An Autocatalytic Network Model of Conceptual Change

Liane Gabora

Department of Psychology, University of British Columbia

Nicole M Beckage

Intel Labs, Hillsboro OR

Mike Steel

Biomathematics Research Centre, University of Canterbury, NZ

Abstract

In Reflexively Autocatalytic Foodset-generated (RAF) networks, nodes are not just passive transmitters of activation; they actively galvanize, or ‘catalyze’ the synthesis of novel (‘foodset-derived’) nodes from existing ones (the ‘foodset’). Thus, RAFs are uniquely suited to modeling how new structure grows out of currently available structure, and analyzing phase transitions in potentially very large networks. RAFs have been used to model the origins of evolutionary processes, both biological (the origin of life) and cultural (the origin of cumulative innovation), and may potentially provide an overarching framework that integrates evolutionary and developmental approaches to cognition. Applied to cognition, the *foodset* consists of information obtained through social learning or individual learning of pre-existing information, and *foodset-derived* items arise through mental operations resulting in new information. Thus, mental representations are not just propagators of spreading activation; they trigger the derivation of new mental representations. To illustrate the application of RAF networks in cognitive science, we develop a step-by-step process model of conceptual change (i.e., the process by which a child becomes an active participant in cultural evolution), focusing on childrens' mental models of the shape of the Earth. Using results from (Vosniadou & Brewer, 1992), we model different trajectories from the Flat Earth model to the Spherical Earth model, as well as the impact of other factors, such as pretend play, on cognitive development. As RAFs increase in size and number, they begin to merge, bridging previously compartmentalized knowledge, and get subsumed by a giant RAF (the maxRAF) that constrains and enables the scaffolding of new conceptual structure. At this point, the cognitive network becomes self-sustaining and self-organizing. The child can reliably frame new knowledge and experiences in terms of previous ones, and engage in recursive representational redescription, and abstract thought. We suggest that individual differences in the *reactivity* of mental representations, i.e., their proclivity to trigger conceptual change, culminates in different cognitive networks and concomitant learning trajectories.

Keywords: autocatalytic network, cognitive transition, cognitive network, conceptual change, creativity, individual differences, mental model, mental representation, process model

An Autocatalytic Network Model of Conceptual Change

Introduction

The aim of this paper is twofold. First, we introduce Reflexively Autocatalytic Foodset-generated (RAF) networks, and the rationale behind their application in cognitive science. Second, we illustrate how they can be used to model cognitive development. What differentiates RAFs from other network science approaches is that nodes are not just passive transmitters of activation; they actively galvanize, or 'catalyze' the synthesis of novel ('foodset-derived') nodes from existing ones (the 'foodset'). The general RAF setting is conducive to the development of efficient (polynomial-time) algorithms for questions that are computationally intractable (NP-hard) (Steel, Hordijk, & Xavier, 2019). These features make RAFs uniquely suited to model how new structure grows out of earlier structure, i.e., generative network growth (Steel et al., 2019). Such generativity may result in phase transitions to a network that is self-sustaining and self-organizing (Hordijk & Steel, 2004, 2016; Mossel & Steel, 2005), as well as potentially able to self-replicate (in a relatively haphazard manner, without reliance on a self-assembly code), and evolve, i.e., exhibit cumulative, adaptive, open-ended change (Gabora & Steel, 2021a; Hordijk & Steel, 2015). Therefore, RAFs have been used to model the origins of evolutionary processes, both biological—the origin of life (OOL) (Hordijk, Hein, & Steel, 2010; Xavier, Hordijk, Kauffman, Steel, & Martin, 2020)—and cultural—the origin of culture (OOC), or more specifically, the kind of cognitive structure capable of generating cumulative, adaptive, open-ended innovation (Gabora & Steel, 2017, 2020a, 2020b, 2021b; Steel, Xavier, & Huson, 2020).

Because RAF nodes modify network structure, the RAF framework is consistent with the goal of understanding not just how networks are structured but also how they dynamically restructure themselves in response to internal and external pressures. In a OOL context, they were used to develop the hypothesis that life began as, not as a single self-replicating molecule, but a set of molecules that, through catalyzed reactions, collectively reconstituted the whole (Kauffman, 1993). Autocatalytic network theory has successfully demonstrated—mathematically or using simulations (Hordijk et al., 2010; Hordijk, Kauffman, & Steel, 2011), and with real biochemical systems (Hordijk & Steel, 2013; Xavier et al., 2020)—how self-maintaining structures that evolve and replicate can emerge from nonliving molecules.

In a cognitive context, autocatalytic models focus not just on network structure per se, but on how the network acquires the capacity to reconfigure itself on the fly in response to current needs. The observation that, like living organisms, cognitive networks are self-sustaining, self-organizing, and self-reproducing (Barton, 1994; Maturana & Varela, 1973; Pribram, 1994; Varela, Thompson, & Rosch, 1991) suggests that cognitive networks constitute a second level of autocatalytic structure.[1] The self-sustaining nature of a cognitive network is evident in the tendency to reduce cognitive dissonance, resolve inconsistencies, and preserve existing schemas in the face of new information. Although the contents of a cognitive network change over time, it maintains integrity as a relatively coherent whole. It's spontaneously self-organizing nature is evident in the capacity to combine remote associates (Mednick, 1962) (such as combining SNOW and MAN to invent SNOWMAN). It reproduces in a piecemeal manner through social learning and imitation, and the propensity to share stories and perspectives. This proposed second level of autocatalytic structure is not merely an extension of organismal needs; indeed, these two levels of endogenous control can be at odds (e.g., a scientist immersed in solving a problem may neglect offspring, or forget to eat.)

Autocatalytic networks have been used to cognitive model phase transitions culminating in behavioral modernity (i.e., the capacity to think and act like modern humans), and the capacity for cumulative, adaptive, open-ended cultural novelty (Gabora, 1998, 1999; Gabora & Steel, 2017, 2020a, 2020b). The OOC involves, not chemical reaction networks, but networks of knowledge and memories, and the products and reactants are not catalytic molecules but mental representations (MRs). 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.

[1] By *cognitive network,* we refer to an individual's web of concepts, language terms, and their associations, as well as knowledge and memories, and how they are structured.

Just as reactions between molecules generate new molecules, interactions between MRs generate new MRs, which, in turn, enable new interactions. Just as a catalyst makes a certain reaction more likely to occur, a question, desire, or external stimulus can trigger or 'catalyze' a new idea, or perspective.

It has been proposed that the origin of a self-modifying autocatalytic cognitive network is the key to cumulative cultural evolution in our species (Gabora & Steel, 2020b), and the source of our generativity (Gabora & Steel, 2021b). Since autocatalytic networks incorporate the 'triggering' or 'catalytic' effect that an observation or piece of information can have on network structure, they provide a means of incorporating creativity into cognitive networks. This requires a network that not only it possesses (for example), MRs of beanbags and chairs (or phones and digital music players) but, when appropriate, actively combines these distant associates to invent the beanbag chair (or the iPhone).

For culture to evolve, it was not sufficient that a self-organizing, generative cognitive network came about once. In keeping with the adage 'ontogeny recapitulates phylogeny,' it must emerge anew in each young child, such that each individual contributes in some way, large or small, to cultural evolution. (Fig. 1). Thus, the OOC is in part an evolutionary story (how did humans evolve the capacity for a second evolutionary process: cultural evolution?), and in part a developmental story (what kind of cognitive developmental change enables a child to become an active participant in this second evolutionary process?). The application of autocatalytic networks to the origin and evolution of distinctively human cognition lays the groundwork for an autocatalytic framework for cognitive development.

Developmental psychology has little to say about a child develops a self-organizing, generative cognitive network. The field has progressed toward understanding the acquisition of specific abilities such as problem solving and theory of mind, but to understand how children come to possess a particular outlook and understanding of the world, we must understand how they organize their cognitive networks (Sizemore, Phillips-Cremins, Ghrist, & Bassett, 2019). Although access to the internal processes of young children is a limiting factor, differences

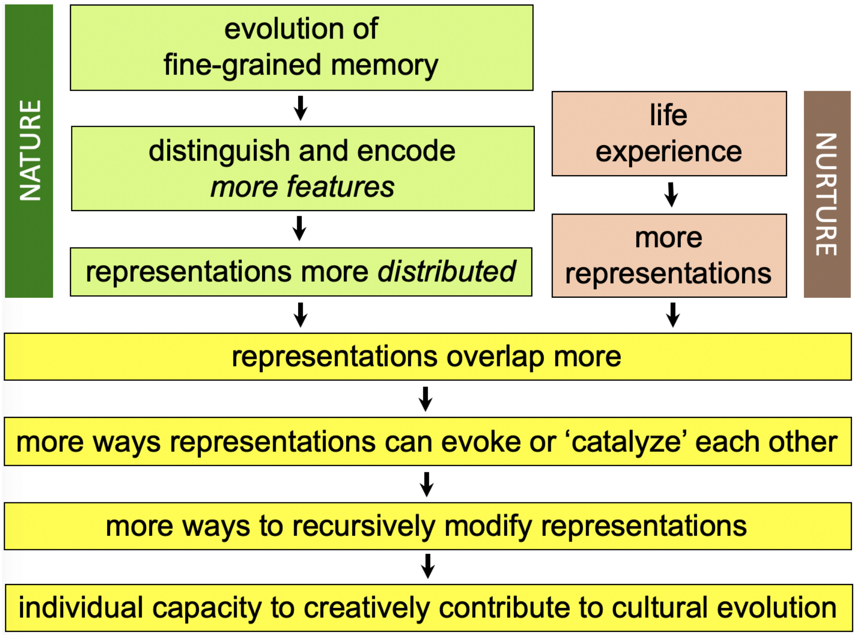
NATURE
evolution of fine-grained memory
distinguish and encode *more features*
representations more *distributed*
life experience
more representations
NURTURE
representations overlap more
more ways representations can evoke or ‘catalyze’ each other
more ways to recursively modify representations
individual capacity to creatively contribute to cultural evolution

Figure 1. Schematic illustration of the proposed changes set into motion through the evolution of finer-grained memory, culminating in an individual’s capacity to become a creative contributor to cultural evolution. For this sequence of events to occur, not only must the individual’s representations be sufficiently distributed, due to the *biologically evolved changes* depicted on the upper left, but they must be sufficiently plentiful, due to the *developmental changes* depicted on the upper right. It is not until one has lived long enough to accumulate a sufficiently rich repository of experience that there is sufficient overlap in ones’ distributed MRs. This provides routes by which MRs can evoke one another, and thus, a means to become a creative, contributing member of society. From Gabora and Steel (2020a). Reprinted with permission.

between learners can be found and quantified, and such differences are important for understanding childrens’ learning and development (Beckage & Colunga, 2019; Beckage, Smith, & Hills, 2011). A child must acquire a coherent, integrated, flexible cognitive network so as to be able to reframe new situations in terms of previous ones, adapt old responses to new tasks, reflect on ideas from different perspectives, and combine knowledge in appropriate and meaningful ways. The child thereby becomes a new ‘cog’ in the cultural evolution machinery. We suggest that RAFs may potentially provide an overarching framework for the emergence of a cognitive network that integrates evolutionary and developmental approaches to cognition.

Situating the Approach in Relation to Network Science

Since knowledge exhibits intermediate modularity, consisting of loosely connected clusters that can be characterized using tools from network science, network science has emerged as a powerful tool for understanding cognition (Baronchelli, Ferrer-i-Cancho, Pastor-Satorras, Chater, & Christiansen, 2013; Borge-Holthoefer & Arenas, 2010; Kumar, Steyvers, & Balota, in press; Siew, Wulff, Beckage, & Kenett, 2019). Though the term 'autocatalytic networks' reflects their initial application to the origin of life (OOL) (Farmer, Kauffman, & Packard, 1986; Kauffman, 1993), autocatalytic networks are a general mathematical setting for studying networks that arose out of earlier work in graph theory (Erdös & Rényi, 1960). Like standard network science approaches in cognitive science, autocatalytic networks represent words or concepts as nodes, and connections between them (by way of free association, shared features, or co-occurrences) as edges. Also, like conventional cognitive networks, RAF networks are decentralized yet hierarchical and can be analyzed with respect to size, density, and connectedness, and the notion of *spreading activation* through nodes of a concept graph is central (Collins & Loftus, 1975; Collins & Quillian, 1969; Koponen, 2021). Finally, like conventional cognitive networks, techniques such as clustering analysis and shortest path distance can be used to draw conclusions about cognition. However, the RAF approach differs from other network approaches to cognition in the following respects:

'Reactions' and 'Catalysis'

A RAF consists of, not only nodes connected by edges, but also *reactions,* or interactions (sometimes depicted as a second kind of node) which generate new nodes (such as the interaction between the concepts BEANBAG and CHAIR, resulting in the invention of the beanbag chair). Thus, although in a biological context the term reaction refers to an interaction between molecules, the term *reaction* is used here in a more general way to refer to an interaction between network elements. In a cognitive network, a reaction may involve *representational redescription* (RR): the re-coding of information in working memory by elaborating, modifying, restructuring, and/or performing mental operations upon it, sometimes in the absence of an external cue (Karmiloff-Smith, 1992). RR can involve a subtle shift of perspective, a flash of insight, or a newly perceived application for an old idea.[2] The question of which concepts participate in a given reaction is discussed and mathematically modeled in (Gabora & Steel, 2020b).

RAFs also have two types of edges: reaction edges and catalysis edges. *Reaction edges* function much like edges in other network science approaches; they can be thought of as representing the 'anatomy' of the network. *Catalysis edges* play a more dynamic role; they can be thought of as representing the 'physiology' of the network. The elements are *catalytic* because they not only participate in certain reactions but also facilitate, or catalyze, other reactions.[3] In origin of life applications, a catalyst speeds up a chemical reaction that would otherwise occur very slowly or not at all. In cognitive models, the notion of catalysis enables us to model how one idea (or perspective, context, or feature of the environment) triggers a mental operation that would otherwise occur very slowly or not at all, such as the combining of two concepts, or the redescription of a mental representation. For example, the galvanizing impact that throwing a beanbag to a baby could have had on the invention of the beanbag chair would be depicted as a catalysis edge. As another example, we would say that the Wright Brothers' invention of wing-warping (which enabled the invention of the airplane) was 'catalyzed' by their observation that bicyclists bank, or lean inward, as they round a corner. As in biology, the cognitive version of a 'catalyzed reaction' may trigger another such reaction, culminating in a *reaction sequence.* In the cognitive case, the reaction sequence is a stream of thought, which may ultimately stem from cognitive dissonance, or a problem, question, or desire. For example, the ultimate source of the cognitive reaction sequence culminating in the beanbag chair would have been the desire to invent a new kind of informal chair, and the ultimate source of the sequence culminating in the concept of wing-warping was the problem of inventing a flying machine. Thus, in RAFs, nodes are not

[2] Creative insights (i.e., those that make significant contributions to culture) often arise subconsciously from just beyond the confines of working memory (Bowers, Farvolden, & Mermigis, 1995).

[3] We note that use of the word 'catalyzing' in a cognitive context extends beyond autocatalytic models of cognition (Beghetto & Jaeger, 2021; Cabell & Valsiner, 2016) (though these other approaches are purely descriptive).

merely passive recipients of spreading activation; they actively redirect it. The resulting network is dynamic both in terms of structure (e.g., new nodes can be generated), and information flow (e.g., either external input, or newly-generated nodes, can result in novel paths of activation).

The rationale for treating MRs as catalysts comes, in part, from the literature on concepts, which provides extensive evidence that when concepts act as contexts for each other, their meanings change (Barsalou, 1982; Hampton, 1988). For example, a child will run from a BEAR, but hug a TEDDY BEAR. TEDDY momentarily reconfigures the cognitive network, altering the perceived meaning of BEAR. Such alterations in meaning are often nontrivial, and defy classical logic (Osherson & Smith, 1981). However, quantum models of concept interactions have provided a means of formalizing the process by which a context (such as the goal of inventing a comfortable chair) creates a transitory 'wormhole' of sorts that spontaneously bridges remote associates (such as BEANBAG and CHAIR) (Aerts, Aerts, & Gabora, 2009; Aerts, Broekaert, Gabora, & Sozzo, 2016; Aerts, Gabora, & Sozzo, 2013). Although the RAF approach is influenced by how context is modeled in quantum approaches to concepts (Aerts et al., 2016, 2013), it is not committed to any formal approach to modeling context. Context is considered to be anything external (e.g., an observation of a scene, object, or person) or internal (e.g., other MRs) that influences the instantiation of a MR in working memory. The extent to which one MR modifies the meaning of another is referred to here as its *reactivity*.

Foodset versus Foodset-derived

Another key feature of RAFs is the distinction between *foodset* items that come into existence outside the network in question, and *foodset-derived* items that come about through interactions between elements within the network in question. Thus, in the context of a cognitive networks, the term *foodset* refers to MRs that are not generated from scratch, i.e., that are innate or obtained through social learning or individual learning (of pre-existing information), while *foodset-derived* refers to MRs that arose through modification of pre-existing MRs (and constitute new information).[4]

In cognitive networks, the distinction between foodset and foodset-derived provides a natural means of grounding abstract concepts in direct experiences; foodset-derived elements emerge through interactions, described as 'reactions,' that can be traced back to foodset items. This in turn enables us to model how new understandings emerge out of available MRs, identifying the required precursor ideas for the emergence of a given idea, and the mental operations carried out by a given individual that result in that new idea. At the level of modeling lineages of cultural descent, the distinction between foodset and foodset-derived makes it possible to tag novel insights with their point of origin, track cumulative change within (and across) individuals, develop cultural evolutionary lineages (Gabora & Steel, 2017, 2020a, 2021a), and follow trajectories in cognitive development step by step.

Potential to Scale Up

To facilitate illustration of the RAF approach, the cognitive development example used in this paper is fairly simple. However, a significant strength of the approach for cognitive science is that RAFs have the potential to scale up. RAF algorithms can be used to analyze and detect phase transitions in vastly complex networks (such as the phase transition from no-RAF to RAF in Kauffman's (Kauffman, 1993) binary polymer model) that have proven intractable using other analytic approaches (Sousa, Hordijk, Steel, & Martin, 2015; Xavier et al., 2020).

Network Science Approaches to Cognitive Development

Of particular relevance to the goal of understanding the emergence of conceptual structure in children is the application of network science to the acquisition of lexical knowledge (Beckage & Colunga, 2016, 2019; Beckage et al., 2011; Stella, Beckage, & Brede, 2017; Steyvers &

[4] This distinction may not be so black and white as portrayed here, but for simplicity, we avoid that subtlety for now. In other words, the approach distinguishes between conceptual shifts that originate within the mind of a given individual, and those that originated prior to their assimilation by that individual. In other words, foodset-derived MRs are generated by the individual, as a result of mental operations such as concept combination, restructuring, deduction, induction, or divergent thinking.

Tenenbaum, 2005). This literature shows that children who acquire language at a slower rate than their peers not only have fewer words, but their words are less connected, suggesting that cognitive network structure affects new concept acquisition. In much of this work, network science is used to model language acquisition trajectories (e.g. Beckage and Colunga (2016); Beckage et al. (2011); Stella et al. (2017); Steyvers and Tenenbaum (2005)), and individual differences in the lexicon (Beckage & Colunga, 2019). These process models address how knowledge is structured, but less attention is paid to how conceptual structure emerges, and how this knowledge is used. These models may or may not generalize to learning more broadly, and are limited in their ability to offer insights beyond lexical acquisition. If one wants to understand concept learning and formation, it is important to study the origin and spontaneous deployment of concepts.

Another area relevant to understanding the emergence of conceptual structure is the application of network science to creativity (Benedek et al., 2017; Benedek & Neubauer, 2013; Kenett & Faust, 2019; Kenett et al., 2018). This research focuses on explaining cognitive processing using an underlying concept network, e.g., work on the combining of knowledge representations to generate novel combinations and solutions (Benedek et al., 2017; Benedek & Neubauer, 2013; Kenett & Faust, 2019; Kenett et al., 2018; Vukić, Martinčić-Ipšić, & Meštrović, 2020). This research has brought to light a wealth of interesting findings, such as that high creative, high intelligence individuals have more richly connected concept graphs, suggesting that they navigate conceptual space more effectively (Benedek et al., 2017; Kenett et al., 2018). A limitation of this direction is that it does not adequately capture the emergence of self-modifying structure, and its role in guiding the spontaneous, context-driven deformation of the cognitive network so as to facilitate insight, as discussed above.

This paper bridges research on the acquisition of lexical knowledge in children with research on creativity. This paper proposes a model of how lexical knowledge self-organizes into a cognitive network, such that existing knowledge can be adapted to new situations, and new experiences can be reframed in terms of old ones.

By studying the emergence and connections between clusters of concepts, we provide a novel interpretation of conceptual change in child development. More specifically, our network analysis offers a new interpretation of experimental results on conceptual change in children's mental representations of the shape of the Earth (Vosniadou & Brewer, 1992) that describes a potential process of learning and acquisition underlying the observed data. Thus, we address the question of how the child reconciles personal lived experience with socially transmitted scientific knowledge to arrive at the conclusion that the world is spherical. Unlike the network science approaches discussed above, the approach used here naturally tags representations with their source, allowing study of how conceptual structures emerge and transform in both individuals and groups.

Reflextively Autocatalytic Foodset-Generated Networks (RAFs)

We now define the term *Reflexively Autocatalytic and Foodset-generated* network (RAF) more precisely (Hordijk & Steel, 2004, 2015, 2016; Steel, 2000; Steel et al., 2019). The term *reflexive* is used in its mathematical sense to mean that each element is related (directly or indirectly) to the whole. As mentioned earlier, the term *autocatalytic* refers to the fact that the whole can be reconstituted through interactions amongst its elements. More precisely, two key properties of such a network are that:

(1) It is *reflexively autocatalytic*: each reaction $r \in \mathcal{R}'$ is catalyzed by at least one element type that is either produced by $\mathcal{R}'$ or is present in the foodset F. This is sometimes referred to as *closure*.

(2) It is *F-generated*: all reactants in $\mathcal{R}'$ can be generated from the foodset F by using a series of reactions only from $\mathcal{R}'$ itself.

A RAF is therefore a non-empty subset $\mathcal{R}' \subseteq \mathcal{R}$ of reactions that satisfies these two properties: it is both reflexively autocatalytic, and F-generated.

The term *catalytic reaction system* has been used to refer to a network of components with catalytic abilities, which may consist of one or more RAFs. The largest RAF in the network, which subsumes all other RAFs, is called the *maxRAF*. The remaining RAFs are called *subRAFs*. The catalytic reaction system is a tuple $\mathcal{Q} = (X, \mathcal{R}, C, F)$ consisting of a set X of types, a set $\mathcal{R}$ of reactions, a catalysis set C indicating which molecule types catalyze which reactions, and a subset F of X called the foodset. A subset $\mathcal{R}'$ of the full reaction set $\mathcal{R}$ of a catalytic reaction system $\mathcal{Q}$ forms a RAF if is simultaneously self-sustaining (by the F-generated condition) and (collectively) autocatalytic (by the RA condition; as each of its reactions is catalyzed by an element of the RAF).

A RAF emerges in a system of interacting components when the complexity of these components reaches a certain threshold (Kauffman, 1993; Mossel & Steel, 2005). In the OOL context, the components are polymers: molecules consisting of repeated units called monomers. In the cognitive context, the components are mental representations (MRs) composed of features or dimensions. RAF theory provides a means of identifying RAFs and related structures, and analysing the probabilities of finding them under different conditions (model parameters). It has proven useful for identifying how phase transitions might occur, and at what parameter values. The phase transition from no RAF to a RAF incorporating most or all of the components depends on (1) the probability of any one component catalyzing a given reaction that forms another, and (2) the maximum number of elements (e.g., monomers or features) per component. The phase transition from no RAF to a RAF has been formalized and analyzed (mathematically and via simulations), and applied to biochemical (Hordijk et al., 2010, 2011; Hordijk & Steel, 2004, 2016; Mossel & Steel, 2005), ecology Cazzolla Gatti, Fath, Hordijk, Kauffman, and Ulanowicz (2018), and cognitive (Gabora & Steel, 2017, 2020a, 2020b) systems.

Hierarchical Structure

If the network contains a RAF, then the collection of all its RAFs forms a partially ordered set (i.e., a poset) under set inclusion, with the maxRAF as its unique maximal element. In other words, a catalytic reaction system need not have a RAF, but when it does, there is a unique maxRAF. Moreover, a catalytic reaction system, 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 (Hordijk & Steel, 2016; Vasas, Fernando, Santos, Kauffman, & Szathmáry, 2012).

There are a number of means by which RAFs can enlarge and combine. The union of any two (or more) subRAFs is a RAF (which explains why there is a unique maximal RAF). These two subRAFs could be disjoint, or could have some reactions in common. By contrast, there may be a large number of minimal RAFs present, i.e., RAFs that cannot be broken down into smaller RAFs. These are called irreducible RAFs, or *irrRAFs*. Another way a subRAF R' can expand is by combining with a 'co-RAF', where a *co-RAF* is any nonempty set of reactions that is not a RAF but, when combined with R', it forms a RAF.

RAF expansion can also be extrinsically driven, for example, by a change in the foodset (e.g., a new environmental stimulus), or in the reactions (such as if participants are instructed to 'think creatively'). In a cognitive development context, this could for example take the form of of a question that probes the child's mental model of the shape of the Earth, such as: if you keep walking on the Earth, would you eventually fall off?

Cognitive RAFs

In a cognitive context, 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 $x \in X_i$. MRs are either *foodset MRs*, or *foodset-derived MRs*. The *foodset* of individual i, denoted F_i, encompasses MRs that are either innate, or that result from direct experience in the world, including natural, artificial, and social stimuli. Thus, F_i has multiple components:

- $\mathbb{S}_i$ denotes the set of MRs arising through direct 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 $\mathbb{S}_i[x_j]$.

  - MRs obtained through individual learning, denoted $\mathbb{S}_i[\ell]$.

- Any *innate knowledge* with which individual i is born, denoted I_i.

F_i 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 that the Earth is spherical, this is an instance of social learning, and the concept SPHERICAL Earth is therefore a member of F_i. It includes everything in the long-term memory of individual i that was not the direct result of individual i engaging in RR. F_i also includes pre-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.

The foodset is related—but not identical—to *core knowledge*: knowledge that emerges over evolutionary time that is shared amongst humans (and potentially other species), as opposed to MRs that are acquired within a human lifespan via learning (Barner & Baron, 2016; Spelke & Kinzler, 2007). The RAF approach makes a finer distinction between knowledge that undergoes RR, resulting in the generation of new information (modeled as a ‘reaction’), and knowledge that does not, which is part of the foodset. For example, if a child learns to distinguish different kinds of trees, that knowledge is part of the foodset. However, if the child independently arrives at a new way of classifying trees based on these distinctions, this knowledge is foodset-derived.

Foodset-derived elements are denoted $\neg F_i$. Thus, $\neg F_i$ 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 $r \in R_i$). 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, foodset-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 thought processes. $\neg F_i$ may include a 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).

A single instance of RR in individual i is referred to as a *reaction,* and denoted $r \in \mathcal{R}_i$. RR is often applied recursively, such that the output of one thought serves as the input to the next. The set of reactions that can be catalyzed 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 $b \in B$. Thus, for example, if a child has the idea of surprising her father by sculpting a dog out of snow, the concepts SNOW and DOG are reactants in A, and the resulting concept SNOW-DOG is a product in B. This conceptual shift, treated as a 'reaction', is 'catalyzed' by the child's desire to surprise her father. It is in this way that the RAF approach tags novelty with its point of origin.

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 $X_i \oplus \mathcal{R}_i$. 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 kinds of interactions are possible between the same MRs. Nevertheless, it makes perfect mathematical sense to talk about $\mathcal{R}_i$ and C_i as sets. In the Appendix we have provided a table (Table 1), which summarizes the terminology and correspondences between the two applications of this abstract mathematical framework: (1) the origin of life, and (2) the cognitive development process that culminates in a new participant in human culture.

A RAF Model of Conceptual Change

A model of conceptual change aims to capture how MRs—consisting of compounds of concepts—are learned, and how their inter-relationships change during development. In capturing how concepts are learned, one must account for how the underlying knowledge base is altered by the addition of new concepts. Our RAF model of developmental change will focus on children's mental models of the shape of the Earth, using data and analysis from a well-known study (Vosniadou & Brewer, 1992). In this study, researchers asked 50 first-, third-, and fifth-grade children increasingly probing questions to ascertain the child's mental representation of the Earth. From the child's responses, the researchers categorized each child according to the kind of mental model of the Earth they held. The different models are illustrated in Figure 2. We describe these categories and the transitions between them using the RAF model. Although children tend to develop increasingly accurate and sophisticated mental models, they do not necessarily transition smoothly between them; each child may have a unique path of conceptual change. Nevertheless, since the RAF approach makes explicit how new ideas grow out of currently available knowledge, and how transitions in conceptualizing and understanding may arise, it is fairly straightforward to parsimoniously account for this data using the RAF approach.

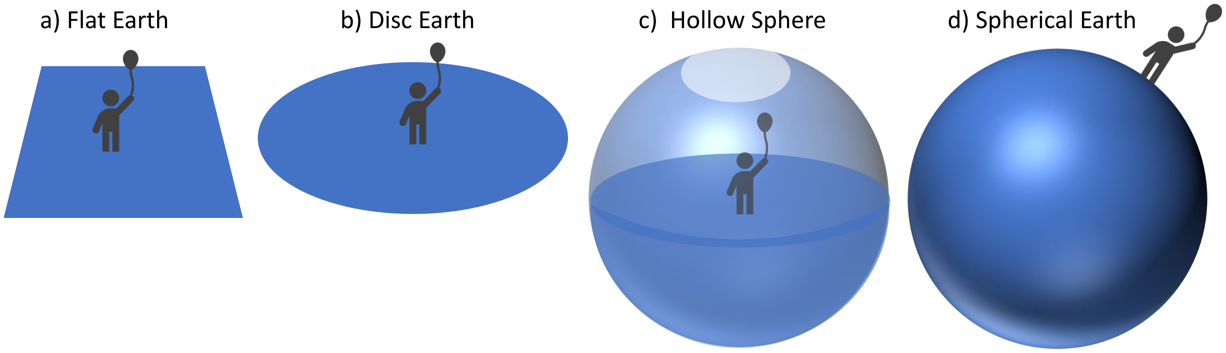

Figure 2. Visual depiction of the mental models of the Earth (a) Flat Earth. (b) Disc Earth: the child understands that Earth is round. (c) Hollow Earth: the child understands that the Earth is spherical, and invents a conception of the Earth that incorrectly reconciles this with the experience of flatness. (d) Spherical Earth: the child realizes that Mars is a planet, and spherical, and Earth is also a planet, so it must be spherical. Dual Earth combines mental representations of the Flat Earth or Disc Earth with concepts of the spherical Earth but maintains two separate representations.

Flat Earth

A young child's direct experience suggests that the Earth is flat; thus, the child's initial mental model is the Flat Earth model, depicted in Figure 2(a). Therefore, the child's earliest mental model of the Earth is described as one for which all mental contents are members of the foodset, as depicted in Figure 3(a).

Where the child in question is denoted individual i, we say that E_i represents the the child's *experience* of being bound to an Earth that appears to be more or less flat. We describe this initial model of the Earth as a flat, solid surface that people live on, as follows:

$$F_i \mapsto F_i \cup \{E_i\}, \text{ where } E_i \in \mathbb{S}_i(\ell) \tag{1}$$

The model described by equation 1 becomes more fleshed-out once the child learns the word 'Earth', as shown in our RAF model in Figure 3(b). Where the caregiver who taught the word 'Earth' to child i is denoted k, we describe this expanded Flat Earth model, denoted E'_i, as follows:

$$F_i \mapsto F_i \cup \{E'_i\}, \text{ where } E'_i \in \mathbb{S}_i(E'_k) \tag{2}$$

Although E'_i is connected by association to concepts such as LIVING, SELF, and Earth, these associations did not arise through abstract thought, but through observation of experiences in the external world.

The Flat Earth model consists only of foodset MRs; the set of derived MRs is empty because there are no MRs playing the role of reactants in $\mathcal{R}$ or catalysts in C. Therefore, $\mathcal{Q} = (X, F)$. Although the Flat Earth model does contain a very simple network, it contains no RAFs because (as we saw earlier), a necessary (though not sufficient) condition for something to be a RAF is that it be a set of one or more RR 'reactions'. Of course, some MRs in a child's Flat Earth model may participate in reactions with MRs that are not related to the concept of Earth, so the child's cognitive network as a whole may contain RAFs. However, to illustrate how

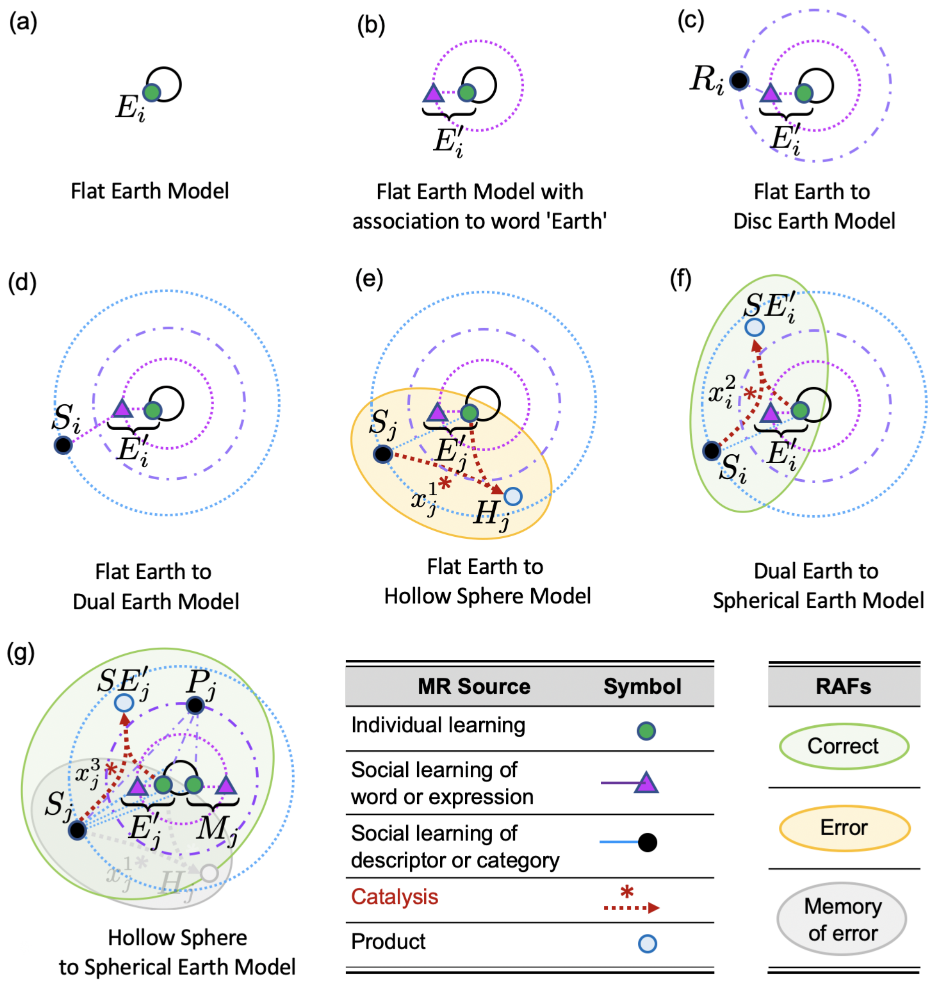
(a)
E_i
Flat Earth Model
(b)
E'_i
Flat Earth Model with association to word 'Earth'
(c)
R_i
E'_i
Flat Earth to Disc Earth Model
(d)
S_i
E'_i
Flat Earth to Dual Earth Model
(e)
S_j
E'_j
x_j^1
H_j
Flat Earth to Hollow Sphere Model
(f)
SE'_i
x_i^2
S_i
E'_i
Dual Earth to Spherical Earth Model
(g)
SE'_j
P_j
x_j^3
S_j
E'_j
M_j
x_j^1
H_j
Hollow Sphere to Spherical Earth Model
MR Source
Symbol
Individual learning
Social learning of word or expression
Social learning of descriptor or category
Catalysis
Product
RAFs
Correct
Error
Memory of error

Figure 3. RAF model of mental models of the Earth. The fact that associations can be context-specific is indicated by way of nested circles differentiated by color and by patterns of dashes and dots. Context-specific associations are indicated by spokes connecting MRs on the relevant circles. Socially transmitted MRs (e.g., MARS, PLANET, or SPHERE) reside on these socially reified layers of abstraction, whereas information obtained through RR or abstract thought need not. (a) Direct experience of living on a flat Earth. (b) Association of the experience of a flat Earth with the word ‘Earth’. (c) Disc Earth: child understands that Earth is round. (c) Dual Earth: the child interprets claim that Earth is spherical to mean that the word Earth refers to two different things. (e) Hollow sphere model: the child understands that the Earth is spherical, and invents a conception of the Earth that incorrectly reconciles this with the experience of flatness. Orange oval represents the resulting unstable RAF. (f) Dual Earth to spherical Earth: child learns that the Earth is a sphere and realizes it must be such a large sphere as to appear flat. Green oval represents the resulting stable RAF. (g) Hollow Earth to Spherical Earth: the child realizes that Mars is a planet, and spherical, and Earth is also a planet, so it must be spherical. Green oval represents the resulting stable RAF. Hollow Earth conception is gray to indicate that, although it may still be accessible in memory, it is understood to be incorrect.

conceptual structure emerges we are modeling an isolated fragment of the whole.

Disc Earth

Some children believe in a variant of the Flat Earth model referred to as the *Disc Earth model*, depicted in Figure 2(b). This model additionally incorporates the socially transmitted information that the Earth is round. This would appear to be a reasonable conception of the shape of the Earth if ‘roundness’ were explained to the child by drawing a circle on a piece of paper, or on a blackboard; the child could quite sensibly interpret this to mean that the Earth is disk-shaped.

The RAF model of the Disc Earth conception is illustrated in Figure 3(c). Just as was done above, to describe transmission of the notion that the Earth is round (disc-shaped), denoted R, from caregiver k to child i, we write:

$$F_i \mapsto F_i \cup \{R_i\}, \text{ where } R_i \in \mathbb{S}_i(R_k) \tag{3}$$

Child i’s conception of the Earth has expanded to include that it is round and finite, though it is incorrect with respect to the assumption of flatness. As with the Flat Earth model, the Disc Earth model contains a very simple network consisting solely of foodset MRs; the set of derived MRs is empty because there are no MRs playing the role of catalysts C or reactants $\mathcal{R}$, and therefore, $\mathcal{Q} = (X, F)$. The disc Earth model contains no RAFs, since there are no RR events.

Dual Earth

Another mental model of the Earth held by some children is the *Dual Earth model* (Vosniadou & Brewer, 1992). The child maintains the mental representation of the flat (and possibly disk-shaped) Earth, but additionally possesses an independent MR of the Earth as a spherical planet, or simply a sphere. For example, Vosniadou and Brewer (1992) quote the following exchange with a first-grader who holds a Dual Earth model:

Experimenter: How come the Earth here is flat but before you said it is round?

> **Child**: Because the Earth is up in the sky and that's (point to the picture of the house) down on the Earth.

These two mental models are related only by way of the word 'Earth,' which appears to be used in two senses (as are other words, such as 'top,' which can refer to a shirt, a kind of toy, or the upper surface of something). The child can refer to each of these mental models when answering questions about the Earth, but has yet to integrate them.

The RAF model of the Dual Earth conception is illustrated in Figure 3(d). The first component—the child's Flat Earth model—is described as per equation 2. We describe the social transmission of the Earth's sphericity, denoted S, from caregiver k to child i as follows:

$$F_i \mapsto F_i \cup \{S_i\}, \text{ where } S_i \in \mathbb{S}_i(S_k) \tag{4}$$

As with the Flat Earth and Disc Earth models, the set of derived MRs is empty because there are no MRs playing the role of catalysts in C or reactants in $\mathcal{R}$; thus, $\mathcal{Q} = (X, F)$. The Dual Earth model contains two separate networks, but no RAFs (again, because there are no RR events).

Hollow Sphere

Some children have what Vosniadou and Brewer (1992) refer to as a *hollow sphere* mental model of the Earth. In this model, the sky constitutes the top hemisphere, the ground where people live is the bottom hemisphere, and people live at the flat interface between the two, as depicted in Figure 2(c). The hollow sphere conception of the Earth is a creative albeit incorrect reconciliation of the Flat Earth model, as described by equation 2, with the socially transmitted information that the Earth is spherical.

The Hollow Earth model illustrates how a catalyzed reaction transforms one or more element(s) of the foodset F_k into a derived MR, i.e., a member of $\neg F_i$. The RAF model of the Hollow Earth mental model is illustrated in Figure 3(e), and it is described as follows. Child j's Flat Earth model is a *reactant*, denoted E'_j. It transforms through RR to the Hollow Earth model, the *product*, denoted H_j. This conceptual shift is provoked, or 'catalyzed' by x_j^1: the desire to reconcile these seemingly contradictory conceptions of the Earth. We describe this as follows:

$$E_j, S_j \in F_j. \tag{5}$$

$$E_j' + S_j \xrightarrow{x_j^1} H_j \in \neg F_j,\ \neg F_j \mapsto \neg F_j \cup \{H_j\}. \tag{6}$$

The set of derived MRs is no longer empty because of the presence of a product, H_j. The Hollow Earth model contains one network, and one RAF, and $\mathcal{Q} = (X, \mathcal{R}, C, F)$. Since this RAF cannot be broken into smaller RAFs, it is an irrRAF. The Hollow Earth model is unsustainable because, sooner or later, the child learns that the Earth is a solid sphere, which is inconsistent with the notion that the upper hemisphere consists of the sky.

Spherical Earth

In (Vosniadou & Brewer, 1992), achievement of the understanding that there is a spherical Earth on which we live was assessed using prompts such as 'Show me where the moon and stars go' and 'If you walked for many days in a straight line, where would you end up?' The Spherical Earth model is illustrated in Figure 2(d). We now consider two feasible routes by which a child could, starting from a Flat Earth model, arrive at a Spherical Earth model: by way of the Dual Earth model, and by way of the Hollow Earth model.

Dual Earth to Spherical Earth

Consider the situation in which child i passes directly from the Dual Earth model described by equation 4 to the mature conception of the Earth as a spherical planet, denoted SE_i, without ever entertaining the Hollow Earth model described by equation 6. This transition, illustrated in Figure 3(f), occurs through integration of two relatively independent conceptions of the Earth: the Flat Earth, model denoted E_i', and the socially transmitted notion that the Earth is spherical, denoted S_i'. The transition to the Spherical Earth model could be described as a RR event catalyzed by the realization that if the spherical Earth were sufficiently large compared to the size of humans, from the human perspective, it would appear flat. We denote this catalyst, x_i^2, and describe the RR event that culminates in the Spherical Earth model, denoted SE_i, as follows:

$$E_i' + S_i \xrightarrow{x_i^2} SE_i,\ \neg F_i \mapsto \neg F_i \cup \{SE_i\} \tag{7}$$

For the Spherical Earth model, the set of derived MRs is no longer empty because there is a product, SE_i. There is also a reactant, and a catalyst, so $\mathcal{Q} = (X, \mathcal{R}, C, F)$. There is one network, and one RAF, an irrRAF. Since each reaction $r \in \mathcal{R}'$ is catalyzed by at least one element type that is either produced by $\mathcal{R}'$ or is present in the foodset F, this model is reflexively autocatalytic (as defined above).

Hollow Earth to Spherical Earth

We now consider the situation in which child j passes from the Hollow Earth model described by equation 6 to the mature conception of the Earth as spherical, denoted SE_j. The child must realize that Earth's sphericity is not composed of two hemispheres, sky and ground; rather, the solid surface of the Earth is spherical.

There are multiple trajectories by which this could be achieved. One consists of learning that (i) Mars is a planet, (ii) a planet is a solid spherical body, (iii) the Earth is also a planet, followed by the realization that (iv) the Earth upon which we walk around and live our lives is a solid sphere. This transition, illustrated in Figure 3(g), is modeled as follows.

Steps (i) to (iii) are modeled as social learning processes (as in equations 2, 3, and 4). The concept that Mars is a planet is denoted M, the concept that a planet is a solid spherical body is denoted P, and the concept that the Earth is also a planet is denoted EP. To describe the transmission of these concepts from caregiver k to child i, we write:

$$F_j \mapsto F_j \cup \{M_j\},\ \text{where } M_j \in \mathbb{S}_j(M_k) \tag{8}$$

$$F_j \mapsto F_j \cup \{P_j\}, \text{ where } P_j \in \mathbb{S}_j(P_k) \tag{9}$$

$$F_j \mapsto F_j \cup \{EP_j\}, \text{ where } EP_j \in \mathbb{S}_j(EP_k) \tag{10}$$

The fourth step is modeled as a RR event (as in equation 6):

$$E'_j + S_j \xrightarrow{x_j^3} SE_j, \ \neg F_j \mapsto \neg F_j \cup \{SE_j\} \tag{11}$$

Further Conceptual Change

We have isolated the example of childrens' conceptions of the shape of the Earth to illustrate how RAFs can describe conceptual change. This example can be described using at most one RAF at a time. However, the RAF approach can scale up to describe the more complex conceptual structure that emerges as children explore their worlds, acquire new information, and creatively play with this knowledge, exploiting affordances, and weaving it into a coherent whole. Consider the situation illustrated in Figure 4, in which child j learns from a teacher that Mars is red, and child j has balloons, one of which is red. Activation of the color RED acts as a catalyst, x_j^4, that stimulates the idea of blowing up a red balloon and pretending it is Mars, thereby generating the product MARS-BALLOON, MB_j. Learning that Mars is red, and conceiving the idea of a MARS-BALLOON, are treated as instances of social learning and RR, respectively, and together, they result in a new RAF. As can be seen from Figure 4, both this new RAF and the earlier RAF are now subRAFs of the maxRAF. Both subRAFs are also irrRAFs, since they cannot be subdivided into smaller RAFs.

Individual Differences

Building on the proposal that the cognitive networks of individuals lie on a spectrum from self-made to socially-made (Gabora, 2019), we suggest that individual differences in reliance on individual learning, social learning, and abstract thought, culminate in RAF networks with

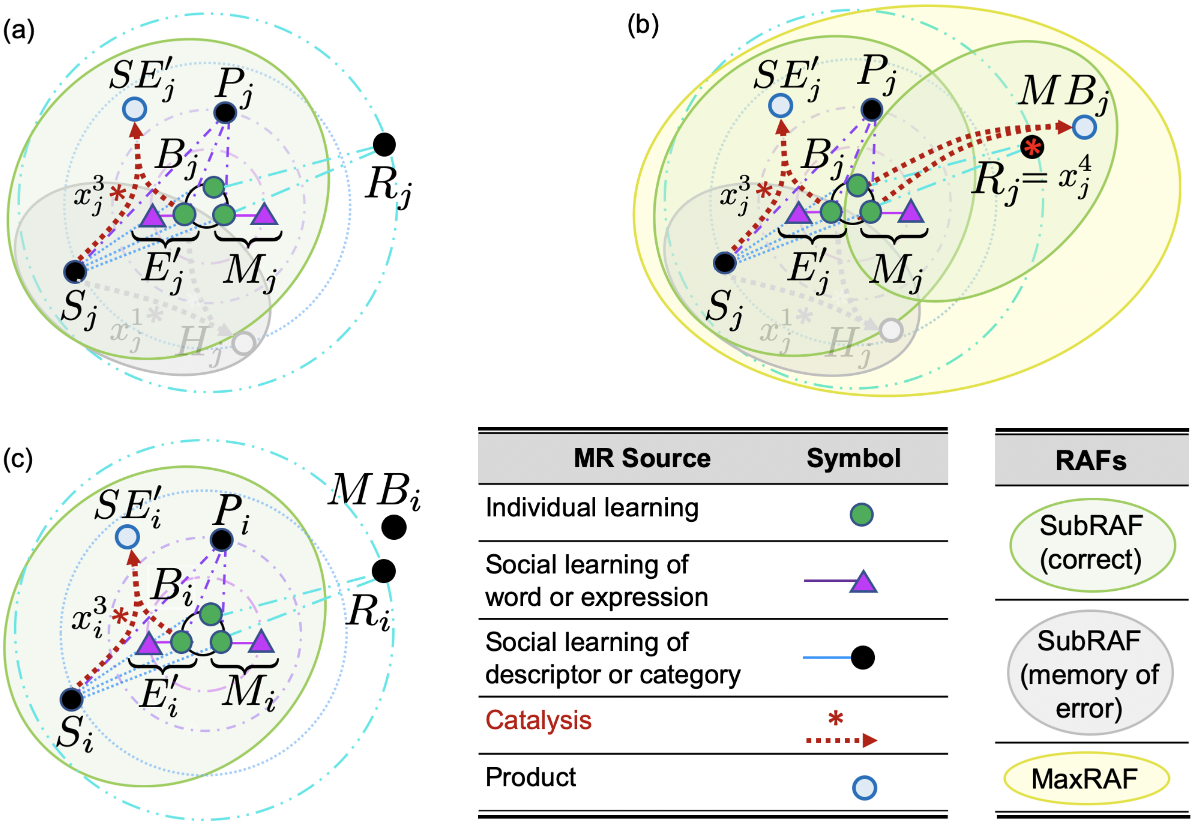

Figure 4. (a) Building on the conceptual structure depicted in Figure 3, individual j acquires through social leaning the knowledge that Mars is red. Individual j has balloons, one of which is red. The knowledge that Mars and the balloon are red is represented by the dashed-and-dotted line connecting M_j and B_j to R_j. (b) The knowledge that Mars is red 'catalyzes' the idea of pretending that a red balloon is Mars; thus, R_j serves as the catalyst, x_j^4. This results in another RAF, which has as its product, MARS-BALLOON, MB_j. This new RAF and the earlier RAF are now both subRAFs (pale green) of the maxRAF (yellow). (c) Since individual i obtained the MARS-BALLOON idea through social learning from j, MARS-BALLOON does not cause formation of another RAF in i.

different structures, which could potentially manifest as personality differences. For example, consider the situation wherein child i (perhaps a sibling, or friend of j) obtains the MARS-BALLOON idea through social learning from j. Since (unlike what took place in the mind of child j), in the mind of child i there is no catalysis event; therefore, MARS-BALLOON does not bring about the formation of another RAF. Although the cognitive networks of both i and j contain roughly the same *concepts* (e.g., RED, SPHERE, MARS-BALLOON, and so forth) and *associations* (e.g., MARS is associated with RED), as illustrated in Figure 4(b) versus 4(c), their RAF structure is quite different. Child j has explored the affordances of both balloons and planets, and discovered that a balloon can represent a planet, and that Mars can be represented by a balloon, but child i has not. This distinction is captured in the RAF model by the fact that in child j, RED acted as a catalyst, MARS and BALLOON have served as reactants, and the new derived MR, MARS-BALLOON has been the product of a conceptual reaction culminating in the generation of a new RAF. None of this occurred in child i. If this leader-follower relationship between child i and child j were to become ingrained, we predict that despite that their cognitive networks would be comparable with respect to their 'node-and-link' structure, when analyzed in terms of RAF structure they would be quite different, as illustrated schematically in Figure 5.

This highlights a key difference between RAFs and other network approaches to cognition. By incorporating 'catalysis,' RAFs can model *how* people develop new representations by looking at existing ideas from new perspectives, or by viewing known concepts from different contexts (or combining them). We predict that, having explored the affordances of both balloons and planets, and having discovered that a balloon can represent a planet, and Mars can be represented by a balloon, child j will be more able than child i to flexibly use these and related concepts. For example, we would expect child j to be more likely than child i to come up with the idea of using a blue balloon to represent Neptune, or if the red balloon were to burst, to come up with the idea of representing Mars with a red beach ball. Importantly, the amount and kind of social learning available, and the order in which MRs are acquired, may send children down distinct developmental trajectories and may even account for clear transition points between mental models.

We posit that the rate at which a child transitions between mental models can be accounted for within the RAF model. One aspect of conceptual change is the ability to reconcile ontological commitments with learned facts or general misconceptions about the world (Brown, 1992; Chi, 2008; Englund, Olofsson, & Price, 2017). In some cases, these facts might be easily reconciled based on the order in which MRs enter the graph, but in other cases, flexibility of thought may play a substantive role.

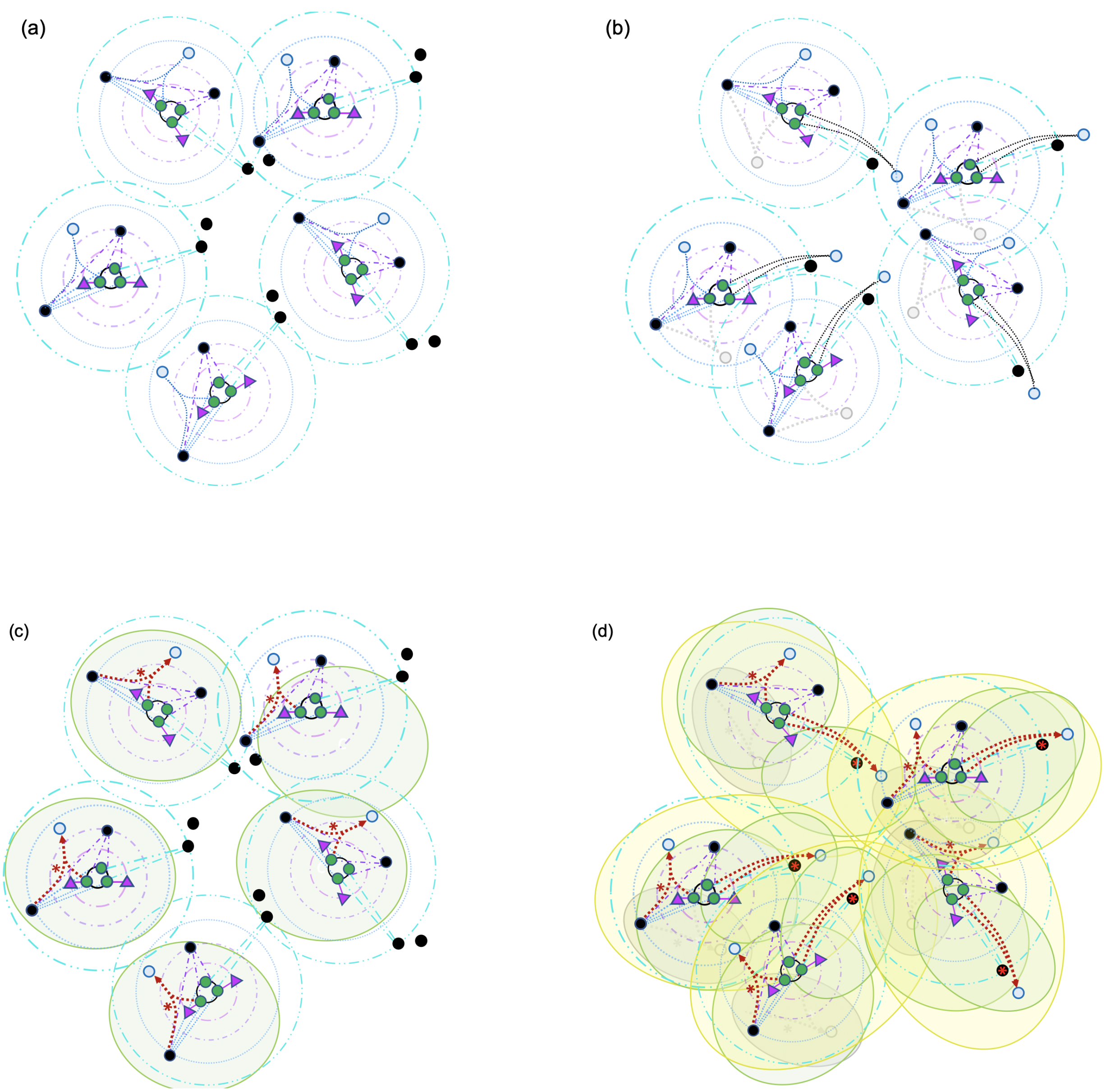

Figure 5. Larger swaths of the cognitive networks of a creative leader, child j, and a follower, child i. Node-and-link depictions of the cognitive networks are given for child i in (a) and for child j in (b). RAF depictions of the cognitive networks is given for child i in (c) and for child j in (d). In terms of nodes and links only, the two cognitive networks appear to be very similar. However, when analyzed in terms of RAF structure, child i's cognitive network is more fragmented than that of child j because there are fewer 'reactions.' For legend, see Figure 4.

Maturity of the Cognitive Network

As RR events increase in frequency, they start to form recursive chains, resulting in streams of thought (Gabora & Steel, 2017), and this is described as follows. We denote each MR in such a process as $m \in M_t$. We say that the content of working memory, denoted $\mathring{w} \in W_t$ is catalyzed by an item $m \in M_t$. This 'reaction' updates the subject of thought, which is now denoted $\mathring{w}' \in W_{t+\delta}$. Each RR step of the stream of thought is denoted:

$$\mathring{w} \xrightarrow{m} \mathring{w}', \text{ and } \mathring{w} \mapsto w \tag{12}$$

The string of these RR steps constituting the stream of thought is described as:

$$\mathcal{C} = \mathring{w}_{t(1)}, \mathring{w}_{t(2)}, \ldots, \mathring{w}_{t(k)} \tag{13}$$

where $\mathring{w}_{t(i)} \in W_{t(i)}$, and where the $t(i)$ values are increasing, such that the product of each reaction, $\mathring{w}_{t(i)}$, serves as the reactant of the next, and brings to mind another item from memory (such as a perspective to be considered), which catalyzes the next reaction. Thus, the stream of thought results in a meta-MR composed of three hierarchically structured MRs, which may themselves be composed of simpler MRs. This is an important step, not just because associations are forged between items in memory, but because abstract concepts tend to become more densely connected through associations than superficial concepts. As extreme examples, the concepts DEPTH and OPPOSITE are relevant to almost every knowledge domain. Concepts such as SPHERE, PLANET, and Earth may be less widely applicable than concepts such as OPPOSITE, but they could potentially be applied to other domains (such as concepts of space, or other forms of life), thereby creating still more associations. Abstract concepts create new affordances for existing concepts (e.g., the concept PLANET could create affordances for GRAVITY that help the child understand other laws of physics, such as FORCE and ACCELERATION). This then further increases the density of associations.

As RAFs continue to increase in size and number, isolated MRs become increasingly rare, as do isolated RAFs. Existing conceptual structure increasingly constrains and enables the scaffolding of new conceptual structure. Reactions at one layer (e.g., factual knowledge, or phonological similarities) may spark chain reactions that lead to activation at other layers (e.g., metacognitive knowledge), a phenomenon discussed elsewhere in relation to self-organized criticality and insight (Gabora & Steel, 2021b). Thus, individual RAFs form, and merge into an expanding maxRAF, which eventually subsumes most of the developing child's conceptual structure.

To formalise this transition mathematically, we first describe a phase-transition result that applies to a general class of large networks. We start with some standard definitions. The set of nodes in any finite network N can be uniquely partitioned into its strongly connected components.[5] Given a network N with n nodes, suppose that for each node u, a directed edge (arc) points to any given node v independently with a fixed probability p. Thus, $\mu = np$ is the expected number of directed edges that lead out of any given node.

For large networks, a phase transition in the structure of N occurs at $\mu = 1$ (Karp, 1990). When $\mu < 1$, the strongly connected components are typically small and isolated, whereas for $\mu > 1$ there is likely to be a single, large, strongly connected component Moreover, the large strongly connected component will be a fixed proportion of the total number of nodes in N. Specifically, if n is the number of nodes in N, then in the limit as n becomes large, the proportion of nodes in N that lie the large component is equal to θ^2 where θ is the unique solution for x in $[0, 1]$ of the equation:

$$1 - x - e^{-\mu x} = 0. \tag{14}$$

Notice that as μ becomes large, the term $e^{-\mu x}$ (for $x \in [0, 1]$) converges to 0, and so Eqn. (14) converges to $1 - x = 0$ which has the solution $x = 1$ (and so $\theta^2 = 1$), which corresponds to the entire network being one single strongly connected component.

This general result is relevant to cognitive networks because if each node corresponds to a small RAF in a cognitive network, and if each directed edge indicates that some element type in

[5] A *strongly connected component* of N is a subset S of nodes of N with the property that for any two nodes u, v in S there is at least one directed path in N from from u to v and at least one directed path in N from v to u (the two paths can involve different sets of nodes).

this RAF catalyses a reaction in another RAF, then (for fixed p) once the network becomes sufficiently large (i.e. n increases so that $\mu = np$ passes the phase transition threshold of 1) the collection of RAFs become linked into a strongly-connected and much larger maxRAF, in which there is a directed path between any two of them. This marks a significant transition, because the child can now reliably frame new knowledge and experiences in terms of previous knowledge and experiences, adapt old responses to new tasks, reflect on ideas from different perspectives, and combine knowledge in appropriate and meaningful ways. The cognitive network becomes self-organizing, and increasingly stable, as genuinely new experiences become rarer; nevertheless, the maxRAF continues to grow as new experiences are assimilated.

General Discussion and Conclusions

This paper used RAFs to model conceptual change, showing how, much as transitions in cognitive evolution were modeled (Gabora & Steel, 2017, 2020a, 2020b), conceptual structure grows, and self-organizing dynamics emerge in a cognitive network. RAFs provide a means of formalizing the notion that cognitive growth may be sparked by cognitive dissonance, questions, or gaps in understanding, which trigger the restructuring or redescription of representations, modeled as 'reactions,' Recursive chains of such cognitive operations are modeled as 'reaction sequences'. Because RAFs tag MRs with their point of origin, it is possible to trace conceptual development within individuals and between them. These features distinguish RAF cognitive models from other cognitive network models (see (Hills & Kenett, 2021)), but we believe them to crucial to understanding the emergence and evolution of cognitive structure.

Using previous research on children's models of the Earth (Vosniadou & Brewer, 1992), we used the RAF approach to parsimoniously explain how transitions between mental models, and different trajectories through these possible models, might exist in a population of learners. The RAF approach enabled us to formulate new hypotheses regarding how individual differences might develop, i.e., some individuals might follow certain trajectories over others based on *what* information they acquire, *how* that information is learned, and *when.* This model was kept simple for the purpose of illustrating the RAF approach. In future work, the approach will be used to analyze vastly larger cognitive networks, capitalizing on its merits relative to other methods for analyzing large networks, and effectiveness for developing efficient (polynomial-time) algorithms for questions that are computationally intractable (NP-hard) (Steel et al., 2019).

Barriers and challenges to RAF models of conceptual development include limitations in childrens' language abilities and attention span. Future research will require the development of methods for identifying what information children have prior to the study (foodset items), assessing what kind of restructuring operations they are capable of, and determining when children produce new ideas (foodset-derived items). Our model did not take into account the role of innate knowledge on the acquisition of a mental model of Earth. In a RAF model, innate knowledge constitutes the initial foodset. Although the model is amenable to the inclusion of such knowledge, the identification of innate knowledge is another potential challenge.

Since innate knowledge is biologically 'expensive,' it may be subject to evolutionary tinkering (i.e., a high mutation rate), so as to continuously assess whether each component is still 'earning its keep.' This leads to the speculative prediction that a mutation that affects innate knowledge may stimulate the individual to develop RAFs that replace that which in others is understood intuitively, thereby rendering normally implicit knowledge explicit, so more amenable to analysis and modification. This is consistent with findings that intellectual giftedness can co-occur with learning disorders (Toffalini, Pezzuti, & Cornoldi, 2017). (Einstein famously said that it was because he lacked the intuitive understanding of space and time that came naturally to others, he was forced to acquire a deeper understanding of these concepts (Einstein, 1949/1999).) The RAF model lends itself to testable hypotheses, such as whether an individual who learned the MR of MARS-BALLOON via RR is more likely to generalize that representation to other types of abstract models than an individual who acquired it through social learning. This hypothesis is reminiscent of those explored in developmental studies of causal reasoning, e.g.(Gopnik & Sobel, 2000; Gopnik, Sobel, Schulz, & Glymour, 2001), where experimenters test generalization and causal understanding when teaching children novel causal relationships via experimentation, instruction, or mimicry. The RAF model offers an *explanation* of how this learning process might happen, and why differences might arise from how things are learned. Other directions for future research include investigating the impact of cognitive development, social environment, and the form and content of socially transmitted information on RAF formation. Narratives that provide catchier or more more compact ways of organizing or understanding information may be describable as RAFs that spread quickly through a kind of cognitive contagion. Another possible direction is to explore whether RAFs could be useful for identifying gaps in knowledge.

We believe the approach holds potential as a unifying formal framework that could bridge theory and findings from archaeology and anthropology with the literature on cognitive evolution and development (Smith, Gabora, & Gardner-O'Kearny, 2018; Voorhees, Read, & Gabora, 2020). Given the successful applications of RAF theory to the origins of both biological evolution and cultural evolution, the approach may provide an integrated theoretical foundation for the origins of evolutionary processes and the self-organizing structures that lie at the heart of both. The RAF approach to cognitive development enables us to address how the child acquires a global pattern of conceptual structure, i.e., how fragments of understanding are woven into an integrated whole. It could be used to model path dependencies in children's learning, or to explore why individuals might differ with respect to their learning trajectories, interests, and aptitudes.

Acknowledgements

This work was supported by a grant (62R06523) from the Natural Sciences and Engineering Research Council of Canada. We thank Megan Foster for some helpful proof-reading suggestions.

Appendix

Table 1 summarizes the terminology and correspondences between the OOL and the cognitive development process that culminates in a new participant in human culture.

Term	**Origin of Life (OOL)**	**Conceptual Development/Origin of Culture**
X_i	all molecule types in protocell i	all mental representations (MRs) in individual i
$x \in X_i$	a molecule in X_i	a 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 representational redescription (RR) in i
$C_i[x]$	reactions catalyzed by x in i	RR events 'catalyzed' by x in i
$(x, r) \in C$	x catalyzes r	x 'catalyzes' 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 R_i (i.e., all B of R_i)

Table 1
Terminology and correspondences between the biological and cognitive/cultural applications of RAFs.

References

Aerts, D., Aerts, S., & Gabora, L. (2009). Experimental evidence for quantum structure in cognition. In P. Bruza, W. Lawless, K. van Rijsbergen, & D. Sofge (Eds.), *International symposium on quantum interaction* (p. 59-70). Berlin: Springer.

Aerts, D., Broekaert, J., Gabora, L., & Sozzo, S. (2016). Generalizing prototype theory: A formal quantum framework. *Frontiers in Psychology (Cognition)*, *7*, 418. doi: 10.3389/fpsyg.2016.00418

Aerts, D., Gabora, L., & Sozzo, S. (2013). Concepts and their dynamics: A quantum theoretical model. *Topics in Cognitive Science*, *5*, 737–772. doi: 10.1111/tops.12042

Barner, D., & Baron, A. S. (2016). *Core knowledge and conceptual change.* New York, NY: Oxford University Press.

Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences*, *17*, 348–360. doi: 10.1016/j.tics.2013.04.010

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & cognition*, *10*, 82–93.

Barton, S. (1994). Chaos, self-organization, and psychology. *American Psychologist*, *49*, 5–14.

Beckage, N. M., & Colunga, E. (2016). Language networks as models of cognition: Understanding cognition through language. *Towards a theoretical framework for analyzing complex linguistic networks*, 3–28.

Beckage, N. M., & Colunga, E. (2019). Network growth modeling to capture individual lexical learning. *Complexity*, *2019*.

Beckage, N. M., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS One*, *17*, e19348.

Beghetto, R., & Jaeger, G. (2021). *Uncertainty: A catalyst for creativity, learning and development.* Berlin, DE: Springer.

Benedek, M., Kenett, Y. N., Umdasch, K., Anaki, D., Faust, M., & Neubauer, A. C. (2017). How semantic memory structure and intelligence contribute to creative thought: a network science approach. *Thinking & Reasoning*, *23*(2), 158–183.

Benedek, M., & Neubauer, A. C. (2013). Revisiting mednick's model on creativity-related differences in associative hierarchies. evidence for a common path to uncommon thought. *The Journal of creative behavior*, *47*(4), 273–289.

Borge-Holthoefer, J., & Arenas, A. (2010). Semantic networks: Structure and dynamics. *Entropy*, *12*(5), 1264–1302.

Bowers, K. S., Farvolden, P., & Mermigis, L. (1995). Intuitive antecedents of insight. In S. Ward & R. A. Finke (Eds.), *The creative cognition approach* (pp. 27–51). Cambridge, US: MIT Press.

Brown, D. E. (1992). Using examples and analogies to remediate misconceptions in physics: Factors influencing conceptual change. *Journal of Research in Science Teaching*, *29*(1), 17–34.

Cabell, K. R., & Valsiner, J. (2016). *The catalyzing mind: Beyond models of causality (volume 11), annals of theoretical psychology.* Berlin, DE: Springer.

Cazzolla Gatti, R., Fath, B., Hordijk, W., Kauffman, S., & Ulanowicz, R. (2018). Niche emergence as an autocatalytic process in the evolution of ecosystems. *Journal of Theoretical Biology*, *454*, 110–117. doi: 10.1016/j.jtbi.2018.05.038

Chi, M. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (p. 61-81). Hillsdale, NJ: Erlbaum.

Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428.

Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247.

Einstein, A. (1949/1999). *Autobiographical notes.* La Salle: Open Court.

Englund, C., Olofsson, A. D., & Price, L. (2017). Teaching with technology in higher education:

understanding conceptual change and development in practice. *Higher Education Research & Development*, *36*(1), 73–87.

Erdös, P., & Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, *5*, 17-61.

Farmer, J. D., Kauffman, S. A., & Packard, N. H. (1986). Autocatalytic replication of polymers. *Physica D*, *22*, 50-67.

Gabora, L. (1998). Autocatalytic closure in a cognitive system: A tentative scenario for the origin of culture. *Psycoloquy*, *9*(67), [adap-org/9901002].

Gabora, L. (1999). Weaving, bending, patching, mending the fabric of reality: A cognitive science perspective on worldview inconsistency. *Foundations of Science*, *3*, 395–428.

Gabora, L. (2019). Creativity and the self-made worldview. In S. Nalbantian & P. Matthews (Eds.), *Secrets of creativity: What neuroscience, the arts, and our minds reveal* (pp. 220–236). Oxford, UK: Oxford University Press. doi: http://doi.org/10.1017/CBO9780511807916.009

Gabora, L., & Steel, M. (2017). Autocatalytic networks in cognition and the origin of culture. *Journal of Theoretical Biology*, *431*, 87–95. doi: 10.1016/j.jtbi.2017.07.022

Gabora, L., & Steel, M. (2020a). Modeling a cognitive transition at the origin of cultural evolution using autocatalytic networks. *Cognitive Science*, *44*, e12878.

Gabora, L., & Steel, M. (2020b). A model of the transition to behavioral and cognitive modernity using reflexively autocatalytic networks. *Proceedings of the Royal Society Interface*, *17*, 20200545. doi: http://doi.org/10.1098/rsif.2020.0545

Gabora, L., & Steel, M. (2021a). Any evolutionary process without variation and selection. *https://www.biorxiv.org/content/10.1101/2020.08.30.274407v1*.

Gabora, L., & Steel, M. (2021b). From uncertainty to insight: An autocatalytic framework. In R. Beghetto & G. Jaeger (Eds.), *Uncertainty: A catalyst for creativity, learning and development.* Springer.

Gopnik, A., & Sobel, D. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205–1222.

Gopnik, A., Sobel, D., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620-629.

Hampton, J. A. (1988). Disjunction of natural concepts. *Memory and Cognition*, *16*, 579–591. doi: 10.3758/BF03197059

Hills, T. T., & Kenett, Y. N. (2021). Is the mind a network? Maps, vehicles, and skyhooks in cognitive network science. *Topics in Cognitive Science*, *115*, 867–872.

Hordijk, W., Hein, J., & Steel, M. (2010). Autocatalytic sets and the origin of life. *Entropy*, *12*(7), 1733–1742. doi: 10.3390/e12071733

Hordijk, W., Kauffman, S. A., & Steel, M. (2011). Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems. *International Journal of Molecular Science*, *12*(5), 3085–3101. doi: 10.3390/ijms12053085

Hordijk, W., & Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, *227*(4), 451-461. doi: 10.1016/j.jtbi.2003.11.020

Hordijk, W., & Steel, M. (2013). A formal model of autocatalytic sets emerging in an rna replicator system. *Journal of Systems Chemistry*, *4*, 3. doi: 10.1186/1759-2208-4-3

Hordijk, W., & Steel, M. (2015). Autocatalytic sets and boundaries. *Journal of Systems Chemistry*, *6:1*. doi: 10.1186/s13322-014-0006-2

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

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* Cambridge, MA: MIT Press.

Karp, R. M. (1990). The transitive closure of a random digraph. *Random Structures and Algorithms*, *1*, 73–93.

Kauffman, S. A. (1993). *The origins of order.* Oxford University Press.

Kenett, Y. N., & Faust, M. (2019). A semantic network cartography of the creative mind.

Trends in cognitive sciences, *23*(4), 271–274.

Kenett, Y. N., Levy, O., Kenett, D. Y., Stanley, H. E., Faust, M., & Havlin, S. (2018). Flexibility of thought in high creative individuals represented by percolation analysis. *Proceedings of the National Academy of Sciences*, *115*, 867–872.

Koponen, I. T. (2021). Systemic states of spreading activation in describing associative knowledge networks: From key items to relative entropy based comparisons. *Systems*, *9*, 20200488.

Kumar, A., Steyvers, M., & Balota, D. A. (in press). A critical review of network-based and distributional approaches to semantic memory structure and processes. *Topics in Cognitive Sciencey*.

Maturana, H., & Varela, F. (1973). Autopoiesis and cognition: The realization of the living. In R. S. Cohen & M. W. Wartofsky (Eds.), *Boston Studies in the Philosophy of Science* (Vol. 42). Dordecht: Reidel.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232.

Mossel, E., & Steel, M. (2005). Random biochemical networks and the probability of self-sustaining autocatalysis. *Journal of Theoretical Biology*, *233*, 327–336. doi: 10.1016/j.jtbi.2004.10.011

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*, 35–58. doi: 10.1016/0010-0277(81)90013-5

Pribram, K. H. (1994). *Origins: Brain and self-organization.* Hillsdale NJ: Lawrence Erlbaum.

Siew, C. S., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, *2019*.

Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R., & Bassett, D. S. (2019). The importance of the whole: topological data analysis for the network neuroscientist. *Network Neuroscience*, *3*, 656–673.

Smith, C., Gabora, L., & Gardner-O'Kearny, W. (2018). The extended evolutionary synthesis facilitates evolutionary models of culture change. *Cliodynamics: The Journal of Quantitative History and Cultural Evolution*, *9*, 84–107.

Sousa, F., Hordijk, W., Steel, M., & Martin, W. (2015). Autocatalytic sets in e. coli metabolism. *Journal of Systems Chemistry*, *6*, 4.

Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96.

Steel, M. (2000). The emergence of a self-catalyzing structure in abstract origin-of-life models. *Applied Mathematics Letters*, *13*, 91–95.

Steel, M., Hordijk, W., & Xavier, J. C. (2019). Autocatalytic networks in biology: Structural theory and algorithms. *Journal of the Royal Society Interface*, *16*, rsif.2018.0808. doi: 10.1098/rsif.2018.0808

Steel, M., Xavier, J. C., & Huson, D. H. (2020). Evolution in leaps: The punctuated accumulation and loss of cultural innovations. *Journal of the Royal Society Interface*, *17*, 20200488.

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, *7*, 1–10.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78.

Toffalini, E., Pezzuti, L., & Cornoldi, C. (2017). Einstein and dyslexia: Is giftedness more frequent in children with a specific learning disorder than in typically developing children? *Intelligence*, *62*, 175–179.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind.* Cambridge MA: MIT Press.

Vasas, V., Fernando, C., Santos, M., Kauffman, S., & Szathmáry, E. (2012). Evolution before genes. *Biology Direct*, *7*(1).

Voorhees, B., Read, D., & Gabora, L. (2020). Identity, kinship, and the evolution of cooperation. *Current Anthropology*, *61*, 194–218.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535–585.

Vukić, Đ., Martinčić-Ipšić, S., & Meštrović, A. (2020). Structural analysis of factual, conceptual, procedural, and metacognitive knowledge in a multidimensional knowledge network. *Complexity*, 9407162.

Xavier, J. C., Hordijk, W., Kauffman, S., Steel, M., & Martin, W. F. (2020). Autocatalytic chemical networks at the origin of metabolism. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *287*, 20192377.