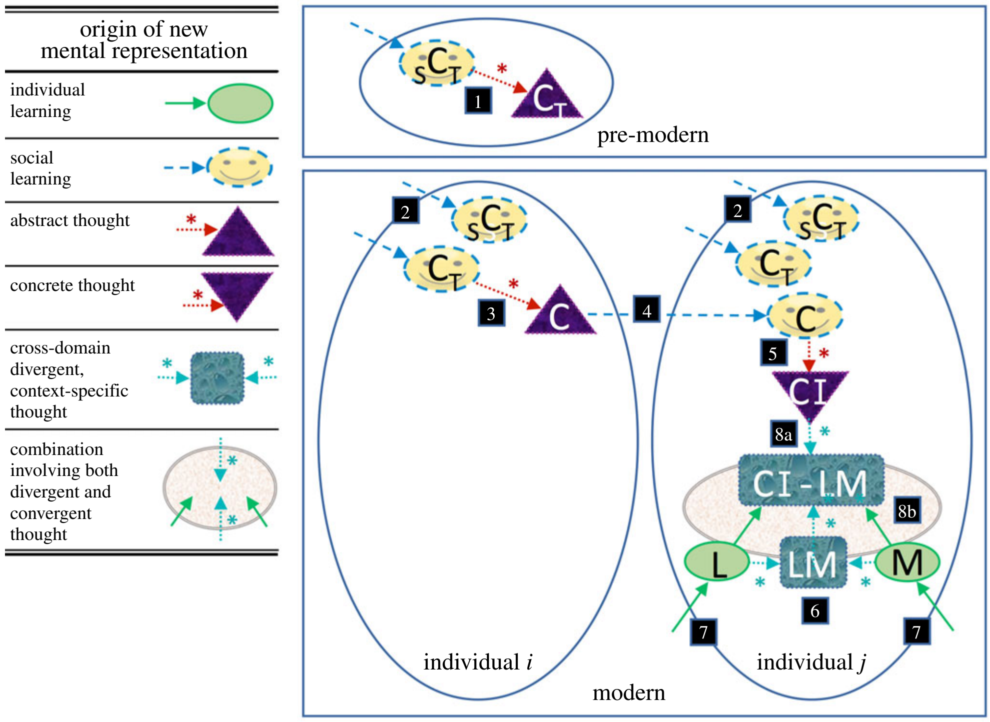

Figure 3. Steps culminating in the creation of the Hohlenstein–Stadel figurine. Meanings of symbols are defined in figure 2 (see text for details).

potentially, any feelings they evoke. The artisan would then have used RR to creatively adapt this technical knowledge to the new task of rendering a figure in ivory.

(viii) **Carve figurine.** The actual carving of the figurine would have required RR in divergent mode to creatively adapt known carving techniques to the new task of rendering the detailed characteristics of the lion and human forms. Engagement in a tactile process meant that thought was concrete, ensuring that the features of the figurine were recognizably human or lion-like.

The entire mental trajectory through the spectrum of thought culminating in the creation of the Hohlenstein–Stadel figurine is depicted in figure 4.

6. The transition to cognitive modernity

In the pre-modern mind, information is thought to have been compartmentalized into domain-specific modules, and RR only operated within particular domains of human inquiry (e.g. 'tools') [7]. Pre-modern cognition was largely (though not entirely) restricted to basic-level categories—an intermediate level of abstraction—without the context-specific RR needed for cross-domain thinking. This was modelled by restricting RAFs to closed subsets of MRs that only reacted with other members of the same subset, resulting in a RAF structure that was transient and fragmented [36,39].

We now describe a simple mathematical model of the formation and persistence of cognitive RAFs culminating in behavioural/cognitive modernity and the cultural transition of the Upper Palaeolithic. Let $\mathcal{W} = \mathcal{W}(t)$ be a continuous measure of the scope (or content) of working memory of an individual at time t (where the continuous variable t varies over the lifetime of that individual). Cognitive processes make use of items obtained through individual learning, social learning or RR, with items persisting in working memory for a short (but variable) time. As in [36], we model this as a non-deterministic process. Let $W = W(t) = \mathbb{E}[\mathcal{W}(t)]$ denote the expected (i.e. mean) value of $\mathcal{W}(t)$. A direct implementation of the dynamical model in [36] then leads to the following nonlinear first-order equation:

$$\frac{\mathrm{d}W}{\mathrm{d}t} = -\mu W + \lambda \mathbb{E}[f(\mathcal{W})] + S, \tag{6.1}$$

where $S = S(t) \geq 0$ is a measure of information that is externally derived, either through individual learning or through social learning at time t, the value $1/\mu$ is the mean time that items remain in working memory, λ describes the rate of RR reactions and f is a certain (unspecified) function that satisfies only the minimal requirements that $f(0) = 0$, $f'(0) > 0$ and f is concave (this last assumption recognizes that $\mathcal{W}(t)$ is bounded). Simple default choices for f would be $f(x) = \min\{x, K\}$ or $f(x) = x(1 - x/K)$, though we do not explicitly assume either of these here.

Crucially, the parameter λ also depends on total memory (a richer memory of knowledge and experiences allows more opportunity to catalyse RR reactions) and it is influenced by the three variables described above (γ_D, γ_A, γ_C) that we propose distinguish pre-modern from modern cognition. More precisely, if $\mathcal{M} = \mathcal{M}(t)$ denotes a continuous measure of the scope of total memory at time t, and $M = M(t) = \mathbb{E}[\mathcal{M}(t)]$ is the expected (mean) value of $\mathcal{M}(t)$, then λ is dependent on M (i.e. $\lambda = \lambda(M)$). Thus, equation (6.1) is coupled to the growth in M, and a simple model for the dynamics of expected total memory M is the first-order linear differential equation

$$\frac{\mathrm{d}(M - W)}{\mathrm{d}t} = \nu W, \tag{6.2}$$

where $\nu \in (0, 1)$ parameterizes the extent to which items in working memory become encoded in long-term memory (which may also depend on time, as ν may vary during the

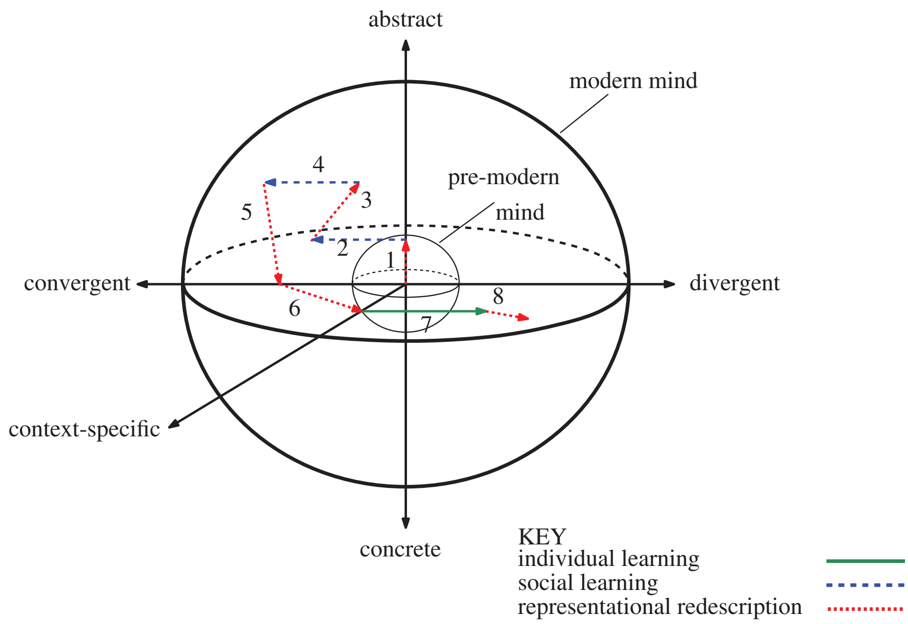

Figure 4. Trajectory through the spectrum of thought culminating in the Hohlenstein–Stadel figurine. Convergent-to-divergent is on the *x*-axis, abstract-to-concrete is on the *y*-axis and the degree of context specificity is on the *z*-axis. Different combinations of these three variables comprise different 'modes' of thought (i.e. ways of navigating memory and processing information). Although the pre-modern mind could, to some degree, form abstractions, its thought trajectories used only a tiny portion of this space, as indicated by the small sphere. It was, therefore, restricted to a single mode of thought. The modern mind could engage in all combinations of these three variables, thereby engaging in many modes of thought, as indicated by the large sphere. The numbered arrows correspond to the eight steps listed in table 5; thus, they depict how the mode of thought shifted over the course of the figurine-making process.

lifetime of the individual). The coupled (nonlinear) system of equations (6.1) and (6.2) leads to certain predictions. In particular, when λ lies below a critical threshold (dependent on the other parameters), CCPs do not form or persist, and thoughts are driven externally (individual learning or social learning). However, once λ passes this threshold, CCPs can form and persist indefinitely, even when the term S (in equation (6.1)) drops to zero. The justification of these two claims and further mathematical details are provided in appendix A.

The ability to shift between convergent and divergent thought, to consider the same item at multiple levels of abstraction and to allow context to bias retrieval from memory by adjusting γ_D, γ_A and γ_C provide distinct and complementary mechanisms for λ to change. If the resulting new MRs are encoded in long-term memory, the positive dependence of λ on M provides more routes for catalysis of CCPs. This increases $M - W$ from equation (6.2), which, in turn, influences the dynamics of W by equation (6.1).

The modern mind could carry out logical operations during convergent thought, and make new connections using divergent thought. Divergent thought could be biased towards a specific need by making thought more contextual. The modern mind could also shift up and down the hierarchy from concrete to abstract. By tuning the mode of thought along the three variables of the above multimodal spectrum to match the situation one is in, the modern mind acquired the capacity to work out how elements of the world were interrelated, and where each element fitted with respect to the whole (i.e. the integrated internal model of the world, or *worldview*). The modern mind could now synthesize different domains of understanding into a coherent web of understandings, using not only basic-level concepts [61] but also higher or lower levels of abstraction, from fine-grained details to the 'big picture', as appropriate.

The worldview of the modern mind is a 'metabolism' in the sense that it has in place entropy-defying processes that maintain its organization. Like the protocell that constituted the earliest structure that could be said to be alive, the structure of the autocatalytic cognitive network as a whole is now maintained through the interactions among its parts. New experiences are interpreted, understood and encoded in memory, in terms of existing cognitive structure already in place.

7. Comparison with other theories

This model builds on the theory that the burst of creativity in the Palaeolithic was due to the onset of contextual focus: the capacity to shift between divergent and convergent modes of thought [57]. That theory is superficially similar to the proposal that the distinguishing feature of human cognition is our capacity for dual processing [8,9].[13] Our model builds on both the contextual focus and dual processing theories by positing that a single-variable spectrum of thought is insufficient to achieve an integrated internal model of the world.

Our model is consistent with Mithen's [93] theory that the transition was due to the connecting of domain-specific information-processing modules, thereby enabling metaphorical thinking and cognitive fluidity: the capacity to combine ideas from different domains, fuse different knowledge processing techniques or adapt a solution to one problem to a different problem. It is also consistent with Coolidge & Wynn's [10] theory that it was due to expanded working memory.[14] Conceptual fluidity and expanded working memory are underwritten by divergent thought but, as explained above, the capacity to engage in divergent thought without the capacity to control *how* divergent one's thinking is would be perilous. Although it is not the focus of this paper, like [10], as well as [65] we are sympathetic with the view that genetic mutation was involved (see [41]).

Our proposal is consistent with the view that complex languages, symbolic representation and myth lay at the heart of this transition [94–96]. However, we put the emergence of a persistent (i.e. stable) and integrated RAF network as central, with language both facilitating and being facilitated by this structure. Given evidence of recursive reasoning well before behavioural modernity, our framework is inconsistent with the hypothesis that the onset of recursive thought enabled mental time travel and cognitive modernity [97,98]; nevertheless, the ability to shift through a multimodal spectrum of thought would have brought on the capacity to make vastly better use of it. The proposal that behavioural modernity can be attributed to the onset of the capacity to model the contents of other minds, sometimes referred to as the 'theory of mind' [5], is somewhat underwritten by recursive RR, since the mechanism that allows for recursion is required for modelling the contents of other minds (though in this case the emphasis is on the social impact of recursion, rather than the capacity for recursion itself). Our proposal is also consistent with explanations for behavioural modernity that emphasize social–ecological factors [6,99], but places these explanations in a broader framework by suggesting a mechanism that aided not just social skills but other skills (e.g. technological) as well.

8. Discussion and conclusion

Formal models exist of many aspects of human cognition, such as learning, memory, planning and concept combination. However, there is little in the way of formal models of how they came to function together as an integrated whole, and how the unique cognitive abilities of *Homo sapiens* came about. RAF networks provide a means of addressing these questions. Building on earlier models of the cognitive transition underlying the earliest origins of human culture and the invention of the Acheulean hand axe, resulting in a transient autocatalytic structure, in this paper, we developed a model of the transition to a persistent, integrated RAF network. We proposed that rapid cultural change in the Middle-Upper Palaeolithic required the ability to not just recursively redescribe the contents of thought but also tailor the 'reactivity' of thought to the current situation. This was accomplished through continuous, spontaneous tuning of three variables that concern not the content of thought *per se*, but how it is processed. The first involves shifting between convergent and divergent processing. The second involves shifting between concrete and abstract representations. The third involves biasing divergent processing according to a pressing need or context. Together, these enabled *Homo sapiens* to reflect on the contents of thought from different perspectives and at different levels of abstraction. This culminated in the crossing of a threshold to conceptual closure and the achievement of self-organizing autocatalytic semantic networks that spanned different knowledge domains, and routinely integrated new information by reframing it in terms of current understandings.

The model is highly simplified, and we do not know that the precise details of the cognitive events modelled here took place (though the model does not hinge on these details). We hope that future research will incorporate inhibition (in conjunction with the existing catalysis), as well as a more sophisticated representation of the interactions among MRs [75,76] and a dynamic representation of context [79,100]. A platform for the computational modelling of RAFs exists (https://github.com/husonlab/catlynet), and we hope to apply it to cognitive RAFs. There remains much work to be done on how cognitive RAFs replicate and evolve (see [33] for informal suggestions in this regard) and on the developmental question of how persistent, integrated RAF networks emerge in the mind of a child.

We also hope that future research will build on the direction taken here by comparing the cognitive RAF model with other standard semantic network models [101–105]. Although these standard semantic networks suffice for modelling semantic structure in individuals, we believe that the RAF approach will turn out to be superior because it distinguishes semantic structure arising through social or individual learning (modelled as food set items) from semantic structure *derived from* this pre-existing material (modelled as non-food set items generated through abstract thought processes that play the role of catalysed reactions). This makes it feasible to model how cognitive structure emerges, and to trace lineages of cumulative cultural change step by step. It also frames this project within the overarching scientific enterprise of understanding how evolutionary processes (be they biological or cultural) begin, and unfold over time.

Data accessibility. This article has no additional data.

Authors' contributions. L.G. provided the conceptual framework and wrote the majority of the text; M.S. contributed mathematical modelling details, including the formal proofs in appendix A.

Competing interests. We declare we have no competing interests.

Funding. L.G. thanks the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding (grant no. 62R06523).

Acknowledgements. We thank the two anonymous reviewers for several helpful suggestions and additional references in an earlier version of this manuscript. We also thank Dietrich Radel for assistance with the formatting of references.

Endnotes

[1]The term 'cultural evolution' is occasionally used in a less restricted sense to refer to the generation and transmission of novelty without the requirement for cumulative, adaptive, open-ended change (e.g. [19]).

[2]Although some attribute cultural evolution uniquely to an increase in the *number* of ideas and cognitive skills (or 'cognitive gadgets') [20], not their interactivity.

[3]For related approaches, see [33,37,38].

[4]Although we use the term 'mental representation', our model is consistent with the view (common among ecological psychologists and in the situated cognition and quantum cognition communities) that what we call mental representations do not 'represent', but instead act as contextually elicited bridges between mind and world.

[5]A more detailed discussion can be found elsewhere [23,41].

[6]Note, however, that we cannot know the extent to which lack of evidence of cultural embellishment in the early record is due to taphonomic biases—i.e. biases in what gets preserved over time (such as lack of preservation of softer materials) [42].

[7]Note that abstract processing is not the same as convergent processing. An item at a particular level of abstraction, such as LION, would, in convergent thought, be held in working memory in a compact manner stripped of details, whereas, during divergent thought, it would be rich in the characteristics of, and feelings evoked by, lions. One might speculate that richly detailed visions of religious deities occur in a mode of thought that is abstract yet divergent.

[8]Note that, in this view, language enhanced not just the ability to communicate and collaborate (thereby accelerating the pace of

cultural innovation) but also the ability to think ideas through for oneself and manipulate them in a manner that was controlled, deliberate and multimodal.

[9]The question of whether the underlying reality being modelled is precisely described by a RAF is more subtle, and beyond the scope of this paper.

[10]This distinction between food set and food set derived may not be so black and white as portrayed here but for simplicity we avoid that subtlety for now.

[11]For example, in a study of the influence of context and mode of thought on the perceived meanings of concepts (as measured by property applicabilities and exemplar typicalities), the concept PYLON was rated low as an exemplar of HAT; however, in the context FUNNY (as in 'worn to be funny'), it was rated high as an exemplar of HAT [79]. Thus, the degree to which PYLON qualified as an instance of a HAT changed depending on the context. The context FUNNY had an even greater effect on the rating of MEDICINE HAT (as in the name of the Canadian town) as an instance of HAT. We say that the *reactivity* was high here because the context exerted a dramatic influence on the perceived meaning of the concept HAT.

[12]We cannot know for certain that it was not invented independently (particularly given the distance between Hohlenstein–Stadel and southern Africa).

[13]Dual processing posits that humans engage in not just a primitive implicit type 1 mode for free association and fast 'gut responses', but also an explicit type 2 mode for deliberate analysis. However, although dual processing makes the split between older, more automatic processes and newer, more deliberate processes, contextual focus theory posits that pre-modern thought was intermediate between two extremes (each valuable in different ways): a divergent mode based on relationships of correlation, and a convergent mode based on relationships of causation. Earlier hominids' memories were coarser-grained, so there were fewer routes for meaningful associations, and less processing of previous experiences. Rather than convergent or divergent processing of previously assimilated material, there was a greater tendency to focus on the here and now, so items in memory tended to remain in the same form as when they were originally assimilated. For a comparison of the divergent thought and dual processing theories, see [92].

[14]Working memory is just the part of memory that is, at any moment, working.

Appendix A. Mathematical details and justification of predictions based on equations (6.1) and (6.2)

In the following arguments, we treat λ as a constant over the short time frame considered in the dynamics of CCPs, since the dependence of λ on M applies over considerably longer time scales. Moreover, in treating λ as a constant and setting $S = 0$, equation (6.1) is technically not an ordinary differential equation (i.e. it is not of the form $\Phi(W, dW/dt) = S(t)$), since $\mathbb{E}[f(\mathcal{W})]$ is not, in general, a function of W. For example, for $f(x) = x(1 - x/K)$, equation (6.1) becomes

$$\frac{\mathrm{d}W}{\mathrm{d}t} = -\mu W + \lambda\left(W\left(1 - \frac{W}{K}\right) - \frac{\mathbb{V}(\mathcal{W})}{K}\right) + S,$$

where $\mathbb{V}(\mathcal{W})$ is the variance of $\mathcal{W}$ at time t. Note also that the dynamics of $\mathcal{W}(t)$ is not determined by the behaviour of $W(t)$; the latter just represents the expected (average) value of the former.

Turning to the first prediction of this model, observe that

$$\mathbb{E}[f(\mathcal{W})] \leq f(\mathbb{E}[\mathcal{W}]) = f(W) \leq f'(0) \cdot W. \tag{A 1}$$

The first inequality in (A 1) is by Jensen's inequality for the concave function f (e.g. [106]). The second inequality also uses the concavity of f together with the conditions $f(0) = 0$ and $f'(0) > 0$. Thus if $\lambda < \mu/f'(0)$, we have

$$\frac{\mathrm{d}W}{\mathrm{d}t} \leq -cW + S$$

for $c = (\mu - \lambda f'(0)) > 0$. Consequently, once S declines to zero, so too does W, and, by the Markov inequality (e.g. [106]), we have (for any $\epsilon > 0$)

$$\mathbb{P}(\mathcal{W}(t) > \epsilon) \leq \frac{\mathbb{E}[\mathcal{W}(t)]}{\epsilon} = \frac{W(t)}{\epsilon} \rightarrow 0,$$

as t increases, which establishes the first prediction.

Now suppose that $\lambda > \mu/f'(0)$. Select $\eta > 0$ sufficiently small so that $\beta := (\mu + \eta)/\lambda < f'(0)$. By the concavity of f, it follows that, for some $\gamma \geq 0$, we have

$$f(x) \geq \beta x, \text{ for all } x \in [0, \gamma], \tag{A 2}$$

since the line $y = \beta x$ and the function $y = f(x)$ both pass through the origin; however, the latter function has a strictly greater slope at the origin.

Now suppose that $\mathcal{W}(t_1) = w$, where $w \in (0, \gamma)$. Considering the process moving forward from time t_1, with the initial condition $\mathcal{W}(t_1) = w$ we then have at $t = t_1$

$$\frac{\mathrm{d}W}{\mathrm{d}t} = -\mu w + \lambda f(w) + S \geq -\mu w + \lambda\beta w = \eta w > 0,$$

where the first inequality is from (A 2) together with $S(t_1) \geq 0$. In summary, when λ passes above the threshold $\mu/f'(0)$ and $\mathcal{W}(t)$ is small but non-zero the expected value of $\mathcal{W}(t)$ begins to increase, owing to CCPs in working memory (and even with $S = 0$).

References

1. Kissel M, Fuentes A. 2018 'Behavioral modernity' as a process, not an event, in the human niche. *Time Mind* **11**, 163–183. (doi:10.1080/1751696X.2018.1469230)
2. McBrearty S, Brooks A. 2000 The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J. Hum. Evol.* **39**, 453–563. (doi:10.1006/jhev.2000.0435)
3. Sterelny K. 2011 From hominins to humans: how *sapiens* became behaviourally modern. *Phil. Trans. R Soc. B* **366**, 809–822. (doi:10.1098/rstb.2010.0301)
4. Davies S. 2019 Behavioral modernity in retrospect. *Topoi* 1–12. (doi:10.1007/s11245-019-09671-4)
5. Tomasello M. 2014 *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
6. Whiten A. 2011 The scope of culture in chimpanzees, humans and ancestral apes. *Phil. Trans. R. Soc. B* **366**, 997–1007. (doi:10.1098/rstb.2010.0334)
7. Mithen S. 2006 Ethnobiology and the evolution of the human mind. *J. R. Anthropol. Inst.* **12**, S45–S61. (doi:10.1111/j.1467-9655.2006.00272.x)
8. Evans J. 2008 Dual-process accounts of reasoning, judgment and social cognition. *Annu. Rev. Psychol.* **59**, 255–278. (doi:10.1146/annurev.psych.59.103006.093629)
9. Nosek BA. 2007 Implicit-explicit relations. *Curr. Dir. Psychol. Sci.* **16**, 65–69. (doi:10.1111/j.1467-8721.2007.00477.x)
10. Coolidge FL, Wynn T. 2005 Working memory, its executive functions, and the emergence of modern thinking. *Camb. Archaeol. J.* **15**, 5–26. (doi:10.1017/S0959774305000016)

11. Bentley RA, Hahn MW, Shennan SJ. 2004 Random drift and culture change. *Proc. R. Soc. Lond. B* **271**, 1443–1450. (doi:10.1098/rspb.2004.2746)
12. Boyd R, Richerson P. 1988 *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.
13. Cavalli-Sforza LL, Feldman MW. 1981 *Cultural transmission and evolution: a quantitative approach*. Princeton, NJ: Princeton University Press.
14. Enquist M, Ghirlanda S, Eriksson K. 2011 Modelling the evolution and diversity of cumulative culture. *Phil. Trans. R. Soc. B* **366**, 412–423. (doi:10.1098/rstb.2010.0132)
15. Gabora L. 2013 An evolutionary framework for culture: selectionism versus communal exchange. *Phys. Life Rev.* **10**, 117–145. (doi:10.1016/j.plrev.2013.03.006)
16. Holden CJ, Mace R. 2003 Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proc. R. Soc. Lond. B* **270**, 2425–2433. (doi:10.1098/rspb.2003.2535)
17. Mesoudi A, Whiten A, Laland KN. 2006 Towards a unified science of cultural evolution. *Behav. Brain Sci.* **29**, 329–347. (doi:10.1017/S0140525X06009083)
18. Powell A, Shennan S, Thomas MG. 2009 Late Pleistocene demography and the appearance of modern human behavior. *Science* **324**, 1298–1301. (doi:10.1126/science.1170165)
19. Whiten A. 2019 Cultural evolution in animals. *Annu. Rev. Ecol. Evol. Syst.* **50**, 1–22. (doi:10.1146/annurev-ecolsys-110218-025040)
20. Heyes C. 2018 Enquire within: cultural evolution and cognitive science. *Phil. Trans. R. Soc. B* **373**, 20170051. (doi:10.1098/rstb.2017.0051)
21. Gysi DM, Nowick K. 2020 Construction, comparison and evolution of networks in life sciences and other disciplines. *J. R. Soc. Interface* **17**, 20190610. (doi:doi.org/10.1098/rsif.2019.0610)
22. Gabora L. 1998 Autocatalytic closure in a cognitive system: a tentative scenario for the origin of culture. *Psycoloquy* **9**.
23. Gabora L, Smith C. 2018 Two cognitive transitions underlying the capacity for cultural evolution. *J. Anthropol. Sci.* **96**, 27–52. (doi:10.4436/jass.96008)
24. Erdös P, Rényi A. 1960 On the evolution of random graphs. *Pub. Math. Inst. Hung. Acad. Sci.* **5**, 17–61.
25. Kauffman SA. 1986 Autocatalytic sets of proteins. *J. Theor. Biol.* **119**, 1–24. (doi:10.3390/ijms12053085)
26. Kauffman SA. 1993 *The origins of order*. Oxford, UK: Oxford University Press.
27. Hordijk W, Steel M. 2015 Autocatalytic sets and boundaries. *J. Syst. Chem.* **6**, 1. (doi:10.1186/s13322-014-0006-2)
28. Xavier JC, Hordijk W, Kauffman S, Steel M, Martin WF. 2020 Autocatalytic chemical networks at the origin of metabolism. *Proc. R. Soc. B* **287**, 20192377. (doi:10.1098/rspb.2019.2377)
29. Smith JM, Szathmary E. 1997 *The major transitions in evolution*. Oxford, UK: Oxford University Press.
30. Hordijk W, Steel M. 2016 Chasing the tail: the emergence of autocatalytic networks. *Biosystems* **152**, 1–10. (doi:10.1016/j.biosystems.2016.12.002)
31. Steel M, Hordijk W, Xavier JC. 2019 Autocatalytic networks in biology: structural theory and algorithms. *J. R. Soc. Interface* **16**, 20180808. (doi:10.1098/rsif.2018.0808)
32. Vasas V, Fernando C, Santos M, Kauffman S, Szathmáry E. 2012 Evolution before genes. *Biol. Direct* **7**, 1. (doi:10.1186/1745-6150-7-1)
33. Andersson C, Törnberg P. 2019 Toward a macroevolutionary theory of human evolution: the social protocell. *Biol. Theory* **14**, 86–102. (doi:10.1007/s13752-018-0313-y)
34. Gabora L. 2000 Conceptual closure: how memories are woven into an interconnected worldview. *Ann. N. Y. Acad. Sci.* **901**, 42–53. (doi:10.1111/j.1749-6632.2000.tb06264.x)
35. Gabora L, Aerts D. 2009 A model of the emergence and evolution of integrated worldviews. *J. Math. Psychol.* **53**, 434–451. (doi:10.1016/j.jmp.2009.06.004)
36. Gabora L, Steel M. 2017 Autocatalytic networks in cognition and the origin of culture. *J. Theor. Biol.* **431**, 87–95. (doi:10.1016/j.jtbi.2017.07.022)
37. Cabell KR, Valsiner J. 2013 *The catalyzing mind: beyond models of causality*. Annals of Theoretical Psychology, vol. 11. Berlin, Germany: Springer.
38. Muthukrishna M, Doebeli M, Chudek M, Henrich J. 2018 The cultural brain hypothesis: how culture drives brain expansion, sociality, and life history. *PLoS Comput. Biol.* **14**, e1006504. (doi:10.1371/journal.pcbi.1006504)
39. Gabora L, Steel M. 2020 Modeling a cognitive transition at the origin of cultural evolution using autocatalytic networks. *Cogn. Sci.* **44**, e12878. (doi:10.1111/cogs.12878)
40. Mithen S. 1996 *The prehistory of the mind: a search for the origins of art, religion, and science*. London, UK: Thames and Hudson.
41. Gabora L, Smith C. 2019 Exploring the psychological basis for transitions in the archaeological record. In *Handbook of cognitive archaeology: psychology in prehistory* (eds T Henley, E Kardas, M Rossano), ch. 12. Abingdon, UK: Routledge/Taylor and Francis.
42. Andersson C, Törnberg A, Törnberg P. 2014 An evolutionary developmental approach to cultural evolution. *Curr. Anthropol.* **55**, 154–174. (doi:10.1086/675692)
43. Tattersall I. 1998 *The origin of the human capacity*. New York, NY: American Museum of Natural History.
44. Mulvaney J, Kamminga J. 1999 *Prehistory of Australia*. Washington, DC: Smithsonian Institution Scholarly Press.
45. Mellars P. 2006 Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* **313**, 796–800. (doi:10.1126/science.1128402)
46. Ambrose S. 2001 Palaeolithic technology and human evolution. *Science* **291**, 1748–1753. (doi:10.1126/science.1059487)
47. Aubert M *et al.* 2018 Palaeolithic cave art in Borneo. *Nature* **564**, 254–257. (doi:10.1038/s41586-018-0679-9)
48. Nelson S. 2008 Diversity of the Upper Palaeolithic Venus figurines and archaeological mythology. *Archeol. Pap. Am. Anthropol. Assoc.* **2**, 11–22. (doi:10.1525/ap3a.1990.2.1.11)
49. Pike A, Hoffmann D, García-Diez M, Pettitt P, Alcolea J, De Balbin R, Zilhão J. 2012 U-series dating of Palaeolithic art in 11 caves in Spain. *Science* **336**, 1409–1413. (doi:10.1126/science.1219957)
50. d'Errico F, Vanhaeren M, Barton N, Bouzouggar A, Mienis H, Richter D, Hublin JJ, McPherron SP, Lozouet P. 2009 Additional evidence on the use of personal ornaments in the Middle Palaeolithic of North Africa. *Proc. Natl Acad. Sci. USA* **106**, 16 051–16 056. (doi:10.1073/pnas.0903532106)
51. Otte M. 2012 The management of space during the Palaeolithic. *Q. Int.* **247**, 212–229. (doi:10.1016/j.quaint.2010.11.031)
52. Erlandson JM. 2001 The archaeology of aquatic adaptations: paradigms for a new millennium. *J. Archaeol. Res.* **9**, 287–350. (doi:10.1023/A:1013062712695)
53. Hovers E, Lani S, Bar-Yosef O, Vandermeersch B. 2003 An early case of color symbolism: ochre use by modern humans in Qafzeh cave. *Curr. Anthropol.* **44**, 491–522. (doi:10.1086/375869)
54. Rappaport R. 1999 *Ritual and religion in the making of humanity*. Cambridge, UK: Cambridge University Press.
55. Bickerton D. 2007 Language evolution: a brief guide for linguists. *Lingua* **117**, 510–526. (doi:10.1016/j.lingua.2005.02.006)
56. Lieberman P. 2015 Language did not spring forth 100,000 years ago. *PLoS Biol.* **13**, e1002064. (doi:10.1371/journal.pbio.1002064)
57. Gabora L. 2003 Contextual focus: a cognitive explanation for the cultural transition of the Middle/Upper Palaeolithic. In *Proc. of the 25th Annual Meeting of the Cognitive Science Society* (eds R Alterman, D Kirsch), pp. 432–437. Hillsdale, NJ: Lawrence Erlbaum Associates.
58. Gabora L. 2010 Revenge of the 'neurds': characterizing creative thought in terms of the structure and dynamics of human memory. *Creat. Res. J.* **22**, 1–13. (doi:10.1080/10400410903579494)
59. Gabora L. 2018 How insight emerges in a distributed, content-addressable memory. In *The Cambridge handbook of the neuroscience of creativity* (eds O Vartanian, J Jung), pp. 58–70. Cambridge, MA: MIT Press.
60. Gabora L. 2019 Reframing convergent and divergent thought for the 21st century. In *Proc. of the 2019 Annual Meeting of the Cognitive Science Society* (eds A Goel, C Seifert, C Freska), pp. 1794–1800. Austin, TX: Cognitive Science Society.
61. Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P. 1976 Basic objects in natural categories. *Cognit. Psychol.* **8**, 382–439. (doi:10.1016/0010-0285(76)90013-X)
62. Barsalou LW. 1982 Context-independent and context-dependent information in concepts. *Mem. Cognit.* **10**, 82–93. (doi:10.3758/BF03197629)

63. Mednick SA. 1962 The associative basis of the creative process. *Psychol. Rev.* **69**, 220–232. (doi:10.1037/h0048850)
64. Pagel M. 2012 *Wired for culture: the natural history of human cooperation*. London, UK: Penguin.
65. Corballis M. 2004 The origins of modernity: was autonomous speech the critical factor? *Psychol. Rev.* **111**, 543–552. (doi:10.1037/0033-295X.111.2.543)
66. Mossel E, Steel M. 2005 Random biochemical networks and the probability of self-sustaining autocatalysis. *J. Theor. Biol.* **233**, 327–336. (doi:10.1016/j.jtbi.2004.10.011)
67. Hordijk W, Hein J, Steel M. 2010 Autocatalytic sets and the origin of life. *Entropy* **12**, 1733–1742. (doi:10.3390/e12071733)
68. Hordijk W, Kauffman SA, Steel M. 2011 Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems. *Int. J. Mol. Sci.* **12**, 3085–3101. (doi:10.3390/ijms12053085)
69. Hordijk W, Steel M. 2004 Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.* **227**, 451–461. (doi:10.1016/j.jtbi.2003.11.020)
70. Cazzolla Gatti R, Fath B, Hordijk W, Kauffman S, Ulanowicz R. 2018 Niche emergence as an autocatalytic process in the evolution of ecosystems. *J. Theor. Biol.* **454**, 110–117. (doi:10.1016/j.jtbi.2018.05.038)
71. Kwantes PJ. 2005 Using context to build semantics. *Psychon. Bull. Rev.* **12**, 703–710. (doi:10.3758/BF03196761)
72. Jones MN, Mewhort DJK. 2007 Representing word meaning and order information in a composite holographic lexicon. *Psychol. Rev.* **114**, 1–37. (doi:10.1037/0033-295X.114.1.1)
73. Kanerva P. 2009 Hyperdimensional computing: an introduction to computing in distributed representations with high-dimensional random vectors. *Cogn. Comput.* **1**, 139–159. (doi:10.1007/s12559-009-9009-8)
74. Griffiths TL, Steyvers M, Tenenbaum JB. 2007 Topics in semantic representation. *Psychol. Rev.* **114**, 211–244. (doi:10.1037/0033-295X.114.2.211)
75. Aerts D, Broekaert J, Gabora L, Sozzo S. 2016 Generalizing prototype theory: a formal quantum framework. *Front. Psychol. (Cognition)* **7**, 418. (doi:10.3389/fpsyg.2016.00418)
76. Aerts D, Gabora L, Sozzo S. 2013 Concepts and their dynamics: a quantum theoretical model. *Top. Cogn. Sci.* **5**, 737–772. (doi:10.1111/tops.12042)
77. Hampton JA. 1988 Disjunction of natural concepts. *Mem. Cogn.* **16**, 579–591. (doi:10.3758/BF03197059)
78. Osherson DN, Smith EE. 1981 On the adequacy of prototype theory as a theory of concepts. *Cognition* **9**, 35–58. (doi:10.1016/0010-0277(81)90013-5)
79. Veloz T, Gabora L, Eyjolfson M, Aerts D. 2011 Toward a formal model of the shifting relationship between concepts and contexts during associative thought. In *Proc. of the Fifth Int. Symp. on Quantum Interaction* (eds D Song, M Melucci, I Frommholz, P Zhang, L Wang, S Arafat), pp. 25–34. Berlin, Germany: Springer.
80. Brockmeier J. 2010 After the archive: remapping memory. *Cult. Psychol.* **16**, 5–35. (doi:10.1177/1354067X09353212)
81. McClelland JM. 2011 Memory as a constructive process: the parallel distributed processing approach. In *The memory process: neuroscientific and humanistic perspectives* (eds S Nalbantian, PM Matthews, JL McClelland), pp. 129–155. Cambridge, MA: MIT Press.
82. Gabora L, Leijnen S, Veloz T, Lipo C. 2011 A non-phylogenetic conceptual network architecture for organizing classes of material artifacts into cultural lineages. In *Proc. of the 33rd Annual Meeting of the Cognitive Science Society* (eds L Carlson, C Hölscher, TF Shipley), pp. 2923–2928. Cognitive Science Society, Psychology Press.
83. Tehrani JJ, Riede F. 2008 Towards an archaeology of pedagogy: learning, teaching and the generation of material culture traditions. *World Archaeol.* **40**, 316–331. (doi:10.1080/00438240802261267)
84. Veloz T, Tempkin I, Gabora L. 2012 A conceptual network-based approach to inferring cultural phylogenies. In *Proc. of the 34th Annual Meeting of the Cognitive Science Society* (eds N Miyake, D Peebles, RP Cooper), pp. 2487–2492. Austin TX: Cognitive Science Society.
85. Kind C, Ebinger-Rist N, Wolf S, Beutelspacher T, Wehrberger K. 2014 The smile of the lion man. Recent excavations in Stadel cave (Baden-Württemberg, southwestern Germany) and the restoration of the famous upper palaeolithic figurine. *Quartär* **61**, 129–145.
86. Porr M. 2010 Palaeolithic art as cultural memory: a case study of the Aurignacian art of Southwest Germany. *Camb. Archaeol. J.* **20**, 87–108. (doi:10.1017/S0959774310000065)
87. Hahn J. 1986 *Kraft und aggression. Die botschaft der eiszeitkunst im Aurignacien Süddeutschlands?* Tübingen, Germany: Archaeologica Venatoria.
88. Stout D, Toth N, Schick K, Chaminade T. 2008 Neural correlates of early stone age toolmaking: technology, language and cognition in human evolution. *Phil. Trans. R. Soc. B* **363**, 1939–1949. (doi:10.1098/rstb.2008.0001)
89. Zhang X, Wang D, Wang T. 2016 Inspiration or preparation? Explaining creativity in scientific enterprise. In *Proc. of the 25th ACM Int. on Conf. on Information and Knowledge Management*, pp. 741–750. New York, NY: ACM.
90. Henley T, Rossano MJ, Kardas E. 2020 *Handbook of cognitive archaeology: a psychological framework*. Abingdon, UK: Routledge/Taylor and Francis.
91. Gao X, Huang W, Ziqiang X, Zhibang M, Olsen JW. 2004 120-150 ka human tooth and ivory engravings from Xinglongdong Cave, Three Gorges Region, South China. *Chin. Sci. Bull.* **49**, 175–180. (doi:10.1360/03wd0214)
92. Sowden P, Pringle A, Gabora L. 2015 The shifting sands of creative thinking: connections to dual process theory. *Think. Reason.* **21**, 40–60. (doi:10.1080/13546783.2014.885464)
93. Mithen S. 1998 *Creativity in human evolution and prehistory*. London, UK: Routledge.
94. Bickerton D. 2014 *More than nature needs: language, mind and evolution*. Cambridge, UK: Harvard University Press.
95. Bickerton D, Szathmáry E. 2009 *Biological foundations and origin of syntax*. Cambridge, UK: MIT Press.
96. Deacon T. 1997 *The symbolic species: the coevolution of language and the brain*. New York, NY: Norton.
97. Corballis M. 2011 *The recursive mind: the origins of human language, thought and civilization*. Princeton, NJ: Princeton University Press.
98. Suddendorf T, Addis DR, Corballis MC. 2009 Mental time travel and the shaping of the human mind. *Phil. Trans. R. Soc. B* **364**, 1317. (doi:10.1098/rstb.2008.0301)
99. Foley R, Gamble C. 2009 The ecology of social transitions in human evolution. *Phil. Trans. R. Soc. B* **364**, 3267–3279. (doi:10.1098/rstb.2009.0136)
100. Howard MW, Kahana MJ. 2002 A distributed representation of temporal context. *J. Math. Psychol.* **46**, 269–299. (doi:10.1006/jmps.2001.1388)
101. Baronchelli A, Ferrer-i-Cancho R, Pastor-Satorras R, Chater N, Christiansen MH. 2013 Networks in cognitive science. *Trends Cogn. Sci.* **17**, 348–360. (doi:10.1016/j.tics.2013.04.010)
102. Beaty RE, Benedek M, Silvia PJ, Schacter DL. 2016 Creative cognition and brain network dynamics. *Trends Cogn. Sci.* **20**, 87–95. (doi:10.1016/j.tics.2015.10.004)
103. Betzel RF, Bassett DS. 2017 Generative models for network neuroscience: prospects and promise. *J. R. Soc. Interface* **14**, 20170623. (doi:10.1098/rsif.2017.0623)
104. Karuza EA, Thompson-Schill SL, Bassett DS. 2016 Local patterns to global architectures: influences of network topology on human learning. *Trends Cogn. Sci.* **20**, 629–640. (doi:10.1016/j.tics.2016.06.003)
105. Medaglia JD, Lynall ME, Bassett DS. 2015 Cognitive network neuroscience. *J. Cogn. Neurosci.* **27**, 1471–1491. (doi:10.1162/jocn_a_00810)
106. Grimmett G, Stirzaker D. 2001 *Probability and random processes*, 3rd edn. Oxford, UK: Oxford University Press.