

Evolved to adapt: A computational approach to animal innovation and creativity

Oren KOLODNY^{1*+}, Shimon EDELMAN², Arnon LOTEM¹

¹ Department of Zoology, Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv, 69978, Israel

² Department of Psychology, Cornell University, Ithaca, NY 14853, USA.

Abstract The production of novel behavioral sequences that gives rise to animal innovation and creativity is one of the most intriguing aspects of behavioral evolution. Numerous studies have recently documented the abundance and diversity of innovative and creative behaviors between and within species, yet the ability to innovate or to act creatively has mainly been described and quantified as a measure of animals' cognitive *ability* without explicit reference to cognitive *mechanisms* that may account for these behaviors. Here we discuss the creative process from a computational point of view and suggest such a mechanistic framework. In light of recent research on human creativity, animal learning, and animal problem solving, we suggest that animal creativity is best understood as the production of context-appropriate novel behavioral sequences, which may be facilitated by the ability to learn the regularities in the environment and to represent them hierarchically, allowing for generalization. We present a cognitive framework that we recently developed, which employs domain-general mechanisms and has been used in the modeling of a range of sequential behaviors, from animal foraging to language acquisition, and apply it to behavioral innovation. In a series of simulations, we show how innovation and creative behavior can be produced by this learning mechanism, as it constructs a network representing the statistical regularities of the environment. We use the simulations to demonstrate the role of particular cognitive parameters in this process and to highlight the effects of the learning dynamics and individual experience on creativity [Current Zoology 61 (2): 350–367, 2015].

Keywords Creativity, Innovation, Evolution, Learning, Animal innovation, Behavioral evolution

1 Introduction

1.1 What is creativity?

We begin by considering a working definition of creativity and innovation in animal behavior. First and foremost, as is the case with any theoretical characterization of behavior, such a definition must be functional (see discussions in Reader and Laland, 2003; Reader and MacDonald, 2003; Shettleworth, 2010). A definition that requires *a priori* that behavior be considered creative only if it is the outcome of a particular cognitive or computational process (e.g. Witt, 2009), would thus be irrelevant to the current discussion, as would be one that depends on subjective data (e.g. questionnaires; see discussion in Fleck and Weisberg, 2004).

Following Kaufman and Kaufman (2004), Simonton (2000), and the discussion in Boden (1995), we consider a behavioral sequence creative if: (1) it is novel, in the sense that it had never been produced by that individual and that individual had never been exposed to the behavior, and (2) it is 'good' in some sense (see below).

This is not to say that creative behavior must be completely novel: nearly all creative behavioral sequences are composed of previously known sub-units. In this spirit, we shall argue that the use of generalization and analogy among known behaviors is a major source of creative behavioral sequences (see, e.g., Edelman, 2008; Hofstadter, 1995; Hofstadter and Mitchell, 1995; Lieven, Tomasello et al., 2003; von Bayern et al, 2009). Following Ramsey et al. (2007), we also limit the scope of novel behaviors to those that are not a direct influence of the environment, i.e. novel behaviors are only those that are not induced in all individuals alike by environmental cues, as may occur with instinctive behaviors and reflexes. Note that this definition is very broad, and should thus be applied carefully to different fields and in different contexts: in some realistic settings of animal behavior, for example, it makes sense to define also a repeated behavior as creative if it is in accord with the definition above, despite the fact that technically only the behavior's first execution satisfies it in full. In the context of automated search algorithms, on

Received Nov. 6, 2014; accepted Feb. 16, 2014.

* Corresponding author. E-mail: okolodny@stanford.edu

+ Current address: Department of Biology, Stanford University, Stanford, CA 94305, USA

© 2015 Current Zoology

the other hand, which seek solutions to given problems in a predefined space of possible “behaviors”, our definition would encompass virtually any solution that the algorithm might produce, and so in that context it might not prove so useful.

The demand that creative behavior be 'good' in some sense is one of the constraints on creativity discussed and debated by Boden (1995) and others (e.g. Kaufman and Kaufman, 2004; Simonton, 2000). In animal behavior, this type of constraint has a natural interpretation: for a behavioral sequence to be creative, it must be adaptive (in the classical evolutionary sense), or more accurately, the tendency to produce this behavior must be adaptive on the average. An adaptive behavioral sequence could be, for example, one that increases the likelihood of finding food, navigating successfully, generating an attractive song or courtship display, or successfully communicating with other individuals.

A concept that is closely related to creativity is *innovation*. Definitions of this term vary widely, as discussed in depth by Reader and Laland (2003). Lately, innovation has been invoked in the context of creative behavior at the level of the population and the study of the spread of new ideas and behaviors in social groups (e.g. Allen et al., 2013; Bateson and Martin, 2013; Hoppitt and Laland, 2013; see discussion in Ramsey et al., 2007 and Logan and Pepper, 2007). Given that our focus here is on the cognitive mechanisms that may underlie creative behaviors, we mostly use the term ‘creativity’; the occasional mention of ‘innovation’ refers simply to creative behavior, possibly in a social context.

One of the several posited essential characteristics of creativity is *insight*. Many authors (e.g. Sternberg and Davidson, 1995) have suggested that insight, as manifested in the subject’s “aha!” feeling, is a cornerstone of creative behavior (see discussions in Boden, 2004; Bowden et al., 2005; and a computational view of the role of surprise in the learning process in Schmidhuber, 2006; 2008; 2010; 2013). We refrain from including insight in our definition of creativity, both because we remain agnostic with regard to its objective nature (in particular, given the possibility that the underlying unconscious dynamics are in fact gradual; see, e.g., Kounios and Beeman, 2009; Simonton, 2000; Weisberg, 2006), and because this subjective experience is usually not well communicated to the observer by animal subjects (Reader and Laland, 2003; Shettleworth, 2010). Moreover, the term 'insight' is frequently used in animal behavior to describe a different phenomenon, in which an animal seems to predict the outcome of a behavior

without trying it out in its entirety beforehand (see, e.g., Seed and Boogert, 2013; Shettleworth, 2012; Thorpe, 1956). That said, we will return to discuss insight from a computational perspective, as it does seem that a certain element of surprise is characteristic of most behaviors that we intuitively view as creative.

1.2 Creativity and innovation in animal behavior

Many instances of animal creative behavior have been documented, such as opening of milk bottles by birds (Fisher and Hinde, 1949), activation of different contraptions that contain food by tits, babblers, mynas, and meerkats (e.g. Griffin et al., 2014; Keynan et al., 2015; Morand-Ferron et al., 2011; Thornton and Samson, 2012), a range of types of tool use by primates and corvids (e.g. Bluff et al., 2007; Boinski, 1988; Goodall, 1986; Seed and Byrne, 2010; Visalberghi, 1987), and surprising methods of food handling by various species (e.g. Hosey et al., 1997; Kawai, 1965). Many of these are reviewed in (Reader and Laland, 2003).

The addition of novel behaviors to an organism's repertoire is a double-edged sword: it may have a significant effect on fitness, in both directions. Much work has been done to understand which individuals are most likely to produce creative behavior and when, showing a range of patterns in various species, that may vary across individuals and contexts (see, e.g., Keynan et al., 2015; Kummer and Goodall, 2003; Lefebvre and Bolhuis, 2003; Morand-Ferron et al., 2011; Sol, 2003).

Importantly, a number of studies have shown a strong effect of the details of the individual's experience, and in particular of the individual's tendency to actively explore its environment and its elements, on its propensity to produce a novel behavior (e.g. Bateson and Martin, 2013; Bekoff and Byers, 1998; Bluff, et al., 2007; Burghardt, 2005; Reader and Laland, 2003).

1.3 What kind of computational process can give rise to creative behavior?

To interact effectively with its environment, it is useful for an organism to have a cognitive representation of the environment, of the possible actions that it can perform in it, and of their possible effects (Byrne and Russon, 1998; Conant and Ashby, 1970). A central cognitive challenge in producing behavior is sequencing actions in a manner that is likely to be adaptive (Savage-Rumbaugh and Lewin, 1994; Byrne, 2002). Our interest focuses on this computational task, as well as on the construction of representations that may facilitate the solution of this computational task. That is, we focus on the learning process that allows to construct a representation of the environment in the learner's memory, in

the form – for example – of a hierarchical network (see below), and that can serve for constructing sequences of actions.

It might seem that for an organism that constantly constructs behavioral sequences, being creative is trivial; this is not the case. The vast majority of sequences that can be produced arbitrarily may be physically nonsensical, and almost all of the remaining sequences are extremely unlikely to be adaptive. The search space is typically huge: the combinatorial space spanned by a mere 6-element repertoire of possible behaviors, in the construction of a 4-element sequence, already contains more than 1,200 possible options. A challenge in constructing creative behavior is thus in conducting a biased search in this vast space and in pruning the tree of possibilities effectively so as to eliminate sequences that are unlikely to be adaptive, or in constructing sequences in a biased manner that would bring this outcome about (see, e.g., Boden, 1995; Edelman, 2008; Forth et al., 2010).

Importantly, the structure of the internal representation that serves as the substrate for the construction of new behavioral sequences (e.g. the links connecting nodes in a network representation) strongly influences its ability to give rise to creative solutions (see, e.g., Boden, 2004). This is true for what the representational structure allows and supports, e.g. whether or not the mechanism supports the construction of categories and of concepts at various levels of abstraction; it is also true for the particular instance of representational structure that an individual managed to learn, e.g. whether or not it is a rich enough representation. Both these aspects are touched upon below.

A number of computational frameworks have been used to model creativity in various settings, aimed at providing insight into human creativity (see, e.g., Forth et al., 2010; Hofstadter and Mitchell, 1995; Johnson-Laird, 1991). These studies underscored the importance of hierarchical representations and of the ability to infer similarity among elements and to generalize or draw analogies accordingly. Typically, they leave the task of evaluating the results (e.g., with regard to their aesthetic impact or ‘goodness’ value) to the subjective opinion of the human users of the models, an approach that cannot reveal the fitness benefits of creativity, and is therefore not very useful for the study of creativity in biological, evolutionary contexts. For instance, the CopyCat model of creative analogy (Hofstadter and Mitchell, 1995) was shown to be capable of producing not only high-probability, conservative, solutions but also ones that had a

low probability, yet were deemed striking by the model’s designers. Whether such creativity is evolutionarily adaptive remains to be shown; this may not be easy to demonstrate in the case of human creativity, but may be possible to study in the context of animal behavior.

Notably, Schmidhuber and others (Schmidhuber, 2006; 2008; 2010; 2013) have developed an elaborate computational theory of creativity, suggesting that the essence of creative behavior is in uncovering a previously unknown regularity in the environment, as signified by a newly available means of compressing its representation, and that the search for such regularities can be cast as intrinsically motivated RL. While it is computationally sophisticated and widely appreciated in cognitive psychology, machine learning, and AI, we do not adopt this view of creativity in the present work, since the requirement that all creative behaviors must lead to improved compression of the world representation rules out many behaviors that we deem creative because they are adaptive in other biologically meaningful respects. Additionally, deciding whether a behavior is creative according to Schmidhuber’s definition requires access to the learner’s internal representation, which is difficult or impossible to achieve in studies of animal behavior.

In the next section we present a framework that allows for hierarchical representation and for generalization in both humans and animals. It has been previously developed to account for a range of learning tasks and production of behaviors (Kolodny et al., 2014a, b), and is offered here as a simple, first-approximation computational model of creativity in animals, and in particular of the capacity for “unlikely” yet useful innovation. We then apply the framework to a number of learning and production challenges in a set of simulations. While some existing computational models of reinforcement learning can be adapted so as to replicate the findings detailed below, the mechanism that we propose differs from these in that it is quite simple and easily evolvable from even simpler building blocks, thereby offering a view of creativity that can be of direct use to ethologists. See further discussion in section 5.

1.4 Learning of behavior: The cognitive framework

Computational cognitive models are not common in the study of animal behavior. Typically it is assumed that much of the studied behavior can be explained by simple associative learning (Heyes, 2012), or, in more complex cases, by backward chaining (Ferster and Boren, 1968; Millenson, 1967; Skinner, 1966). In many cases, the learning or decision task that the animal faces

is viewed as a choice among a small number of options such as whether to prefer to peck at one spot or the other, or whether to forage in one of two localities – challenges that can be met by relatively simple learning rules (Bush and Mosteller, 1951; Giraldeau, 1997; Harley, 1981; McNamara and Houston, 1987; Trimmer et al., 2012). Elsewhere (Kolodny et al., 2014a) we have argued that these models and tasks have taught us much about animal learning and behavior, but may be insufficient for handling the broader challenges that animals face in their environment (see also Fawcett et al., 2013; McNamara and Houston, 2009). Relevant to this paper is the fact that such models leave little room, if any, for creativity on the part of the animal. Moreover, laboratory and field studies that do yield examples of animal creativity are typically "model-free" in that they do not attempt to put forth a mechanistic account of the behaviors in question (e.g. Griffin et al., 2013; Aplin et al., 2014).

To fill this gap, we present here an explicit cognitive computational model that learns and represents the world and can accommodate a wide range of behaviors. Importantly, the model's ability to construct sequences in general can also provide a framework for studying the construction of context-appropriate novel sequences of behavior, and hence creativity. This cognitive-computational approach makes it possible to go beyond the description of creative behaviors and the comparison of quantifiable descriptive traits, to the modeling of the cognitive mechanisms that underlie the behavior. This, in turn, leads to possible refinement of the questions that are asked and the learning challenges that are studied, so as to uncover the details of the actual cognitive mechanisms and of the processes involved.

2 The model

The model we propose, LEBS (an acronym for Learning and Exploration of Behavioral Sequences), allows an organism to construct a "grammar of behavior": a representation of elements that comprise the modality of interest and the possible sequential relations among them. This model is based on simple principles of associative learning, but despite its simplicity, it can give rise to many abilities that are thought of as "higher cognition", such as generalization, analogy, and abstrac-

tion. It is an instance of a framework that we have applied with success to learning and behavior in various modalities and environments^①, from learning in songbirds and their composing of song sequences from syllables (Menyhart et al., under revision), through learning regularities in a foraging environment composed of various objects and optimizing the choice of foraging routes among them (Kolodny et al., in review; Kolodny et al., 2014a), to the learning of human language (Kolodny et al., 2014b). For generality, in this paper we apply the model to an abstract modality, most readily envisioned as a modality of possible physical actions. The framework was developed with special attention to biological realism and evolutionary plausibility, considering how simple mechanisms can incrementally evolve into complex ones, as explored and discussed in (Kolodny et al., 2014a, b).

2.1 Learning and representation

A LEBS learner starts out with a clean slate; with time it gradually encounters units that comprise the learned modality. These are extracted from the stream of sensory and motor data and are committed to memory; they are introduced into an internal representation, which is structured as a directed graph: a set of nodes, connected by edges that represent the statistical temporal/sequential relations between them (see simplified example in Fig. 1).

The extraction of units from the data stream is done in a number of ways. First, the data stream may be naturally divisible into basic elements, such as phonemes in language, syllables in birdsong, physical objects in the environment, or movements of body parts (see, e.g., Byrne, 1999). In this case, the learner adds to its representation every new element as it is encountered (henceforth EBL: element-based learning). A second method of extraction of units is applicable also to data streams that are not easily broken down into their constituent units: the learner searches for recurring sequences in the data stream and adds those to its representation (henceforth RBL: recurrence-based learning). This search is conducted only along the sequence of input acquired within a rather short time-window, allowing it to be efficient computationally and memory-wise. This turns out to have an additional functional advantage: a sequence that recurs within a short time-window is un-

^① The LEBS (Learning and Exploration of Behavioral Sequences) model is a version of the framework dubbed CBLS (Kolodny et al., under review) or U-MILA (Kolodny et al., 2014), run with a set of default parameter values.

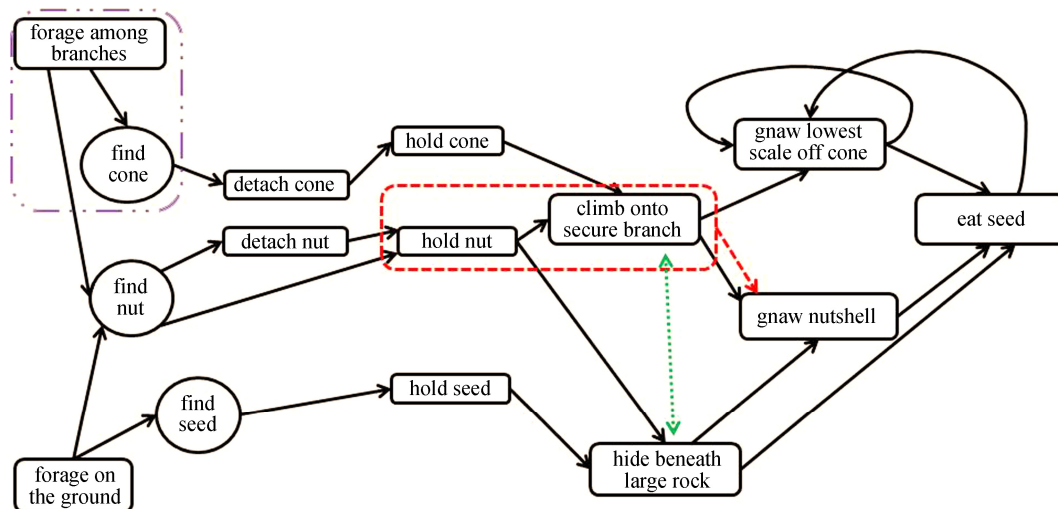


Fig. 1 An illustration of an internal representation of a black rat's actions in the context of handling of food

Nodes denote actions or occurrences and solid edges denote temporal relations. An example of a high-order hierarchy (chunk – see text for definition) is shown in red (dashed) line, signifying that alongside separate representations of holding a nut and of climbing onto a secure branch, this learner has a combined representation of the two actions as a single unit, allowing it to represent the fact that the chunk is always followed by the action of gnawing the nutshell. See main text for a discussion of the need for such units. An example of a bi-directional link, representing high similarity between units, is denoted in green (dotted) line. See detailed explanation at the end of this section. The purple (dash-dot) high-order unit denotes potential entrenchment and is referred to in the results section. For clarity, edge weights are not shown.

likely to be coincidental, and is more likely to reflect a stable and potentially meaningful unit in the learned modality (Goldstein et al., 2010; Lotem and Halpern, 2008, 2012; Onnis et al., 2008a, 2008b). Thus, every such recurring unit is added tentatively to the representation (but decays if it does not recur at a frequency above a certain threshold). Note that in this manner a learner may add to its repertoire long sequences ("chunks") which are themselves composed of smaller units (see discussion below). The third method by which units may be added to the representation is via a bottom-up construction mechanism: if two previously-learned units follow one another frequently (above a certain statistical threshold), a sequence (chunk) which is a concatenation of the two will be added to the representation.

The high-order units, hierarchies, that are added to the learner's repertoire by the two latter mechanisms represent encapsulations of behavioral sequences. This may be advantageous in several respects. First, a certain high-order unit may have a different meaning, reflected in its statistical relations with other units, from a mere concatenation of its constituents. As (a simplified) example, consider a pecking action that can be a precursor of food collection or of an attack of an opponent. The latter is usually preceded by a wing stroke while the former is not. In this case, it would be worthwhile to code the sequence of $\langle \text{wing stroke} + \text{peck} \rangle$ as a unit in

the repertoire, allowing the learner to code its statistical temporal relations with other units separately from those of a peck that is not preceded by a wing stroke (i.e. $\langle \text{wing stroke} + \text{peck} \rangle \rightarrow \langle \text{opponent} \rangle$ versus $\langle \text{peck} \rangle \rightarrow \langle \text{food} \rangle$). Other advantages of high-order encapsulation of sequences are reduced processing time or computational effort and thus faster reaction times, reduced probability of unplanned diversions from an action-plan, and a means of representing highly stereotyped behaviors in which small deviations from the sequence are costly. These advantages are not unique to animal learning and behavior, in which these features may carry high adaptive value: such high-order encapsulation of action-sequences are also commonly used in control structures and flow charts in many industries, from automated production of goods, through robotics, to aircraft control (Harel, 1987, 2007).

Importantly, high-order encapsulations of sequential behaviors can be viewed as an abstraction of the behavior, allowing the learner to process the sequences efficiently, easily learn the similarity between complex behaviors, and combine complex sequential behaviors, in turn, into higher order hierarchical sequences. This idea can be illustrated intuitively on the example of the seemingly simple task of throwing a stick. On the face of it, this requires grasping the stick, picking it up, and flinging it away. In reality, though, each of these steps is composed of a multitude of sub-tasks, such as reaching

for the stick, opening the fingers, and closing them at the right timing, each in itself a complex challenge that involves multiple sensory and motor systems and perfect coordination among them. The abstraction to a small number of complex units that are highly automated (grasp, pick up, fling away) makes a real-time execution of such a complex sequence possible.

As a learner encounters previously-learned units, it updates the representation with regard to their frequency of occurrence and it learns their statistical relations with other units: edges between units denote which ones had been encountered following one another, and each edge is weighted to reflect the frequency with which this happened. An edge's weight thus constitutes an estimate of the actual (real-world) statistical relation between the two nodes that the edge connects.

These concepts are illustrated in Fig. 1, which depicts a rat's simplified representation of actions that are related to food handling (Aisner and Terkel, 1992). Note that multiple high-order chunks and links that represent high similarity among nodes are to be expected in such a representation, if the animal is capable of learning such properties. For visual clarity only a few examples of each type are depicted. Note also that the depicted actions are in reality composed of multiple sub-units; any such representation requires a certain level of abstraction, which, we suggest, is also to be found in real-life cognitive representations (Kolodny and Edelman, submitted).

2.2 Generalization and analogies

Learners in our framework can generalize and draw analogies among units in a rather simple manner based on the internal representations that they construct. They do so by calculating the extent of similarity between pairs of units, as reflected by the contexts in which they are encountered, arriving at an estimated similarity index: If the edge profile of unit A is similar to that of unit B, it may be reasonable to generalize and expect that unit A may follow elements that had preceded unit B, even if A had not been encountered in this context (see details and method of calculating the similarity index in the Electronic Supplementary Material, ESM). Thus, for example, the simple representation in Fig. 1 suggests that *<climb onto secure branch>* and *<hide beneath large rock>* have a high similarity index, as half of their incoming and outgoing links to other nodes are identical (note that a realistic representation would be richer and the similarity index accordingly more reliable). Given this similarity, the rat, while holding a cone and planning its next step, may replace the *<climb onto secure*

branch> with *<hide beneath large rock>*. The realized sequence should be regarded as creative: it is both novel and may be highly adaptive, depending on the rat's location. The decision of whether or not to generalize depends on a parameter in our model (P_{gen}), which is the probability of similarity-based generalization being used in the production of sequences. As we shall further demonstrate below, the ability and the tendency to generalize (based on P_{gen}) may play a crucial role in a learner's ability to act creatively.

2.3 Sequence production

The representation produced as described above can be useful for predicting what is likely to follow, for making choices among multiple options, for the assessment of novel data in relation to existing memories as in the assessment of a courtship sequence, and for a range of other objectives (see also Kolodny et al., 2014a, b). Most importantly for the present study of creative behavior, however, the representation can be used for the production of behavioral sequences. The graph-based representation in the form of directed edges between nodes that represent meaningful units lends itself naturally to producing sequences, including novel ones: a sequence can be composed simply by following a path over the edges of the graph. Various algorithms can be used for this purpose; one simple option is to start out at a certain node (in the simulations below all sequences start with a special node representing "begin"), and choose an edge from among the outgoing edges of that node, taking into account the edges' relative weights (see ESM and Kolodny et al., 2014b for computational details). This choice leads to the next node in the constructed sequence, and the process can be iterated again and again (in the following, a trajectory terminates upon reaching the special node 'end'). In the results section, we present a number of alternative algorithms, alongside the model parameters that govern their action, and their effect is demonstrated.

3 Methods: Examining the Effect of Learning on Creativity in a Simple Cognitive Framework

To demonstrate the use of the proposed cognitive framework for the study of creativity, we ran simulations in which learners were exposed to a set of input sequences and were then required to produce sequences of their own. The input sequences in each simulation followed a set of rules regarding the sequential relations among different elements that can presumably be expe-

rienced in nature, in a scenario – for example – of a young animal watching or joining the behavior of experienced individuals. These take the form of finite automata, and are depicted in Fig. 2.

Importantly, the input datasets never included all sequences that are acceptable (i.e., behaviorally adaptive) in the learned modality, thus leaving room for the production of novel sequences by the learner. The production of a novel adaptive sequence would constitute creative behavior. Note that we henceforth refer to sequences that are acceptable, i.e., that follow the regularities that underlie the sequences in the input datasets, as ‘adaptive’ or simply as ‘acceptable’. These behavioral sequences should be viewed as sequences whose execution is likely to increase the learner’s fitness in the classic sense, for instance by increasing its foraging success. Note also that the definition of a creative sequence requires that it had never been produced before by that individual, and thus it would seem that at best each creative sequence should be viewed as such in our simulation only on the first time that it is produced by the learner. This is not the case, because the learning mechanisms were disabled once the exposure of the learners to input sequences ended and before they started to produce sequences of their own. Produced sequences can thus be viewed as though they occurred immediate-

ly at the end of the learning from the input sequences, independent of one another.

The behavioral sequences produced by the learner depend on the model’s parameters that govern the production of behavioral sequences and, obviously, on what the learner had learned: the structure of the directed-graph representation that it had constructed. Through adjustment of the model’s parameters, existing explanations for creative and innovative behaviors can be accommodated by our computational framework. For example, attributing an organism’s reluctance to act creatively to factors such as neophobia, conservatism, or functional fixedness (e.g. Brosnan and Hopper, 2014; Overington et al., 2011) may be made more explicit by suggesting how these factors act as parameters in a computational model such as ours. Similarly, differences in these parameters’ values may shed light on why the tendency to innovate may be context-dependent, task-specific, or vary across individuals (e.g. Cole et al., 2011; Griffin et al., 2013; Griffin et al., 2013; Laland and Reader, 1999; Morand-Ferron et al., 2011).

To explore and to demonstrate the potential effects of the learning experience and of the model’s parameters on animal creativity, each of the following sections describes simulation results in which one or more of these factors were altered.

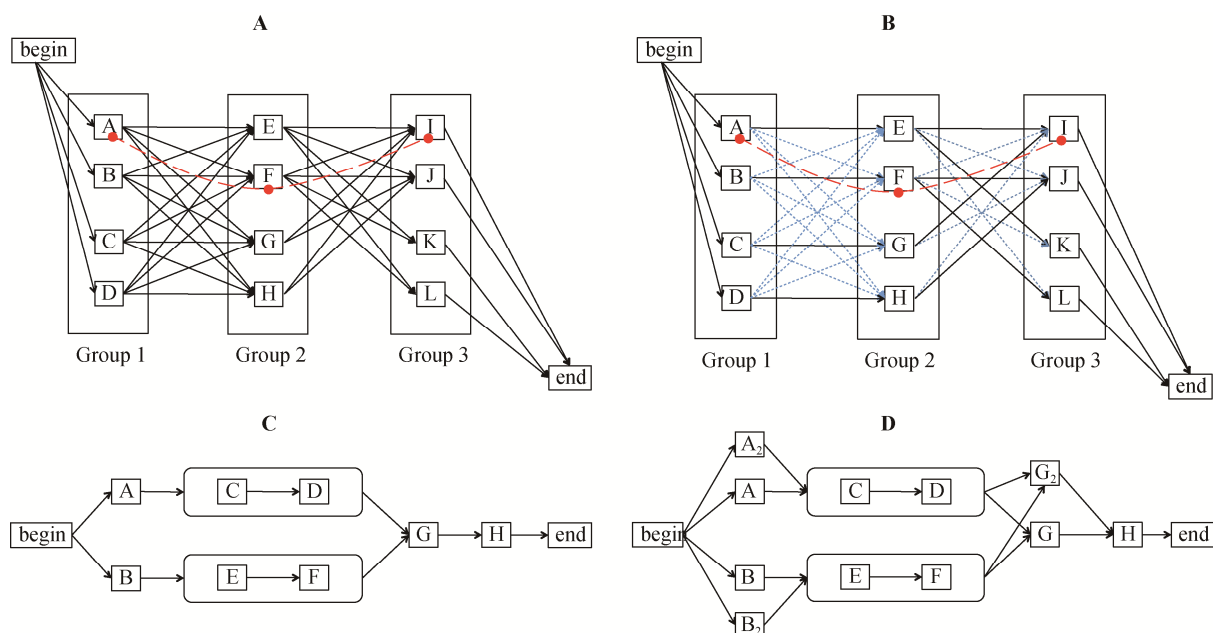


Fig. 2 The rules that we used to construct the training sets in the simulations, presented in the form of finite automata: each trajectory along directed edges from *begin* to *end* appeared in the simulations’ training phases with some probability

A. Automaton 1, used to produce training sets for simulations 1–3. **B.** Automaton 2, used to produce training sets for simulations 4–6, see more details in the main text. **C.** Automaton 3, used to produce the training sets for simulation 7. **D.** Automaton 4, used to produce the training sets for simulation 8. Note that the rules described by these automata do not produce all the acceptable (adaptive) sequences; see main text for further details and for an interpretation of each automaton’s regularities in the context of defining creative behavior.

4 Results and Discussion

4.1 Simulation 1: two routes may lead to creativity

In this simulation, adaptive behavioral sequences are defined as all (and only) sequences of the form “*begin {Group 1} {Group 2} {Group 3} end*”, in Fig. 2A. The sequences that comprise the 10,000-sequence training set are those produced by following trajectories along black (solid) edges in Fig. 2A, with the exception of the sequence “*begin A F I end*”, all instances of which were removed from the training set (depicted as a red, dashed line). The learner is exposed to the training set, and is required in the test phase to construct 100,000 sequences. We then evaluate how many of the produced sequences are adaptive and whether any of those are creative (i.e. both acceptable and novel).

This simple framework is useful to explore creativity, as it allows the learner to potentially construct a very wide range of behaviors, while setting clear constraints on what is deemed adaptive. Moreover, given the training set, it is straightforward to identify which potential sequences should be regarded as creative, i.e. both novel and adaptive. The possible creative sequences are therefore be: “*begin A F I end*” and the sequences of the form “*begin {Group 1} {G,H} {K,L} end*”.

We trained and tested two learners in this simulation; both use only the EBL mechanism, i.e. they construct a graph whose nodes are only the base-elements that they encounter while training (and no high-order units). In the first learner (L1), the generalization parameter (P_{gen}), which sets the extent to which similarity-based generalization is used in the production of sequences, was set to 0 (i.e. no generalization), while in learner 2 (L2) this parameter was set to a value of 0.2. This implies that during sequence production, each node that is initially chosen to take part in the sequence may be replaced by another node, at a probability of 0.2. The choice of a “similar” node to be used from all candidate nodes is proportional to its similarity index; see details in section 1 and 2 of the ESM.

All the 100,000 sequences produced by each of the learners were adaptive, i.e. of acceptable form in both length and structure. Nearly 1% of the sequences produced by L1 (that is incapable of generalization) were of the form *begin A F I end*, demonstrating that at least some creative behaviors do not require the ability to generalize per se (for full numerical results see Table 1 in the ESM). An examination of the simulation's logbook shows that these novel sequences were constructed simply by the learners' combining learned se-

quences with one another: *F* had been encountered following *A* and *I* had been encountered following *F*, so their concatenation was straightforward even though the three had never been encountered following one another as a full sequence. In fact, the simple representation constructed by L1 does not enable it to avoid the construction of such sequences, because no long-distance data beyond first-order transitions regarding the encountered sequences is recorded. Thus, one route to creativity that does not require generalization is a simple concatenation of learned sequences of two nodes (bigrams) into novel sequences of three nodes.

Further examination of the simulation results shows that no sequences of the novel form “*begin {group 1} {G,H} {K,L} end*” were produced by L1, while nearly 6.5% of the sequences produced by L2 were in this form. This result suggests a second route to creativity, which does require generalization. Indeed, an examination of the simulation's logbook revealed that L2 had inferred that *I* and *J* are similar to *K* and *L*, as all four have a large overlap in their edge profiles, and in its production process it occasionally replaced *I* or *J* by *K* or *L*.

4.2 Simulation 2: The effect of the generalization parameter on creative output

This simulation was identical to simulation 1, but in this case the learners that were trained and tested differed from one another by the value of their generalization parameter. This parameter determines the learner's tendency to try out new combinations based on its inference of similarity among learned units, and as expected we found that the higher the tendency to generalize, the higher the probability that novel sequences of the form “*begin {Group 1} {G,H} {K,L} end*” would be constructed (Fig. 3). Note that in this simulation, generalization never led to the construction of maladaptive sequences because of the extremely simple structure of the learned modality. In other conditions, including most realistic scenarios, it is quite likely that over-generalization would frequently lead to the construction of maladaptive sequences (see examples in simulations 7 and 8).

4.3 Simulation 3: A third (but risky) route to creative behavior through random choice of nodes in sequence production

Trial and error of completely random behavioral sequences may lead to sequences that answer the criteria for creative behavior. It is expected that such activity would also give rise to many non-adaptive sequences, some of which might even be detrimental, setting a realistic constraint on such behavior. Nevertheless, it seems

that this is the gist of the strategy chosen by some animals in certain contexts (e.g. in the exploration of novel types of food, in the interaction with unfamiliar objects in the environment, or in young individuals' attempts to communicate with adult individuals). To model this, we repeated the previous simulations with a learner whose generalization parameter was set to 0, but with a different parameter, P_{random} , set to 0.1. In production, with a probability of 0.1 in each step of constructing the sequence, this learner inserts to the sequence a randomly chosen node from its internal representation. This is done regardless of whether or not the unit represented by that node had been encountered previously in the newly constructed context.

Not surprisingly, more than a quarter of the sentences that this learner produced were not acceptable. At the same time, 0.7% and 0.3% of the produced sequences were, respectively, of the creative forms “*begin A F I end*” and “*begin {Group 1} {G,H} {K,L} end*” (see table 1 of the ESM). This suggests that in contexts in which the cost of ill-formed sequences is low while the potential advantage of uncovering novel adaptive sequences is high, some moderate level of random trial and error in the construction of sequences may be adaptive and may circumvent the need for informed generalization, which may have a high cognitive cost.

4.4 Simulation 4: The risk of entrenchment

This simulation was similar to the previous ones, but the automaton used to generate the training set was slightly different (Fig. 2B, henceforth 'automaton 2'). In this automaton, most edges are assigned very low probability, and thus occur rarely in the training set (dotted

edges in Fig. 2B), while few edges occur very frequently (solid edges in Fig. 2B). In constructing the training set, 95% of the steps are along the solid edges. The length of the training remained as before, and the number of encounters with each element was maintained similar to the previous experiments. Despite these, the frequent recurrence of certain edges and not others led to entrenchment of certain sequences in the learner's representation, which significantly lowered the probability of construction of some creative sequences (see below). We again trained and tested learners L1 and L2, which differ from one another in that L2 uses similarity-based generalization.

As before (and as expected), L1 produced no sequences of the form “*begin {Group 1} {G,H} {K,L} end*”, but as opposed to its behavior following training with sequences derived from automaton 1 (Fig. 2A), in which nearly 1% of the produced sequences were “*begin A F I end*”, in this simulation only 0.01% of the sequences were of this form (this and all other differences reported as significant refer to two-tailed chi-squared tests with $df = 1$ and $p < 0.001$). Learner L2 produced slightly more sequences of this form (0.08%), but also showed a great reduction compared to its production of such sequences previously (1.1%). Yet, L2 production of sequences of the form “*begin {Group 1} {G,H} {K,L} end*” did not change significantly and certainly did not decline in comparison to simulation 2 (6.6% versus 6.5%, see table 1 in the ESM).

Analysis of these results and of the process that brought them about offers a few insights. The creative sequences whose production was significantly reduced

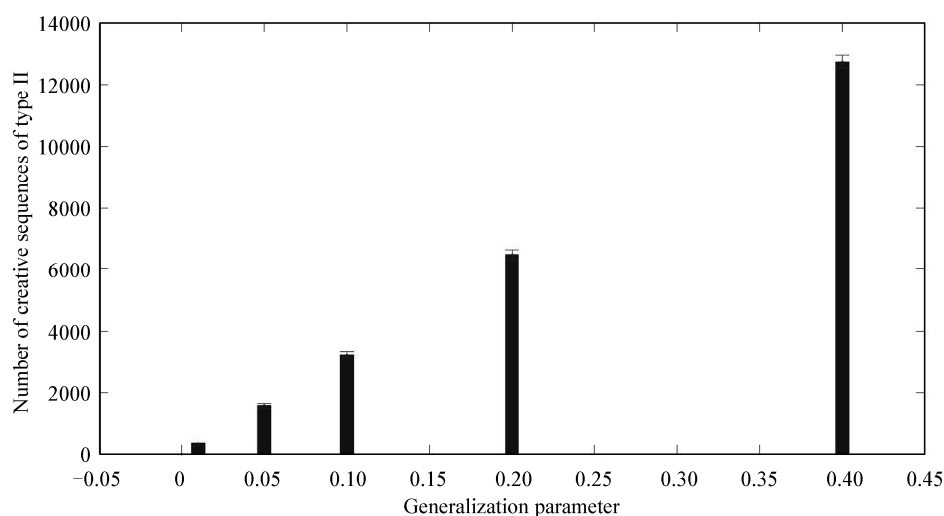


Fig. 3 The number of creative sequences produced by EBL learners which differ by their values of the generalization parameter (simulation 2)

Each learner produced an output of 100,000 sequences. Error bars depict \pm two standard deviations derived from a bootstrap analysis. See ESM for details.

are of the form that can be constructed without need for similarity-based generalization (analogy). This is a result of entrenchment of certain sequences: a tendency to follow only particular trajectories along the learners' internal representations. The vast experience that the learners had accumulated in encountering the sequences *A E*, *F J*, and *F L* lowered the probability that *A F* and *F I* would be produced, lowering the production probability of *A F I* to near zero. We find that L2's tendency to use generalizations at a certain probability creates a "rescue" effect: L2 also showed a significant reduction in the production of these sequences, but it was alleviated somewhat (a near 10-fold reduction, as opposed to a near 100-fold reduction in L1) by the construction of these sequences via generalization: although F occurred rarely following A and I rarely followed F, L2 could infer that qualitatively their edge profiles were similar to those of other elements in their groups, and it generalized accordingly. These findings suggest that rich experience, in which the various elements occur in a wide range of contexts that reflect the range of possible roles that each of them may assume, may be critical for the production of creative behavior through similarity-based generalization.

The production of creative sequences that are dependent on generalization was not altered by the change in the training set that induced entrenchment. This is due to a combination of two factors. First, even rare occurrences affect the assessment of similarity among nodes, and may do so strongly (see the calculation of the similarity index in the ESM). Second, the implementation of the learners that was used is such that the rate of using generalizations depends on the generalization parameter and is insensitive to the similarity index: the level of similarity may influence which node will be chosen as a replacer but not whether a replacement will be executed. Jointly these factors lead to the maintenance of a high production level of novel sequences of the form "*begin {Group 1} {G,H} {K,L} end*" also under conditions that induce entrenchment.

Note that there are other, quite reasonable, ways to model the use of generalization. One scheme that comes to mind is a model in which the extent of generalization depends on the learner's assessment of the probability that the generalization would be adaptive, which is determined by the level of similarity between units: the higher the similarity, the higher the perceived probability of a generalization's success. To find out whether in such a scheme the frequency of creative sequences would be reduced under conditions that induce entrenchment,

we implemented such a learner and trained it on the training set produced from automaton 2. As expected, the frequency of creative sequences of both forms was lower than in L2 (0.025% and 4.5%, respectively; see Table 1 in the ESM under "altered generalization", and see the details of the generalization process in the ESM).

4.5 Simulation 5: Production strategies that can bypass the entrenchment problem and increase creativity

Entrenchment of certain combinations may have advantages (as will be discussed in the next section), but as shown above, it may impair creativity. This phenomenon was demonstrated most clearly with regard to the construction of the sequence "*begin A F I end*" by L1 following training on the automaton 2 training set. To examine possible production strategies on the part of the learner that might overcome this challenge, we implemented two learners similar to L1, but whose preferences regarding which edges to follow in the internal representation during construction of new sequences were altered: one learner (L3) ignores edge weights, and the other (L4) prefers to make use of rarely encountered sequences. As expected, the frequencies of instances of the sequence "*begin A F I end*" constructed by the two learners were much higher than in the previous results: 1.5% and 3.9% respectively (see table 1 of the ESM). The potential risk of using these strategies is that rare occurrences may be less reliable or less frequently adaptive in common conditions. This implies that such strategies are likely to evolve only when the benefit from producing novel adaptive sequences greatly exceeds the cost of producing maladaptive ones.

4.6 Simulation 6: High-order units may impair creativity via increased entrenchment

Fig. 1 illustrates an example of a potential entrenchment following the inclusion of high-order units in the repertoire: had the rat never encountered nuts while foraging among branches in the trees (i.e. the weight of the edge among these units were zero), it is likely that the sequence *forage among branches => find cone* would have become encapsulated as a single unit, as illustrated in purple (dash-dot line). This would lead to a situation in which foraging in the trees is associated solely with search for cones, and the rat would have been likely not to expect other food sources that may be found there. This potential pitfall is demonstrated explicitly in the following simulation.

So far the learners in the simulations we described used only the EBL mechanism to add units to their re-

pertoire. To explore the effects of the representation of high-order units in the repertoire, we conducted a simulation in which four learners that implement the RBL mechanism (recurrence-based learning, see section 2.1) were trained on sequence sets derived from automata 1 and 2. For simplicity, all explorations of the effects of high-order chunks on creativity that are presented in this paper were conducted using RBL, rather than bottom-up chunking (see section 2.1) or a combination of the two.

As discussed above, it may be reasonable for an organism to prefer the use of encapsulated sequences of behavior (i.e. chunks) over the construction of sequences from basic elements. The extent to which the learner gives precedence to longer units in its construction of sequences is determined by a parameter, given values of 0, 1, 2, 3 respectively, that differed among the learners in this simulation.

The generalization parameter in all learners was set to zero, and so none of them produced sequences of the form "*begin {Group 1} {G,H} {K,L} end*". All learners produced some instances of the sequence "*begin A F I end*" when trained on the dataset derived from automaton 1, but all produced significantly fewer such sequences than the learners in previous simulations, with a clear decreasing trend among them: the higher the preference for longer units, the fewer creative sequences were produced (the frequencies of the creative sequence in the learners' outputs were, respectively, 0.15%, 0.05%, 0.02%, 0.004%). This result was pronounced when trained on the dataset from automaton 2: the frequency of *A F I* sequences in the output was between 0.002% and 0% in all four learners.

These results are easy to explain: the only acceptable 3-element sequence that does not occur in the training is *A F I*; all others recur in it, and the training is long enough for most or all triplets to recur within a short time-window and thus be recognized as units by the learner. In production, even when no particular preference is assigned to long units, many of the composed sequences are simply instances of these 3-element units, and thus *A F I* has an extremely decreased representation in the output. This trend is even more extreme following training on automaton 2 data, because in it the sequence *A F* is rare and unlikely to be recognized as a unit by the learners; thus alongside sequences that are instances of 3-element units and do not include *A F I*, sequences that initiate with a 2-element unit also never result in a sequence that initiates with *A F*.

This demonstrates that one major effect of constructing high-order units may be to decrease creativity by

boosting the learners' tendency to be conservative and by intensifying the effect of entrenchment (see also Lotem and Halpern 2008 for possible causes and implications of this effect in impairing generalization and creativity in autism).

4.7 Simulation 7: Can high-order units contribute to the production of creative behavior?

As noted earlier, high-order units (chunks) may be useful as a realization of action-plans, i.e. a sequence of behaviors that it is worthwhile to encapsulate as a unit. In other cases the high-order units may simply play the role of units whose breaking down into their individual constituents strips them of their meaning. This suggests that although learning of high-order units may lead to entrenchment as shown above, there may be cases in which it is essential in order to bring about creative behavior.

To further explore the effects of chunking, we constructed a novel training set based on the simple automaton described in Fig. 2C (automaton 3). Using this dataset we trained and tested L1, L2 and three RBL learners whose generalization parameters were set to 0, 0.1 and 0.2 respectively. In this automaton the only equivalence groups among units are $\{A, B\}$ and $\{C D, E F\}$, and thus creative behaviors would be the construction of the novel sequences "*begin A E F G H end*" or "*begin B C D G H end*". These can be brought about only by means of drawing the analogy between the units *C D* and *E F* and not by constructing a new combination of previously encountered sequences.

On reviewing the simulation results we found that none of the five learners constructed either of the creative sequences. An inspection of the simulation log-books shows that the reasons vary among learners, L1 and the first RBL learner do not generalize, and their failure to act creatively in this setting is expected. L2 does not represent high-order units, and thus failed to learn that *C D* and *E F* are units whose constituents are non-separable, demonstrating that without the ability to learn high-order chunks, creativity may be severely limited in some settings. An inspection of the internal representations of the latter RBL learners showed the reverse obstruction: they had learned the sequences *A C D G H* and *B E F G H* as whole units, and did not recognize that these sequences constitute whole action plans and not non-separable units, i.e. they too did not learn that *C D* and *E F* are independent units and that they are similar to one another. This is a second demonstration of the hazards of entrenchment, which could have been overcome only by a richer experience with

each of the units that compose this modality. This is demonstrated in the following simulation.

4.8 Simulation 8: How variation sets that facilitate correct segmentation contribute to the production of creative behavior based on high-order units

In simulation 7 we have seen that a failure to segment the units $\langle C D \rangle$ and $\langle E F \rangle$ from the longer sequences $\langle A C D G H \rangle$ and $\langle B E F G H \rangle$ precluded the learner's ability to recognize their similarity and to use them for generalization. This connection between segmentation problems and generalization failures has been suggested in the past (Lotem and Halpern, 2008), as well as the need to provide appropriate distribution of data input in order to allow proper segmentation based on alignment and comparison of data strings (Goldstein et al., 2010; Lotem and Halpern, 2012; Onnis et al., 2008a). In fact, as described earlier, the design of our RBL learner was inspired by this idea of identifying units that recur within a limited time window as an efficient and reliable method for data segmentation (see Introduction and Kolodny et al., 2014b). The recurrence of units in multiple contexts, like variants on a theme, within a short time window is known as a 'variation set', and is common in child-directed speech (Waterfall, 2006; Brodsky, et al., 2007), and instrumental in improving adults' learning of linguistic units (Onnis et al., 2008b). Thus, by providing additional data in our simulation, in the form of variation sets, we predicted that the segmentation of $\langle C D \rangle$ and $\langle E F \rangle$ will become possible, which may then allow the learner to recognize the similarity between them and to creatively generate useful novel sequences that are based, in this case, on high-order units.

To that end we constructed training sets based on automaton 4 (see Fig. 2D). In this automaton the equivalence groups (groups containing similar units) are $\{A, A_2, B, B_2\}$, $\{C D, E F\}$ and $\{G, G_2\}$, and creative sequences are the sequences of the forms "*begin A/A₂ E F G/G₂ H end*" and "*begin B/B₂ C D G/G₂ H end*". Note that now segmentation of $\langle C D \rangle$ and $\langle E F \rangle$ may be possible by their recurrence in sequences such as "*A₂ C D G*" and "*A C D G₂*" (or "*B E F G*" and "*B₂ E F G₂*").

The results of this simulation show that as opposed to simulation 7, the RBL learners whose generalization parameters were greater than zero succeeded in constructing creative sequences (at a frequency of 0.2% and 0.4% of their overall produced sequences). This was possible because in this simulation these learners learned $C D$ and $E F$ as units whose similarity to one

another is high. The crucial difference from the previous dataset lies in that automaton 4 leads to the construction of different sequences in which each of these two units occurs in multiple contexts, allowing the learners to correctly segment the entire sequences through alignment and comparison during the search for recurring units. In addition to allowing for improved segmentation and correct recognition of the modality's underlying units, these multiple contexts lead to the building up of a rich representation of each unit's edge profiles, increasing the learners' ability to infer which units are similar to one another.

5 General Discussion

To help us understand the mechanisms underlying animal creativity we suggested in this paper to make use of a computational cognitive framework that we previously developed for analyzing and reproducing a range of behavioral phenomena, from animal foraging to the learning of human language (Kolodny et al., 2014a, b). Common to these phenomena is the need to learn regularities in the environment, to represent them hierarchically, and to use this representation for the production of behavioral sequences. Animal creativity is viewed in this framework as the production of novel behavioral sequences that are likely to be adaptive.

This framework differs from most existing computational models in that it was designed with an eye towards biological realism in a number of respects, such as the use of limited computation and memory resources, the exclusive reliance on incremental (as opposed to batch) learning, the application of domain-general mechanisms, and the attention to the possible evolutionary factors that may have led to its development (Kolodny et al., 2014a; and see further details and discussion in Kolodny, Lotem, and Edelman, 2014). It is particularly suitable for the modeling of animal behavior due to its simplicity and generality, lending itself easily to the modeling of a wide range of behaviors, including novel behaviors whose definition as creative is independent of internal cognitive processes that are not accessible to the observer.

Our simulations, conducted within this framework, highlighted a few different mechanisms that can produce creative behaviors and elucidated some constraints and intricacies. First (in simulation 1), we demonstrated two main routes to creativity: novel concatenation of previously learned units and similarity-based generalization (also referred to as the use of analogies). We then showed how creativity may be enhanced by increasing

the propensity to generalize (simulation 2), or, in the absence of generalization, by simply introducing some random choices into the process of sequence production (simulation 3). Note that these two mechanisms correspond to different types of creativity. In the first one, novelty is canalized to similar objects or actions and is therefore likely to be context-appropriate and adaptive. In the second mechanism, animals may 'gamble' on the next step, which may be successful in only a small proportion of the cases and may therefore be adaptive only when the cost of errors is low, as in the context of exploratory or play behavior (Bateson and Martin, 2013; Bekoff and Byers, 1998; Burghardt, 2005; Heinrich and Smolker, 1998).

A general important constraint on creativity is demonstrated in our model by the problem of entrenchment (simulations 4 and 5). While the ecological costs and benefits of creative or innovative behaviors have been discussed in some depth (e.g. Kummer and Goodall, 2003; Lee, 2003; Reader and Laland, 2003), the problem of entrenchment deals with the tradeoff arising from the working principles of the cognitive mechanism that produces the creative sequences, as described throughout this paper: on the one hand, following the weights in the network that represent the statistical frequencies of sequences in nature is necessary for producing the most probable sequences that are likely to be adaptive (Goldstein et al., 2010, Lotem and Halpern, 2012; Kolodny et al., 2014a, b). On the other hand, relying on the most frequent sequences necessarily causes entrenchment and prevents creativity.

Our model offered three partial solutions to this problem. First, and perhaps most important, it shows that similarity-based generalization may partly overcome entrenchment by generalizing from common to rare nodes (or from rare to common nodes) while still adhering to the use of weights (see simulation 4). Second, the risk of overgeneralization may be minimized by a mechanism that adjusts the propensity to generalize according to the level of similarity between nodes (the second part of simulation 4). Finally, in simulation 5 we showed that an alternative solution to the entrenchment problem (that does not involve generalization) is simply to ignore the weights in the network, or to invert their values and choose the rare instead of the common options. As discussed earlier, this route to creativity is risky but may occasionally be adaptive. More generally, the extent to which the production of sequences is determined by the weights in the network illustrates a spectrum of strategies, which range from

conservatism and rigidity on the one side to creativity on the other.

The last three simulations of our paper (6 to 8) addressed the more realistic and cognitively advanced form of our framework that accommodates high-order units (chunks or hierarchies). Our results show that the effect of high-order units on creativity may be complex. On the one hand, creating long chunks can limit creativity because it does not allow novel concatenations of nodes that are 'locked' within the chunks, which intensifies the effect of entrenchment (see simulation 6). On the other hand, high-order units are critical for representing real high-order entities in the environment, and when these are represented correctly in the network, generalization and creativity may be most powerful (see simulation 8). Yet, in order to correctly represent high-order units in the network, we have seen that the segmentation process must work properly and this was affected by the amount of experience and the availability of data strings with partly shared sequences (variation sets). This result suggests that advanced creative behaviors, such as those required for sequential problem solving or tool use (see, e.g., Bentley-Condit and Smith, 2010; Wimpenny et al., 2009) cannot be generated solely by the simple mechanisms suggested earlier, but must also be supported by extensive experience that allows the animal to construct a rich network of high-order units, across which generalizations can be made. Indeed, it seems that most evidence for highly creative behaviors comes from species where exploratory and play behavior is extensive (see, e.g., Diamond and Bond, 1999; Heinrich and Smolker, 1998; Pellegrini and Smith, 2005).

The ability to generalize and draw analogies between learned units, which we have found to be crucial to the construction of certain creative behaviors, plays a central role in models of human creativity. The model in (Hofstadter and Mitchell, 1995), for example, deals with this concept elaborately; Hofstadter and Mitchell suggest that the "depth" of a notion, i.e. its degree of abstraction, is related to the degree of surprise in the creative outcome that is based on an analogy to that notion. This "depth" of notion may find its counterpart in our framework in the extent of similarity among units that are used in an analogy, and perhaps more important, in the level of hierarchies being compared and generalized. For example, a young Arabian babbler that learns to turn over small branches on the ground in order to find food may generalize this behavior to little stones or leaves that may look very different from branches. This

creative behavior may be explained in a straightforward manner by our framework, as branches, stones and leaves are simple units that are likely to occur in similar contexts and thus have a high index of similarity among them, all together suggesting that the depth of notion required for such an analogy is not great. On the other hand, a New Caledonian crow that upon being faced with a challenge made a hook from a metal wire and used it to pull out a food item from a tube (Weir et al., 2002) seems to be using a more sophisticated process of sequence production. This process may be based on multiple analogies that are based on high-order units (sequences of actions) and perhaps on employing only liberal constraints on the extent of similarity that units must exhibit to be allowed to replace one another in behavioral sequences.

While excluded from our modeling here, the element of surprise or insight is a recurring theme in the discussion of human creativity, which we find interesting (see, e.g., Boden, 2004). The intuitive notion of creativity is bound to the feeling of surprise on encountering or coming up ourselves with a certain new behavior that is appropriate in the given context, and this is true for many animal behaviors that have been described as creative. The process of constructing a creative behavioral sequence can be viewed from a computational perspective, as noted before, as a search in the vast possible space of sequential behaviors for one that would be adaptive in a certain context. This search may produce multiple candidate sequences, among which only one or very few can eventually be carried out. This requires an assessment and decision mechanism, perhaps one that involves an internal simulation to infer the possible outcomes of each sequence (an example for such a choice mechanism is implemented in our framework in (Kolodny et al., 2014)). We suggest that the feeling of 'insight' (as used by Boden, 2004) or surprise accompanies the cases in which this process yields a sequence whose initial probability of being constructed was low or whose initial assessment in the choice procedure was not promising, but which was eventually chosen and turned out to be successful. Similarly, it might be elicited by a sequence that contains a far-reaching analogy that had not been used before. While the "feeling" of insight or surprise is currently not being produced by our model, the circumstances under which it may be elicited can be fully specified in our current framework, as in the version developed in (Kolodny et al., 2014a) that includes a choice procedure based on the learner's assessment of an option's likelihood of success. The

adaptive function of the feeling of insight or surprise deserves further discussion. We suggest briefly that like other emotional states, it may be related to drawing attention and assigning more weight in memory to rare but important discoveries (see, e.g., Lotem and Halpern, 2012; Schmidhuber, 2010; 2013).

Notably, some specific suggestions regarding the nature of the process that gives rise to creative behaviors in humans are particularly in line with our approach. Thus, Bowden et al. (2005) suggest that many creative solutions to problems follow a re-phrasing of the problem. Similarly, the literature review offered by Edelman (Edelman, 2008) suggests that creative behavior can be characterized as the outcome of running into an impasse in trying to cope with a situation in a certain way of thought, and the need to backtrack one's train of thought and try a different approach. In terms of our model, such re-phrasing or trying different approaches may initially be analogous to multiple searches along different paths of nodes and links in the network, but then may include repeated searches after modifying the production strategies or the generalization parameters in our model, or may even involve some re-segmentation of the data in different ways.

Finally, we would like to stress that the computational framework we presented here to study animal creativity was kept simple in order to make it generally applicable and suitable for the study of a wide range of phenomena in cognition and behavior. It was constructed in a modular and flexible way so as to allow the accommodation and integration of more detailed models of particular aspects of the cognitive process. We hope that future research will both benefit from the generality of the present framework and realize its potential for elaboration. The software used in this study is available upon request.

Acknowledgements We would like to thank Corina Logan and an anonymous reviewer for their comments, which helped improve this manuscript. OK was partially supported by a Dean's scholarship from the Faculty of Life Sciences at Tel-Aviv University and by a Wolf Foundation award. AL and OK were partially supported by the Israel Science Foundation grant no. 1312/11.

References

- Aisner R, Terkel J, 1992. Ontogeny of pine-cone opening behavior in the black rat *Rattus rattus*. *Animal Behaviour* 44 (2): 327–336.
- Allen J, Weinrich M, Hoppitt W, Rendell L, 2013. Network-based diffusion analysis reveals cultural transmission of lobtail

- feeding in humpback whales. *Science* 340 (6131): 485–488.
- Aplin LM, Farine DR, Morand-Ferron J, Cockburn A, Thornton A et al., 2014. Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*. June 23; Doi: 10.1038/nature13998.
- Bateson PPG, Martin P, 2013. *Play, Playfulness, Creativity and Innovation*. New York: Cambridge University Press.
- Bekoff M, Byers JA, 1998. *Animal Play: Evolutionary, Comparative, and Ecological Perspectives*. Cambridge: Cambridge University Press.
- Bentley-Condit VK, Smith EO, 2010. Animal tool use: Current definitions and an updated comprehensive catalog. *Behaviour* 147 (2): 185–A132.
- Bluff LA, Weir AA, Rutz C, Wimpenny JH, Kacelnik A, 2007. Tool-related cognition in new caledonian crows. *Comparative Cognition and Behavior Reviews* 2: 1–25.
- Boden M, 1995. Creativity and unpredictability. *Stanford Humanities Review* 4: 123–139.
- Boden MA, 2004. *The Creative Mind: Myths and Mechanisms*. 2nd edn. London, New York: Routledge.
- Boinski S, 1988. Use of a club by a wild white-faced capuchin *Cebus capucinus* to attack a venomous snake *Bothrops Asper*. *American Journal of Primatology* 14 (2): 177–179.
- Bowden EM, Jung-Beeman M, Fleck J, Kounios J, 2005. New approaches to demystifying insight. *Trends in Cognitive Sciences* 9: 322–328.
- Brodsky P, Waterfall HR, Edelman S, 2007. Characterizing Motherese: On the computational structure of child-directed language. In: McNamara DS, Trafton JG ed. *Proceedings of the 29th Cognitive Science Society Conference*. Austin, TX: Cognitive Science Society, 833–838.
- Brosnan SF, Hopper LM, 2014. Psychological limits on animal innovation. *Animal Behaviour* 92: 325–332.
- Burghardt GM, 2005. *The Genesis Of Animal Play: Testing the Limits*. Cambridge, London: MIT.
- Bush RR, Mosteller F, 1951. A mathematical model for simple learning. *Psychol. Rev.* 58 (5): 313–323.
- Byrne RW, 2002. Imitation of novel complex actions: What does the evidence from animals mean? *Advances in the Study of Behavior* 31: 77–195.
- Byrne RW, 1999. Imitation without intentionality: Using string parsing to copy the organization of behaviour. *Animal Cognition* 2: 63–72.
- Byrne RW, Russon, AE, 1998. Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences* 21 (5): 667–721.
- Cole EF, Cram DL, Quinn JL, 2011. Individual variation in spontaneous problem-solving performance among wild great tits. *Animal Behaviour* 81 (2): 491–498.
- Conant RC, Ashby WR, 1970. Every good regulator of a system must be a model of that system. *International Journal of Systems Science* 1 (2): 89–97.
- Diamond J, Bond AB, 1999. *Kea, Bird of Paradox: The Evolution and Behavior of A New Zealand Parrot*. Berkeley: University of California Press.
- Edelman S, 2008. *Computing the Mind: How The Mind Really Works*. New York: Oxford University Press.
- Fawcett TW, Hamblin S, Giraldeau LA, 2013. Exposing the behavioral gambit: The evolution of learning and decision rules. *Behavioral Ecology* 24 (1): 2–11.
- Ferster CB, Boren MCP, 1968. *Behavior Principles*. New York: Appleton-Century-Crofts.
- Fisher J, Hinde RA, 1949. The opening of milk bottles by birds. *British Birds* 42: 347–357.
- Fleck JI, Weisberg RW, 2004. The use of verbal protocols as data: An analysis of insight in the candle problem. *Memory and Cognition* 32 (6): 990–1006.
- Forth J, Wiggins GA, McLean A, 2010. Unifying conceptual spaces: Concept formation in musical creative systems. *Minds and Machines* 20 (4): 503–532.
- Giraldeau LA, 1997. The ecology of information use. In: Krebs JR, Davies NB ed. *Behavioural Ecology: An Evolutionary Approach*. Oxford: Blackwell Science.
- Goldstein MH, Waterfall HR, Lotem A, Halpern J, Schwade J et al., 2010. General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences* 14: 249–258.
- Goodall J, 1986. *The chimpanzees of Gombe: Patterns of behavior*. Cambridge: Belknap Press of Harvard University Press.
- Griffin AS, Diquelou M, Perea M, 2014. Innovative problem solving in birds: A key role of motor diversity. *Animal Behaviour* 92: 221–227.
- Griffin AS, Guez D, Lermite F, Patience M, 2013. Tracking changing environments: Innovators are fast, but not flexible learners. *PLoS ONE* 8 (12).
- Griffin AS, Lermite F, Perea M, Guez D, 2013. To innovate or not: Contrasting effects of social groupings on safe and risky foraging in Indian mynahs. *Animal Behaviour* 86 (6): 1291–1300.
- Harel D, 1987. Statecharts: A visual formalism for complex-systems. *Science of Computer Programming* 8 (3): 231–274.
- Harel D, 2007. Statecharts in the making: A personal account. In: HOPL III: Proceedings of the Third ACM SIGPLAN Conference on History of Programming Languages. New York: ACM, (5–1–5–43).
- Harley CB, 1981. Learning the evolutionarily stable strategy. *Journal of Theoretical Biology* 89 (4): 611–633.
- Heinrich B, Smolker R, 1998. Play in common ravens *Corvus corax*. In: Beckoff M, Byers JA ed. *Animal Play: Evolutionary, Comparative, and Ecological Perspectives*. Cambridge: Cambridge University Press.
- Heyes C, 2012. Simple minds: A qualified defense of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1603): 2695–2703.
- Hofstadter D, 1995. *Fluid Concepts and Creative Analogies*. London: Harvester Wheatsheaf.
- Hofstadter DR, Mitchell M, 1994. The Copycat project: A model of mental fluidity and analogy-making. *Advances in Connectionist and Neural Computation Theory* 2: 31–112.
- Hoppitt W, Laland KN, 2013. *Social Learning: An Introduction to Mechanisms, Methods, and Models*. Princeton, NJ: Princeton University Press.
- Hosey GR, Jacques M, Pitts A, 1997. Drinking from tails: Social learning of a novel behaviour in a group of ring-tailed lemurs *Lemur catta*. *Primates* 38 (4): 415–422.
- Johnson-Laird PN, 1991. Jazz improvisation: A theory at the computational level. In: Howell P, West R, Cross I ed. *Representing Musical Structure*. London: Academic Press, 291–325.
- Kaufman JC, Kaufman AB, 2004. Applying a creativity framework to animal cognition. *New Ideas in Psychology* 22 (2):

- 143–155.
- Kawai M, 1965. Newly-acquired pre-cultural behavior of the natural troop of Japanese monkeys in Koshima islet. *Primates* 6 (1): 1–30.
- Keynan O, Ridley A, Lotem A, 2015. Social foraging strategies and acquisition of novel foraging skills in cooperatively breeding Arabian babblers. *Behavioral Ecology* 26 (1): 207–214.
- Kolodny O, Edelman S, The problem of concurrent multimodal serial order in behavior. submitted.
- Kolodny O, Edelman S, Lotem A. Evolution of proto-linguistic abilities as a by-product of learning to forage in structured environments. Under review.
- Kolodny O, Edelman S, Lotem A, 2014a. The evolution of continuous learning of the structure of the environment. *Journal of the Royal Society Interface* 11 (92): 20131091.
- Kolodny O, Lotem A, Edelman S, 2014b. Learning a generative probabilistic grammar of experience: A process-level model of language acquisition. *Cogn Sci.* 30 June; DOI: 10.1111/cogs.12140.
- Kounios J, Beeman M, 2009. The Aha! Moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science* 18 (4): 210–216.
- Kummer H, Goodall J, 2003. Conditions of innovative behaviour in primates. In: Reader SM, Laland KN ed. *Animal Innovation*. New York: Oxford University Press.
- Laland KN, Reader SM, 1999. Foraging innovation is inversely related to competitive ability in male but not in female guppies. *Behavioral Ecology* 10 (3): 270–274.
- Lee PC, 2003. Innovation as a behavioural response to environmental challenges: A cost and benefit approach. In: Reader SM, Laland KN ed. *Animal Innovation*. New York: Oxford University Press.
- Lefebvre L, Bolhuis JJ, 2003. Positive and negative correlates of feeding innovation in birds: Evidence for limited modularity. In: Reader SM, Laland KN ed. *Animal Innovation*. New York: Oxford University Press.
- Lieven E, Tomasello M, Behrens H, Speares J, 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30: 333–370.
- Logan CJ, Pepper JW, 2007. Social learning is central to innovation, in primates and beyond. *Behavioral and Brain Sciences* 30 (4): 416–417.
- Lotem A, Halpern, J, 2008. A Data-Acquisition Model for Learning and Cognitive Development and Its Implications for Autism (Computing and Information Science Technical Reports). New Heaven : Cornell University.
- Lotem A, Halpern JY, 2012. Coevolution of learning and data-acquisition mechanisms: A model for cognitive evolution. *Philosophical Transactions of the Royal Society B-Biological Sciences* 367 (1603): 2686–2694.
- McNamara JM, Houston AI, 1987. Memory and the efficient use of information. *Journal of Theoretical Biology* 125 (4): 385–395.
- McNamara JM, Houston AI, 2009. Integrating function and mechanism. *Trends in Ecology and Evolution* 24 (12): 670–675.
- Menyhart O, Kolodny O, Goldstein MH, DeVoogd TJ, Edelman S. Like father, like son: Zebra finches learn structural regularities in their tutors' song. Under review.
- Millenson JR, 1967. *Principles of Behavioral Analysis*. New York: Macmillan.
- Morand-Ferron J, Cole EF, Rawles JEC, Quinn JL, 2011. Who are the innovators? A field experiment with 2 passerine species. *Behavioral Ecology* 22 (6): 1241–1248.
- Onnis L, Waterfall HR, Edelman S, 2008a. Learn locally, act globally: Learning language from variation set cues. *Cognition* 109: 423–430.
- Onnis L, Waterfall HR, Edelman S, 2008b. Variation sets Facilitate Artificial Language Learning. Proc. 30th Cognitive Science Society Conference. Washington, DC.
- Overington SE, Cauchard L, Cote KA, Lefebvre L, 2011. Innovative foraging behaviour in birds: What characterizes an innovator? *Behavioural Processes* 87 (3): 274–285.
- Pellegrini AD, Smith PK, 2005. *The Nature of Play: Great Apes and Humans*. New York: Guilford Press.
- Ramsey G, Bastian ML, van Schaik C, 2007. Animal innovation defined and operationalized. *Behavioral and Brain Sciences* 30 (4): 393–437.
- Reader SM, Laland KN, 2003. *Animal Innovation*. Oxford: Oxford University Press.
- Reader SM, Laland KN, 2003. Animal innovation: An introduction. In: Reader SM, Laland KN ed. *Animal Innovation*. New York: Oxford University Press.
- Reader SM, MacDonald K, 2003. Environmental variability and primate behavioural flexibility. In Reader SM, Laland KN ed. *Animal Innovation*. New York: Oxford University Press.
- Savage-Rumbaugh S, Lewin R, 1994. *Kanzi: The Ape at the Brink of the Human Mind*. New York: Wiley.
- Schmidhuber J, 2006. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science* 18 (2): 173–187.
- Schmidhuber J, 2008. Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In: Pezzulo G, Butz MV, Sigaud O, Baldassarre G ed. *Anticipatory Behavior in Adaptive Learning Systems*. Berlin-Heidelberg: Springer, 5499: 48–76.
- Schmidhuber J, 2010. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development* 2 (3): 230–247.
- Schmidhuber J, 2013. Maximizing fun by creating data with reducible subjective complexity. In: Baldassarre G, Mirolli M ed. *Intrinsically Motivated Learning in Natural and Artificial Systems*. Berlin-Heidelberg Springer, 95–128.
- Seed A, Byrne R, 2010. Animal tool-use. *Curr. Biol.* 20 (23): R1032–1039.
- Seed AM, Boogert NJ, 2013. Animal cognition: An end to insight? *Current Biology* 23 (2): R67–R69.
- Shettleworth SJ, 2010. Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences* 14 (11): 477–481.
- Shettleworth SJ, 2012. Do animals have insight, and what is insight anyway? *Can. J. Exp. Psychol.* 66 (4): 217–226.
- Simonton DK, 2000. Creativity, cognitive, personal, developmental, and social aspects. *Am. Psychol.* 55 (1): 151–158.
- Skinner BF, 1966. *The Behavior of Organisms*. New York: Appleton-Century-Crofts.
- Sol D, 2003. Behavioural innovation: A neglected issue in the ecological and evolutionary literature? In: Reader SM, Laland KN ed. *Animal Innovation*. New York: Oxford University Press.

- Sternberg RJ, Davidson JE, 1995. *The Nature of Insight*. Cambridge, Mass.: MIT Press.
- Thornton A, Samson J, 2012. Innovative problem solving in wild meerkats. *Animal Behavior* 83 (6): 1459–1468.
- Thorpe WH, 1956. *Learning and Instinct in Animals*. London: Methuen.
- Trimmer PC, McNamara JM, Houston AI, Marshall JAR, 2012. Does natural selection favour the Rescorla-Wagner rule? *Journal of Theoretical Biology* 302: 39–52.
- Visalberghi E, 1987. Acquisition of nut-cracking behavior by 2 capuchin monkeys *Cebus apella*. *Folia Primatologica* 49 (3/4): 168–181.
- von Bayern AMP, Heathcote RJP, Rutz C, Kacelnik A, 2009. The Role of experience in problem solving and innovative tool use in crows. *Current Biology* 19 (22): 1965–1968.
- Waterfall HR, 2006. *A Little Change is A Good Thing: Feature Theory, Language Acquisition and Variation Sets*. Chicago: University of Chicago.
- Weir AAS, Chappell J, Kacelnik A, 2002. Shaping of hooks in new Caledonian crows. *Science* 297 (5583): 981–981.
- Weisberg RW, 2006. *Creativity: Understanding Innovation in Problem Solving, Science, Invention, and the Arts*. Hoboken, N.J.: Wiley.
- Wimpenny JH, Weir AAS, Clayton L, Rutz C, Kacelnik A, 2009. Cognitive processes associated with sequential tool use in new Caledonian crows. *PLoS ONE* 4 (8): e6471.
- Witt U, 2009. Propositions about novelty. *Journal of Economic Behavior and Organization* 70 (1–2): 311–320.

Supplementary

1 Generalization and analogy based on nodes' similarity

Learners generalize and draw analogies among units by assessing the extent of similarity between pairs of units, as reflected by the contexts in which they are encountered. If the edge profile of unit A is similar to that of unit B, it may be reasonable to generalize and expect that unit A would be productive following elements that had preceded unit B, even if A had not been encountered in this context. This is a simplification of realistic assessments of similarity, which are most likely based on information from multiple modalities and on multiple aspects within each modality. The value of similarity between two units is taken to be the inner product of their transition probability vectors:

$$\text{Similarity}(A, B) = \frac{\sum_{x \in X} (\text{Weight}(x \Rightarrow A) \cdot \text{Weight}(x \Rightarrow B))}{\sqrt{\sum_{x \in X} \text{Weight}(x \Rightarrow A)^2 \cdot \sum_{x \in X} \text{Weight}(x \Rightarrow B)^2}} + \frac{\sum_{x \in X} (\text{Weight}(A \Rightarrow x) \cdot \text{Weight}(B \Rightarrow x))}{\sqrt{\sum_{x \in X} \text{Weight}(A \Rightarrow x)^2 \cdot \sum_{x \in X} \text{Weight}(B \Rightarrow x)^2}}$$

where A and B are the two units between which the similarity is calculated, X denotes all vertices in the graph, and \Rightarrow denotes a directed link. A non-existent directed link is treated as having a weight of zero. Intuitively, this expression simply calculates the extent of overlap between the two units' links, taking into account not only the existence of a link but also its weight: if all their links' weights are identical, the similarity between the two units is 1, and if no links are shared at all, the similarity index is zero.

2 Sentence production

To produce a sequence, the learner traverses the graph representation that it had constructed beginning at the special node $\langle \text{begin} \rangle$ and ending the sequence with the special node $\langle \text{end} \rangle$. At each node ('the focal node'), the next node to be appended to the sequence is chosen from among the nodes that the outgoing edges of the focal node go to. It is drawn with proportion to the edges' relative weights, i.e. when choosing among a number of possible nodes, those that had followed the focal node more frequently in the training phase are the most likely to be chosen.

A number of parameters may alter this process:

1) If P_{random} is set to a non-zero value, then at each choice of a new node to be appended to the sequence, there is a probability of P_{random} that a randomly chosen node from the representation will be chosen and appended (see simulation 3).

2) If the parameter that sets the preference to high-order units in producing sequences is set to a value other than zero (see simulation 6), the draw among candidate nodes that was described above changes slightly, and is executed after each edge's weight is updated to its weight in the representation, multiplied by its length raised to the power of the preference parameter. Thus, if the preference parameter is set to 2, each edge weight is updated to be $\text{original_weight} \times \text{length}^2$. The length of a node is the number of elementary units from which it is composed.

3) If the generalization parameter (P_{gen}) is set to a non-zero value, at each draw of a new node to be appended, there is a probability of P_{gen} that after the node is chosen (the 'candidate node'), it will be replaced by a node similar to it. The choice of which node to replace it with from among the nodes similar to it is also done by a draw, in which each node in the representation might be chosen at a probability that is proportional to its similarity to the node that is being replaced. If all nodes' similarity to the node that was initially chosen is zero, no replacement takes place.

Simulation 4 notes an alternative mode of generalization (and the simulation results following its use are found under "altered generalization" in table 1): in this mode, the probability that generalization will take place is affected by P_{gen} alongside the extent of similarity that the candidate node has to other units. Instead of choosing to replace the candidate node at a probability of P_{gen} , it is replaced at a probability of $P_{\text{gen}} \times SC$. SC is a measure of how similar are other nodes to the candidate node: it is equal to the sum of the similarity assessments between the candidate node and all other nodes, divided by the number of other nodes. Thus, if all other nodes in the representation had a similarity index of 1 to the candidate node, there would be a P_{gen} probability that it would be replaced, while if the candidate node has a similarity index of 0 to all other nodes, there would be no chance for it to be replaced.

3 Details of the analysis in Fig. 3

The data presented in Fig. 3 is found in numerical form in table 1 of the ESM.

Note that the results provide a near-perfect fit to a linear curve due to the large sample size, of 100,000 output sequences, and to the fact that the generalization parameter's effect on the probability of producing a creative sequence is, in the simple modality studied here, strictly linear. This is not the case in more complex modalities (results not shown).

The error bars depict two standard deviations above and below the simulation result. Standard deviations were calculated via a bootstrap analysis: 100,000 sequences were drawn at random with repeats from among the original 100,000 sequences produced by the learner. The number of creative sequences in each such sample was calculated. This process was repeated 10,000 times, giving rise to a vector of 10,000 counts of creative sequences. The standard deviation of this vector was calculated. This is equivalent to the standard deviation of a simple binomial draw whose parameter is the number of creative sentences in the learner's output, divided by the size of the output, 100,000.

Supplementary material, Table 1: full simulations' results

	Simulation 1		Simulation 2					
	L1	L2	Pgen=0	Pgen=0.01	Pgen=0.05	Pgen=0.1	Pgen=0.2	Pgen=0.4
Form of sequence, as regular expression: BEGIN [ABCD] [EFGH] [IJKL] END BEGIN A F I END BEGIN [ABCD] [GH] [KL] END Unacceptable sequences Overall number of sequences produced	100000 978 0 0 100000	100000 1074 6493 0 100000	100000 978 0 0 100000	100000 917 307 0 100000	100000 964 1575 0 100000	100000 1015 3212 0 100000	100000 1074 6493 0 100000	100000 1206 12752 0 100000
	Simulation 3		Simulation 4		Simulation 5			
Form of sequence, as regular expression: BEGIN [ABCD] [EFGH] [IJKL] END BEGIN A F I END BEGIN [ABCD] [GH] [KL] END Unacceptable sequences Overall number of sequences produced	Prandom=0.1 73189 687 337 26811 100000	L1 100000 13 0 0 100000	L2 100000 76 6588 0 100000	Altered Generalization 100000 25 4485 0 100000	L3 100000 3869 0 0 100000	L4 100000 1521 0 0 100000		
	Simulation 6							
Form of sequence, as regular expression: BEGIN [ABCD] [EFGH] [IJKL] END BEGIN A F I END BEGIN [ABCD] [GH] [KL] END Unacceptable sequences Overall number of sequences produced	A1, pref=0 100000 150 0 0 100000	A1, pref=1 100000 47 0 0 100000	A1, pref=2 100000 18 0 0 100000	A1, pref=3 100000 4 0 0 100000	A2, pref=0 100000 2 0 0 100000	A2, pref=1 100000 2 0 0 100000	A2, pref=2 100000 0 0 0 100000	A2, pref=3 100000 0 0 0 100000
	Simulation 7							
Form of sequence, as regular expression: BEGIN [AB] C D G H END BEGIN [AB] E F G H END BEGIN A E F G H END BEGIN B C D G H END Unacceptable sequences Overall number of sequences produced	L1 50285 49715 0 0 100000	L2 40019 39949 0 0 20032 100000	RBL, Pgen=0 50164 49836 0 0 0 100000	RBL, Pgen=0.1 44917 44583 0 0 10500 100000	RBL, Pgen=0.2 40617 40239 0 0 19144 100000			
	Simulation 8							
Form of sequence, as regular expression: BEGIN [AB](2)? C D G(2)? H END BEGIN [AB](2)? E F G(2)? H END BEGIN A(2)? E F G(2)? H END BEGIN B(2)? C D G(2)? H END Unacceptable sequences Overall number of sequences produced	L1 49780 50220 0 0 100000	L2 39799 40033 0 0 20168 100000	RBL, Pgen=0 49831 50169 0 0 0 100000	RBL, Pgen=0.1 43550 45137 111 110 11313 100000	RBL, Pgen=0.2 38841 40196 164 199 20963 100000			