Neural Networks 22 (2009) 674–686

Contents lists available at ScienceDirect

Neural Networks

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

2009 Special Issue

Neural dynamics of idea generation and the effects of priming

Laxmi R. Iyer [a], Simona Doboli [b], Ali A. Minai [c,*], Vincent R. Brown [d], Daniel S. Levine [e], Paul B. Paulus [f]

[a] Department of Computer Science, University of Cincinnati, Cincinnati, OH 45221, United States
[b] Department of Computer Science, Hofstra University, Hempstead, NY 11549, United States
[c] Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati, OH 45221, United States
[d] Department of Psychology, Hofstra University, Hempstead, NY 11549, United States
[e] Department of Psychology, University of Texas at Arlington, Arlington TX 76019, United States
[f] College of Science, University of Texas at Arlington, Arlington TX 76019, United States

ARTICLE INFO

Article history:
Received 6 May 2009
Received in revised form 31 May 2009
Accepted 25 June 2009

Keywords:
Ideation
Brainstorming
Semantic cognition
Creativity
Priming
Cognitive dynamics
Working memory
Multi-level attractor networks

ABSTRACT

Idea generation is a fundamental attribute of the human mind, but the cognitive and neural mechanisms underlying this process remain unclear. In this paper, we present a dynamic connectionist model for the generation of ideas within a brainstorming context. The key hypothesis underlying the model is that ideas emerge naturally from itinerant attractor dynamics in a multi-level, modular semantic space, and the potential surface underlying this dynamics is itself shaped dynamically by task context, ongoing evaluative feedback, inhibitory modulation, and short-term synaptic modification. While abstract, the model attempts to capture the interplay between semantic representations, working memory, attentional selection, reinforcement signals, and modulation. We show that, once trained on a set of contexts and ideas, the system can rapidly recall stored ideas in familiar contexts, and can generate novel ideas by efficient, multi-level dynamical search in both familiar and unfamiliar contexts.

We also use a simplified continuous-time instantiation of the model to explore the effect of priming on idea generation. In particular, we consider how priming low-accessible categories in a connectionist semantic network can lead to the generation of novel ideas. The mapping of the model onto various regions and modulatory processes in the brain is also discussed briefly.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The ability to generate relevant ideas in familiar and novel contexts is a central characteristic of the human mind, and has been studied extensively through behavioral experiments in the context of brainstorming (Osborn, 1957). These experiments have uncovered various social and cognitive factors that influence idea generation (Coskun, Paulus, Brown, & Sherwood, 2000; Dugosh & Paulus, 2005; Nijstad & Stroebe, 2006; Paulus & Brown, 2003; Paulus & Dzindolet, 1993), and have shown that priming with hints during brainstorming can enhance both the number and the quality of the generated ideas (Coskun et al., 2000; Dugosh, Paulus, Roland, & Yang, 2000; Nijstad, Stroebe, & Lodewijkx, 2002). Understanding these factors is crucial to developing better brainstorming protocols, and for explaining the idea generation process in human cognition.

Several brainstorming models based on associative memory have been developed (Brown, Tumeo, Larey, & Paulus, 1998; Nijstad & Stroebe, 2006; Paulus & Brown, 2003), and while they account for results observed in behavioral experiments, they provide limited insight into the underlying neural processes. The SIAM model by Nijstad and Stroebe (2006) is a flow-chart diagram of the search process for ideas, and is based on the free-recall model (SAM) by Raaijmakers and Shiffrin (1981). While the model describes the logical interplay between search cues, associative semantic memory, learning of retrieved ideas and storage in working memory and episodic memory, it is a high-level model that does not explain the process of how new ideas are generated. The associative model by Brown et al. (1998) and Paulus and Brown (2003) represents semantic knowledge as a network of categories and the retrieval of ideas from it as a stochastic Markov-type process. This model has been very successful in explaining brainstorming experiments (Coskun et al., 2000; Dugosh & Paulus, 2005; Dugosh et al., 2000; Paulus, Nakui, Brown, & Putman, 2006) and in predicting factors that would enhance brainstorming productivity. The model is able to emulate short-term memory effects and attention to others' ideas, and can model different styles of ideation (e.g. divergent and convergent thinking). The main shortcoming of this model too is the abstract representation of individual ideas, precluding the explicit consideration of ideas,

* Corresponding author. Tel.: +1 513 556 4783; fax: +1 513 556 7326.
E-mail address: Ali.Minai@uc.edu (A.A. Minai).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.06.019

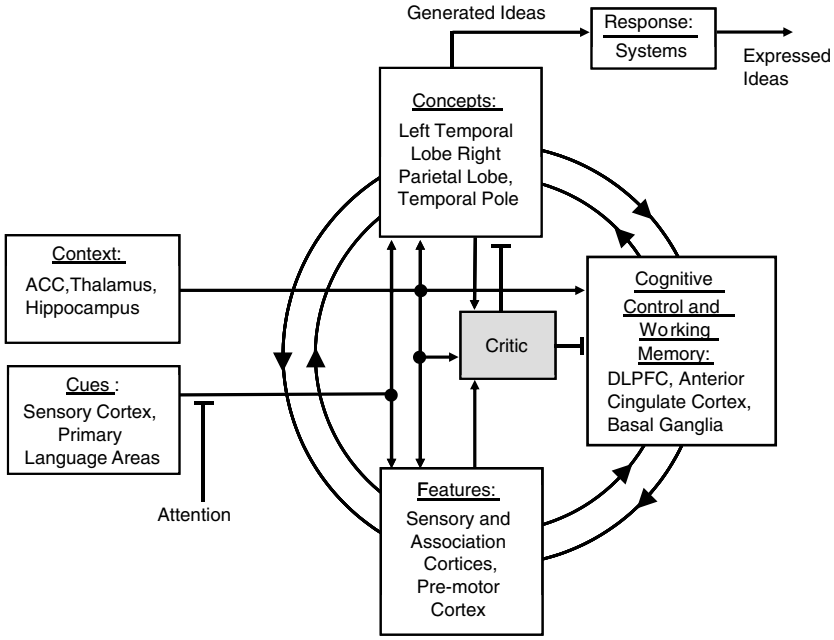

Fig. 1. Architecture of the hypothesized model for idea generation and its approximate mapping onto brain regions (Abbreviations: VLPFC — ventrolateral prefrontal cortex; DLPFC — dorsolateral prefrontal cortex; ACC — anterior cingulate cortex).

assessing their quality or novelty, and modeling the dynamics of the idea generation process.

Our group has recently proposed a connectionist model for the dynamics of idea generation (Brown & Doboli, 2006; Doboli & Minai, 2005; Doboli, Minai, & Brown, 2007; Iyer, Minai, Doboli, & Brown, 2007, 2008), motivated by experimental results on the neurobiology of semantic cognition (Caramazza & Mahon, 2003; Damasio, 1989; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Kellenbach, Brett, & Patterson, 2001; Martin, 2007; Patterson, Nestor, & Rogers, 2007; Warrington & Shallice, 1984), and theories of semantic organization (Andrews, Vigliocco, & Vinson, in press; Blei, Ng, & Jordan, 2003; Burgess & Lund, 1997; Griffiths, Steyvers, & Tenenbaum, 2007; Landauer & Dumais, 1997; McRae, de Sa, & Seidenberg, 1997; Mueller & Shiffrin, 2006; Steyvers & Tenenbaum, 2005; Verguts, Ameel, & Storms, 2004; Vigliocco, Vinson, Lewis, & Garrett, 2004), as well as insights from other connectionist models of semantic information processing (Kruschke, 1992; McClelland & Rogers, 2003; Moss, Hare, Day, & Tyler, 1994). The model is now being extended to simulate and explain actual data from behavioral experiments. In this paper, we apply the model to explain experimental effects of priming during the ideation process (Coskun et al., 2000; Dugosh & Paulus, 2005; Dugosh et al., 2000; Nijstad et al., 2002).

2. The idea generation process

Semantic information in the brain is represented at several levels, ranging from combinations of sensorimotor features (Martin, 2007; Warrington & Shallice, 1984), through amodal concepts (Kellenbach et al., 2001; Patterson et al., 2007), to semantic categories (Caramazza & Mahon, 2003). Considerable evidence now supports the idea that semantic processing involves several cortical functional networks that process and integrate information at all these levels (Damasio, 1989; Damasio et al., 1996, 2004; Martin, 2007). The areas involved include the left temporal lobe, the prefrontal cortex, the anterior cingulate cortex, the orbitofrontal cortex and parts of the occipital cortex. Regions of the right hemisphere temporal and parietal cortices are also involved (Bowden, Jung-Beeman, Fleck, & Kounios, 2005), and may provide a crucial non-linguistic component needed for the intuitive generation of novel ideas (Bowden et al., 2005; Duch, 2007; Heilman, Nadeau, & Beversdorf, 2003; Schilling, 2005). The flow of information in these cortical networks is controlled by switching processes in the basal ganglia (Graybiel, 1995; Houk, 2005), and is modulated by dopaminergic (Apicella, 2007; Schultz, 2000) and noradrenergic (Aston-Jones & Cohen, 2005) signals reflecting judgments of value.

The organization and use of semantic knowledge has also been studied extensively by researchers in linguistic cognition and computational linguistics. Several models of meaning have been developed based either on sensorimotor experiential features (Andrews et al., in press; McRae et al., 1997; Verguts et al., 2004), or on more abstract features derived from the distribution of words in text corpora (Blei et al., 2003; Burgess & Lund, 1997; Griffiths et al., 2007; Landauer & Dumais, 1997). Both approaches typically use multi-level representations of semantic knowledge in terms of features, concepts, categories, topics, etc.

Following the semantic networks approach (Boden, 1995; Mednick, 1962), we postulate that concepts are the key elements of the semantic space, and that ideas are *combinations of concepts* that arise through the dynamics of networks linking them with each other based on previous learning. This dynamics is modeled as an itinerant flow (Tsuda, 2001) where groups of co-activated concepts arise as resonant metastable patterns of activity in concept space (Doboli, Brown, & Minai, 2009; Iyer, Minai, Doboli, Brown, & Paulus, 2009; Minai, Iyer, Padur, & Doboli, 2009), and the itinerant dynamics generating them can be seen as a self-organized search process. This dynamics is modulated over time by factors such as external context, recently generated ideas and evaluative feedback from a critic to make the search more efficient.

The broad architecture of our idea generation model is shown in Fig. 1. Motivated by results from neuroscience and theories of semantic cognition, the model represents semantic knowledge in terms of features, concepts and categories. It also incorporates ideas from reinforcement learning (Sutton & Barto, 1998) and selection-based control (Graybiel, 1995; Houk, 2005). The figure indicates how the system's components might map qualitatively to specific brain regions: The encoding of features

is identified with lower-level association cortices and pre-motor regions (Martin, 2007; Warrington & Shallice, 1984); ideas and concepts are associated with higher-level association regions (Damasio et al., 1996, 2004; Kellenbach et al., 2001; Martin, 2007; Patterson et al., 2007); working memory is associated with the dorsolateral prefrontal cortex (Durstewitz, Seamans, & Sejnowski, 2000; Fuster, 2000; Goldman-Rakic, 1995), while the basal ganglia are hypothesized to function as the "gate" for the internal attention function of cognitive control (Frank & Claus, 2006; Graybiel, 1995; Houk, 2005). The system can be seen as comprising two complementary sub-systems: (1) An *idea generation sub-system* (IGS) that embeds semantic knowledge and whose dynamics generates ideas as conceptual combinations; and (2) A *critic*, which receives the generated ideas and produces evaluative feedback based on its domain knowledge about the given context. The key features of the IGS are the dynamic interactions among the different semantic levels and attractor dynamics within each level. Given a context, the task of the IGS is to produce ideas that meet specific quality criteria as evaluated by the critic using its internal model. Thus, the IGS basically learns to adapt its dynamics to the critic's liking.

In this paper, we present two instantiations of this broad model. The first one is a system with multiple interacting semantic levels, which is used to generate ideas in familiar and novel contexts. The second one is a simplified continuous-time version of the system that models the dynamics of idea generation as abstract conceptual combinations, and simulates the effect of external priming targeted to more or less accessible conceptual categories.

3. Multi-level idea generation model

Our first model includes the feature, concept and cognitive control components of the system shown in Fig. 1. It receives a context input and evaluative feedback from a critic. The central principle is that the search for ideas should emerge from the natural dynamics of the system under appropriate stimulus situations, i.e., when the system is given a task/problem context, its natural response should be to search for and generate relevant ideas. To be effective, the emergent search process must be both systematic and flexible, striking an appropriate and adaptive balance between the *exploitation* of prior knowledge and the *exploration* of new possibilities (Sutton & Barto, 1998). Purely random search is clearly not efficient, but nor is a rigid, stereotypical search that only follows previously known trajectories. The search should also not get trapped easily in sub-optimal situations, or ignore useful information that is available.

To meet these imperatives, we model the search for ideas as a two-level process:

- *Dynamic configuration of an Effective Search Domain (ESD)*: At this level, the system identifies a subset within the concept space within which to focus its search in the expectation that this subspace is disproportionately dense in good ideas for the given context. This ESD is instantiated as a slowly adaptive selectional bias created by attractor dynamics in the cognitive control regions and projected onto the space of concepts over the course of the search process.
- *Itinerant search within the ESD*: Once the ESD is selected, ideas emerge from the rapid dynamics of the activity in the ESD as groups of co-active concepts that persist temporarily as resonant metastable patterns, and dissolve. If the ESD is dense in good ideas, the search is much more effective than if it had occurred over the entire semantic space.

Thus, ideas are generated by the interaction of attractor dynamics at two spatial and temporal scales — the coarser, slower scale for adapting the ESD, and the finer, faster scale for the itinerant attractor dynamics within the ESD, which can be seen as forming a high-dimensional continuous attractor (Wu, Hamaguchi, & Amari, 2008) or invariant subspace for conceptual dynamics. The overall effectiveness of the search process depends on configuring a productive ESD and on generating good search trajectories within it. Three sources of information shape the search process at both levels:

- Associative and inferred relationships among semantic units (e.g., concepts) encountered during previous experience in specific contexts. These are represented by synaptic weights.
- Similarity between the given context and previously known contexts.
- Real-time evaluative feedback on the generated ideas from the critic, modulating the search dynamics through reward signals.

These types of information interact with inherently stochastic activity in the system, resulting in a search that, while broad and stochastic, has a strongly directed systematic component. From a theoretical perspective, the model conceptualizes the brain as a dynamical system with emergent coordination among its components at multiple spatiotemporal scales (Bressler & Kelso, 2001; Freeman, 2000; Singer et al., 1997; Tononi, Edelman, & Sporns, 1998). It belongs to the broad class of neurodynamical models characterized by a combination of resonance (Carpenter & Grossberg, 1987; Carpenter, Martens, & Ogas, 2005) and metastability (Freeman, 2000; Tsuda, 2001), where the requisite coordination arises from the circulation of information within interacting networks relaxing towards – but never completely finding – equilibrium.

3.1. System architecture

The connectivity of the system discussed in this section is a partial implementation of the large model in Fig. 1. It is shown in Fig. 2 with the IGS and the critic delineated. The critic is currently implemented as a rule-based oracle simulating both internal evaluation and feedback from the environment or a human evaluator in an actual ideation scenario. The IGS is a multi-layer connectionist network configured using a database (or *training set*) of contexts and relevant ideas from these contexts. After configuration, the system is required to do the following:

- *Recall*: Automatically recall familiar ideas in familiar contexts.
- *Context-specific generalization over idea space*: Generate relevant new ideas in familiar contexts.
- *Generalization over contexts:* Search efficiently for good ideas in unfamiliar contexts within the system's knowledge domain.
- *Learning*: Incorporate new contexts and ideas into the system's structure so that they become familiar over time.

3.1.1. The Idea Generation Sub-system (IGS)

The IGS is a multiply recurrent connectionist system that embeds semantic knowledge, and is driven by an external *context input* specifying the task or problem at hand.

The context input, $X = \{x_i\}$, is encoded through N_x binary bits grouped together into two subsets: (1) The *type bits* indicate the type or domain of the context, e.g., eating, vacation, shopping, etc.; (2) The *attribute bits* encode the descriptive attributes of the context, e.g., inexpensive, exciting, short, etc. In any given context, only one type bit is set to 1, but any number of attribute bits can be set to 1. Opposite or mutually exclusive attributes are encoded by separate bits (Verguts et al., 2004). This representation allows the

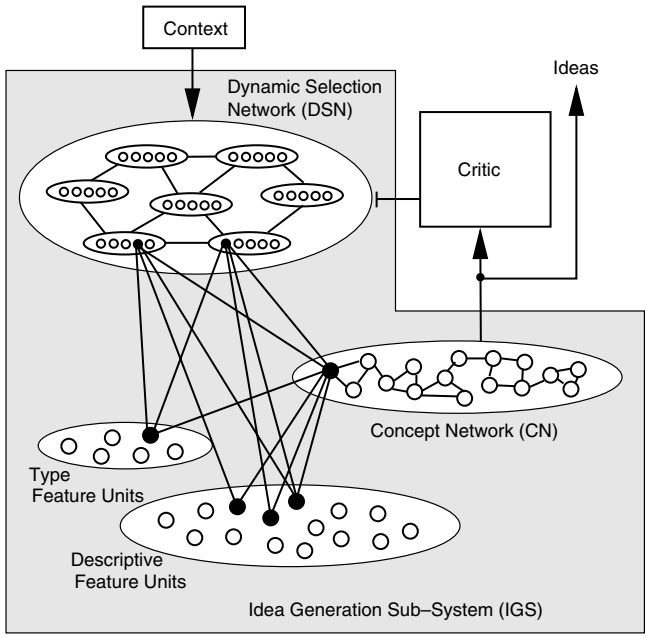

Fig. 2. Architecture of the multi-level model, showing the Idea Generation Subsystem (IGS) and the critic. Inter-layer connections are shown only for a single concept unit for clarity. Other concept units are connected to feature and category units in a similar way (see text for details). The large ovals in the DSN are modules, with circles indicating selector units. The architecture and connections of the DSN are only shown schematically. The interneurons modulating learning (shown in Fig. 3) are not shown here for clarity. All connections shown without arrows are bi-directional, though their strengths are not necessarily the same in each direction.

similarity between different contexts to be represented naturally by Hamming distance.

The fundamental semantic elements of the IGS are *concepts* (corresponding to nouns and verbs), encoded by *concept units* in the *Concept Network* (CN). Each concept maps onto a set of *features*, which are represented by *feature units* in the *Feature Layer* (FL). Concepts that are similar on specific subsets of features are grouped together into non-exclusive *categories* represented by *selector units* in the *Dynamic Selection Network* (DSN), which provides the selective bias to configure the EDS, thus implementing the functions of internal attention and working memory (Dehaene, Kerszberg, & Changeux, 1998; Durstewitz et al., 2000; Fuster, 2000; Goldman-Rakic, 1995; Houk, 2005; Newman & Grace, 1999).

The N_c concept units in the CN provide amodal representations of concepts, and are connected recurrently based on their joint participation in known good ideas (see below). Each concept also has a distributed *feature representation* with N_f features divided into two subsets: (a) A set F^T of N_f^T *type features* denoting functional types for concepts, e.g., "accommodation", "food", "mode of transportation", etc.; and (b) a set F^D of N_f^D *descriptive features* denoting attributes such as "expensive", "large", "strenuous", "red", etc., with distinct elements for opposite or mutually exclusive descriptive features (Verguts et al., 2004). The type features are binary, indicating whether a concept is of a certain functional type, while descriptive features are real-valued, indicating the degree to which the features hold for a concept. Thus, the feature representation for concept i is given by $\Phi_i = \{\Phi_i^T \ \Phi_i^D\}$, where $\Phi_i^T = \{\phi_{ij}^T\} \in \{0, 1\}^{N_f^T}$ and $\Phi_i^D = \{\phi_{ij}^D\} \in \Re^{N_D}$ are the vectors of values taken for the concept by type and descriptive features, respectively. Most concepts are associated with only one type feature. The *spanning subspace* for concept i is defined as $\Psi_i = \{j \in F_t^D \ | \ \phi_{ij}^D > 0.5\}$, while the spanning subspace for all concepts of type k is given by $\Psi(k) = \bigcup_{i|\phi_{ik}^T=1} \Psi_i$. There is a feature unit in the FL for each feature, giving N_f units divided into N_f^T *type feature units* and N_f^D *descriptive feature units*. Concept unit i and feature unit j are connected by reciprocal connections with weights $w_{ij}^{cf} = w_{ij}^{fc} = \phi_{ij}^{\diamond}$, where $\diamond$ denotes t or d.

The N_s selector units of the DSN are organized into M_s modules called *templates*. Each selector unit, i, is tuned to a set of features, $\Xi_i = \{f_i^t; f_{i_1}^d, f_{i_2}^d, \ldots, f_{i_q}^d\}$, where the first element is a type feature and the rest are descriptive features, and q is a small positive integer. The set of features Ξ_i is termed the *focus set* for selector unit i, which groups together all concepts of type f_i^t that have sufficiently high values of $f_{i_1}^d, f_{i_2}^d, \ldots, f_{i_q}^d$, e.g., "cheap, fast, long-distance transport" or "expensive, urban accommodation". The set of concepts covered by selector unit i is termed its *member set* and denoted Γ_i. For convenience, we term the semantic entity represented by selector unit i as its associated category. There are several identically tuned selector units in the DSN for each category, but never in the same module. Each module groups m_a selector units of distinct types (i.e., with distinct type features) into a densely connected local network.

Each selector unit, i has four sets of connections: Reciprocal binary gating connections with all concept units, j in Γ_i, binary connections, w_{ij}^{sf} from the feature units, $j \in \Xi_i$, real-valued modifiable associative connections, w_{ij}^{sx} from the context bits, $j \in X$, and initially weak real-valued connections, w_{ij}^{ss}, from other selector units, $j \in DSN$, that are modified both during initial learning and during search (see below).

The context bits and selector units can be seen as representing the highly cue-, feature- and context-selective cell assemblies in the prefrontal cortex (Constantinidis, Franowicz, & Goldman-Rakic, 2001; Fuster, 2000; Goldman-Rakic, 1995) and the basal ganglia (Graybiel, 1995).

Initial configuration of the IGS:

Configuration of the IGS begins with a given set of concepts – the *concept vocabulary* – already defined in terms of a given set of features comprising the *feature set*. These are currently developed manually, though work is underway to automate this process. The system is trained using a *training set*, which includes a set of contexts, $G = G^1, \ldots, G^m$, and a set of ν_u relevant ideas, J_k^u for each context G^u. Each J_k^u is encoded as a N_c-bit binary vector indicating the concepts comprising it.

The purpose of initial configuration is to embed the semantic information in the training set. It is done in two stages as follows:

Stage I: During this stage, the given set of concepts and features is used in conjunction with the training set of ideas to obtain a diverse repertoire of categories and templates. The concepts corresponding to a particular type are considered together and clustered in the spanning subspace for the type. The clusters identified are then checked against the ideas in the training set to see if they are disproportionately associated with certain contexts, and clusters that do so are used as seed categories. Any concepts not included in a cluster are given their own singleton clusters. More clusters are then generated heuristically by taking subspaces of the selected clusters. Once the categories are defined, a large number of selector units are created with multiple instances of each category, and grouped together randomly into incipient modules, making sure that multiple categories of the same type are not included in the same module.

Stage II: In this stage, the sub-system comprising the concept units, the selector units and the context input is trained on the ideas in the training set using neural learning methods. Feature units are excluded from the learning process during this stage. Learning occurs at two levels:

Learning in the DSN: The ideas in the training set are projected onto the selector units in random order through the concept network while the corresponding context pattern is clamped on the context

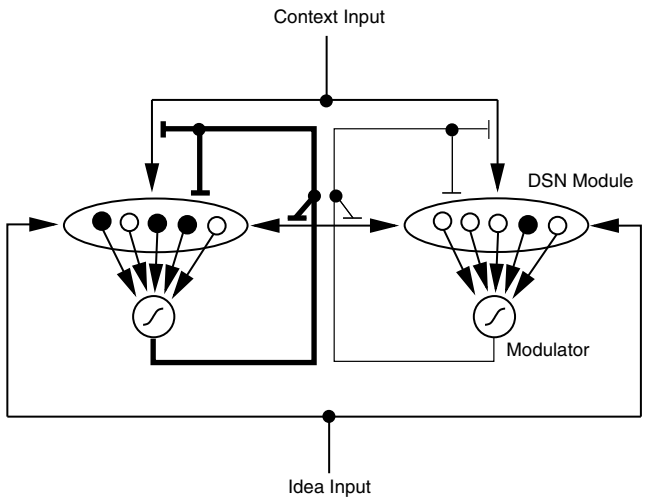

Fig. 3. Learning in the DSN: Synaptic modification for all connections is modulated nonlinearly by the total activity of the module, ensuring that highly activated modules learn disproportionately more. This figure shows two modules corresponding to those shown in the DSN in Fig. 2. Arrows indicate information signals while T-endings denote connections modulating the learning rate. The bolder modulatory signal for the left module indicates that its neurons undergo greater synaptic modification because the module is more active than the module on the right.

inputs. Each idea is presented to the system for several consecutive steps called a *presentation cycle*. At the start of a presentation cycle, the selector units' activity is determined completely by the training idea input. The network is then allowed to relax based on the recurrent connectivity of the DSN, but with no influence from the context input. Neural firing at this time is competitive, with the K_s most excited units allowed to fire. The resulting activity pattern at the end of the learning cycle is embedded auto-associatively into the network by the Hebbian rule: $w_{ij}^{ss}(t+1) = w_{ij}^{ss}(t) + \eta_{ss}(t) z_j^s(t) z_i^s(t)(1 - w_{ij}^{ss}(t))$. It is also associated with the context input pattern using a presynaptically-gated Hebbian rule: $w_{ij}^{sx}(t+1) = w_{ij}^{sx}(t) + \eta_{sx}(t) x_j(t)[z_i^s(t) - w_{ij}^{sx}(t)]$, where t indexes the presentation cycle, x_j is the activity of the jth context bit and z_i^s that of the ith selector unit. The learning rates for the two associations (which can be different) are modulated nonlinearly by the activity of the module to which the postsynaptic unit belongs: $\eta(t) = \eta_{fixed} S[\sum_k z_k^s(t)]$, where k indexes all units in the module for unit i, η_{fixed} is a fixed number between 0 and 1, and $S[.]$ is a highly nonlinear sigmoid function. This is modeled as modulatory feedback through a module-specific interneuron, as shown in Fig. 3. This soft-competitive learning at the module level ensures that highly active modules learn much more than those with low activity, establishing a preferential link between contexts and ideas that are appropriate in those contexts.

Initially, all the recurrent weights in the DSN are low, but as they change with learning, they play an increasing role in the process. Ultimately, each context becomes associated with a *graded attractor* comprising a strongly connected core and an increasingly weakly connected periphery of selector units. The strength of association between units of the attractor associated with each context depends on the frequency of their participation in ideas for that context. When a context is presented during recall, it activates its associated attractor from the core outward, depending on the level of lateral inhibition in the DSN.

Learning in the Concept Network: The concept network is configured concurrently with the DSN, using the ideas from the training set. As each idea is presented, the recurrent connections between the concept units it activates are modified auto-associatively to give $w_{ij}^{cc} \sim P[z_i^c(t) = 1,\ z_j^c(t) = 1]$, producing preferentially linked concept groups representing pieces of viable ideas over all trained contexts. Some of these groups are explicitly present in the training ideas while others arise emergently as the system is trained.

IGS unit activities

The units in the three components of the IGS function as follows:

Concept units: The net *synaptic input* to concept unit i is given by:

$$y_i^c(t) = \bigvee_{j|i\in\Gamma_j} z_j^s(t) \left[\gamma^{cf} \sum_{j\in FL} w_{ij}^{cf} z_j^f(t-1) + \gamma^{cc} \sum_{j\in CN} w_{ij}^{cc} z_j^c(t-1) + \gamma_{noise}^c \eta_i(t) \right] \quad (1)$$

where the outputs from units in the CN, the FL and the DSN are denoted by z_j^c, z_j^f and z_j^s, respectively, w_{ij}^{cf} are binary weights to concept unit i from feature units, and w_{ij}^{cc} are weights from other concept units, j, $\eta_i(t)$ is uniform white noise between 0 and 1, and γ^{cs}, γ^{cc}, and γ_{noise}^c are fixed gain parameters.

The *state* of concept unit i at time t is given by:

$$u_i^c(t) = (1-\alpha^c)u_i^c(t-1) + \alpha^c y_i^c(t); \quad 0 < \alpha^c < 1. \quad (2)$$

Activity in the CN is competitive, so that the K_c non-refractory units with the highest $u^c(t) > 0$ can fire at time t. Once fired, unit i may remain active for a random *activity duration* $\varphi \sim U[\varphi_{\min}^c, \varphi_{\max}^c]$ if $u_i^c(t)$ remains sufficiently high. After that, i enters a *refractory period* of random duration $\rho \sim U[\rho_{\min}^c, \rho_{\max}^c]$. The set of K_c co-active units at any time comprise an *idea*, but the idea is recognized by the system only if it persists for τ_A successive time steps, termed the *awareness threshold*. Since only co-active concepts with strong mutual weights can persist in the face of competition and the weights depend on the joint utility of concepts in prior ideas, most ideas that reach awareness comprise concepts that "make sense" together. The gating from the DSN ensures that only concepts belonging to the member set of an active selector unit are eligible for activity, which defines the current ESD. Finally, disynaptic recurrent input via the FL allows active concepts to bias successor concepts similar to the currently active ones, making it possible to generate analogous ideas successively.

Feature units: Feature unit i receives input only from concept units: $u_i = \sum_{j\in CN} w_{ij}^{fc} z_j^c$, and has output $z_i^f = 1$ if $u_i > 0.5$ and 0 else. Thus, if the current idea has sufficient support for a feature, it is active.

Selector units: Selector units use a model similar to the concept units. The net synaptic input to selector unit i is given by:

$$y_i^s(t) = \gamma^{sx}(t) \sum_{j\in X} w_{ij}^{sx} x_j(t) + \gamma^{ss}(t) \sum_{j\in DSN} w_{ij}^{ss}(t) z_j^s(t-1) + \gamma^{sf}(t) \sum_{j\in FN} w_{ij}^{sf} z_j^f(t-1) + \gamma^{sc}(t) \sum_{j\in CN} w_{ij}^{sc} z_j^c(t-1) + \gamma_{noise}^s \eta_i(t) \quad (3)$$

where x_j are context bits, η is uniform white noise, γ^{sx}, γ^{sc}, γ^{ss}, and γ_{noise}^s are time-dependent gain parameters, and the remaining items are as defined earlier. Note that the recurrent weights in the DSN are modified during the search process, as discussed below. The *state* of selector unit, i at time t is given by:

$$u_i^s(t) = (1-\alpha^s)u_i^s(t-1) + \alpha^s y_i^s(t); \quad 0 < \alpha^s < 1 \quad (4)$$

where $\alpha^s \ll \alpha^c$, ensuring that the DSN activity changes much more slowly than CN activity, as required for working memory functionality (Durstewitz et al., 2000). As in the CN, selector units fire competitively, with the $K_s(t)$ non-refractory units with the highest $u^s(t) > 0$ activated at time t. Like concept units, selector

units also have random activity duration, $\varphi \sim U[\varphi^s_{\min}, \varphi^s_{\max}]$ and refractory period, $\rho \sim U[\rho^s_{\min}, \rho^s_{\max}]$, both of which are much longer than for concept units.

A key feature of the DSN is modulation of the competitive inhibition, $I^s(t)$, which determines $K_s(t)$, so that different numbers of selector units can be active at different times (see below for details).

IGS dynamics: The FL in the current model does not have any intrinsic dynamics, and serves mainly as an intermediate layer providing disynaptic linkage for CN units and DSN units. The DSN has slowly itinerant attractor dynamics biased by the context input and with varying activity levels determined by the modulated inhibition. Finally, the CN has faster itinerant attractor dynamics confined to the current ESD, with small groups of concepts coming on, either persisting as ideas (if mutually compatible) or fading away quickly.

3.1.2. The critic

While familiar contexts are expected to yield good ideas automatically, effortful search (Dehaene et al., 1998) is needed to find a productive ESD in unfamiliar contexts. The critic mediates this process by providing evaluative feedback on the generated ideas. This, in turn, modulates competitive inhibition on the selector units and drives the search process as discussed in the next section.

In the current model, the critic is implemented as a self-consistent rule-based system with extensive domain knowledge. This models the internal evaluative processes hypothesized to reside in the basal ganglia (Graybiel, 1995; Houk, 2005), the anterior cingulate cortex (ACC) (Botvinick, Cohen, & Carter, 2004), the orbitofrontal cortex (OFC) (Bechara, Damasio, & Damasio, 2000), etc., as well as external evaluation from the environment. The idea generation system's performance is judged by how well it can infer the preferences of this critic based only on reinforcement feedback for the ideas the system generates — over both familiar contexts (those used in training) and unfamiliar ones.

As the system generates ideas, the critic evaluates them based on five criteria:

Admissibility, A, is the fraction of required concept types included in the idea.
Coherence, C, is 1 if the concepts in the idea are compatible, 0 else.
Quality Q, is a function of the fraction of *quality features* included in the idea, and their strength. Quality features are defined in the critic.
Novelty N, is 0 if the idea occurs in the training set; 0.2 if the idea shares three concepts with a training set idea; 0.7 if it shares two; and 1 if it shares one or none.
Efficiency, $E = 1 - R$, where *redundancy*, R, is the fraction of concepts in the idea that are unnecessary, e.g., a *vacation* context requires a place to stay, but not more than one. However, multiple activities on a *vacation* may not be redundant.

Two evaluative signals are generated for each idea as follows. A *reward* value, r, is calculated as: $r = AC(aQ + (1 - a)N)$, where parameter a indicates the relative importance of quality and novelty. Also, an *adjusted reward* r_c, is computed as follows: (1) If $Q > 0.8$ and $E > 0.6$, $r_c = r$; (2) If $Q \leq 0.8$ and $E = 1$, $r_c = r$; and (3) Else, $r_c = 0$. Their use is discussed in the next section.

3.2. Idea search dynamics

After the system is configured, it is tested by giving it both familiar and unfamiliar contexts as input. The unfamiliar contexts differ from familiar ones in their features, but still use the same feature space.

To start off, the *top-down gains*, γ^{sx} and γ^{fc}, are high while the other connections are relatively ineffective. This causes the external context to activate the DSN, which biases the concept network to define an initial ESD. As ideas are generated andevaluated, a *satisfaction level*, $\rho(t)$, is calculated as: $\rho(t) = \rho(t-1) + \Delta\rho(t)$, with

$$\Delta\rho(t) = H[r(t)](1 - \rho(t-1))r(t) - (1 - H[r(t)])\beta\rho(t-1) \quad (5)$$

where $H[\ .\ ]$ is the Heaviside step function.

The search dynamics emerges from the interaction of the natural IGS dynamics described earlier with three modulatory processes driven by evaluative feedback from the critic. Each of these is described next.

(a) **Modulation of DSN inhibition:** During the search process, if the satisfaction level, $\rho(t)$ is low, the inhibition $I^s(t)$ is lowered to expand the ESD by recruiting a larger set of active selector units corresponding to the periphery of the attractor activated by the context, and when $\rho(t)$ increases, the ESD is narrowed again by increasing $I^s(t)$. The value of $\rho(t)$ is checked every T_s time steps ($T_s = 500$ in the simulations below), and I^s is changed as $I^s(t+1) = I^s(t) + \Delta I^s(t)$, with:

$$\Delta I^s(t) = \begin{cases} \epsilon(\rho(t) - \min(I^s(t), \rho(t))), & \text{if } \rho(t) \geq \omega \\ -\delta_I & \text{if } \rho(t) < \omega \end{cases} \quad (6)$$

where $\omega \in (0, 1)$ is a *frustration threshold*, and $\epsilon, \delta_I \in (0, 1)$ are increment and decrement parameters. The activity level in the DSN is then changed as $K_s(t+1) = K_s(t) + \Delta K_s(t)$, where:

$$\Delta K_s(t) = \begin{cases} -\min\left[\lceil \delta_K^- \Delta I^s(t) \rceil, K_s(t) - K_s^{\min}\right], & \text{if } \rho(t) \geq \omega \\ \delta_K^+ & \text{if } \rho(t) < \omega \end{cases} \quad (7)$$

where δ_K^- and δ_K^+ are integers, and $K_s^{\min}$ is a minimum activity level for the DSN. The new levels of K_s and I^s then remain unchanged for the next T_s time steps.

(b) **Modulation of gains:** Initially, the *top-down gains*, γ^{sx} and γ^{fc}, are high while the other connections are relatively ineffective. However, the *bottom-up gains*, γ^{sc}, γ^{cf} and γ^{sf} are modulated in synchrony with K_s. Thus, as the search gets more relaxed without sufficient reward, influence from the internal structure of the system – encoding previously inferred semantic relationships – increases.

(c) **Modification of DSN weights:** Whenever an idea is rewarded, i.e., $r(t) > 0$, the connections between active selector units associated with its constituent concepts have the connections between them potentiated temporarily. This is a *working memory* effect that may use mechanisms such as those recently hypothesized by Mongillo, Barak, and Tsodyks (2008) corresponding to dopamine-mediated increase in the excitability of the selector units (Apicella, 2007; Durstewitz, Kelc, & Güntürkün, 1999; Schultz, 2000) due to the reward signal.

The interplay between effects (a) and (c) above is crucial. The initial activity pattern in the DSN (and therefore, the initial ESD in the CN) is based on the similarity of the given context with previously known ones. If that does not yield good ideas, i.e., high $\rho(t)$, lowered I^s causes the DSN activity pattern – and therefore the ESD – to expand in a way that is based on the prior connectivity of the DSN, i.e., the semantic knowledge embedded during training. Eventually, this begins to yield good ideas, but with low efficiency because the ESD is much broader than it needs to be. However, the specific reward-dependent modification of DSN weights ensures that when the activity level in the DSN contracts again due to increased $\rho(t)$ and I^s, the units that survive are likelier to be the ones *actually responsible* for the recent rewards. Going back and forth over the expansion–contraction cycle has a ratchet-like (or

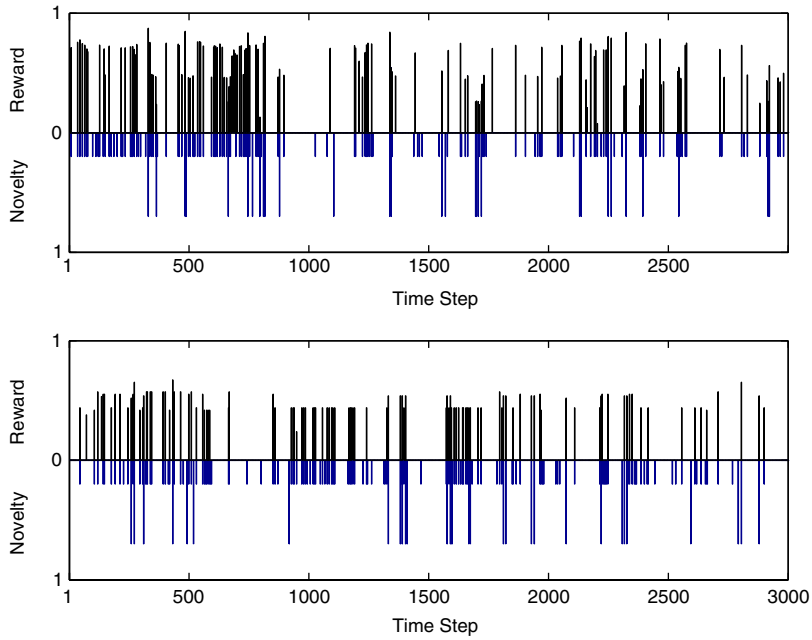

Fig. 4. Reward and Novelty time series for two familiar contexts: (*Romantic vacation* and *Adventurous vacation*). Both quantities are between 0 and 1, but Novelty is shown pointing downwards to allow plotting both quantities on the same graph, making it possible to assess the overall quality and novelty of each idea. Only rewards for unique ideas are shown. See text for discussion.

annealing-like) effect, leading to the configuration of better and better ESDs. This effect is illustrated in Fig. 6 later in this paper.

The increased potentiation in the DSN due to effect (c) is temporary, as working memory should be. but it leaves behind a residual effect that, over several runs, slowly reconfiguring the DSN to embed information about the novel contexts it encounters. Also, the recurrent weights between active concept units are potentiated slightly during the search process whenever the adjusted reward, $r_c(t) > 0$, i.e., when a rewarded idea is sufficiently good and efficient. These concepts are thus more likely to be co-active in the future. Over time, this modifies the joint utility-based connectivity of the concept layer.

3.3. Simulation results and discussion

The model described above has been instantiated using a training set based on 5 contexts, 108 concepts and 1000 training ideas. All contexts were drawn from the general domain of *outings and vacations*, and are: (1) *School field trip*; (2) *Guys' day out*; (3) *Romantic vacation*; (4) *Adventurous vacation*; and (5) *Winter vacation*. The system has 560 selector units grouped into 140 modules of 4 units each. Typical concepts include `five-star hotel`, `skiing`, `hitch-hiking`, `art museum`, etc., while typical categories include `expensive accommodation`, `romantic activity`, etc.

After configuration, the system was tested by running it for the 5 familiar contexts, one unfamiliar context (*Nature-related vacation*) similar to the familiar ones, and one highly novel context (*Physically active outing*). In each case, the system was allowed to run for 3000 time steps, and the ideas generated were evaluated on all five metrics listed above. The composite reward, r, captures the results, and Fig. 4 plots the reward and novelty time-series for two familiar contexts using $a = 0.8$. Thus, ideas are rewarded mostly for quality, but with a significant component for novelty. Only rewards for unique ideas are shown. As can be seen, the system was able to generate good ideas rapidly for the familiar contexts, but most of the initial ideas were not novel, i.e., the system worked mainly through exploitation of embedded knowledge. This corresponds to the so-called automatic response generation (Dehaene et al., 1998). Once exploitable ideas were exhausted, the search moved on to more novel good ideas, which were found — more so for the first context than the second. This shows that the system can generalize to novel ideas within a familiar context, but only after first exploiting known ones.

Fig. 5 is analogous to Fig. 4, but this time for the two novel contexts. For the slightly novel context (top panel), the results are similar to those for familiar ones, albeit less productive — as is to be expected. The search first exploits what it can before exploring for more novel ideas. It finds many novel ideas, but none of them is especially good. For the highly novel context, in contrast, exploitation fails utterly, forcing the system to explore by expanding the search until it begins to discover good ideas. When they are found, they are highly novel, as they are expected to be in a highly novel context. Fig. 6 shows a more detailed view of the search in the highly novel context over a longer duration. Beginning with only 3 active selector units, insufficient reward causes the search to expand until step 1500. At that point, the search suddenly begins finding good ideas, indicating that the correct set of categories is now included in the active group. This leads to an increase in inhibition and exploration of several category combinations at that level. However, the flow of good ideas dries up by step 2600, causing inhibition to go down again until good ideas reappear after step 3500. This time, when inhibition goes up, at step 4000, it captures a sparse set of highly productive categories. The process of ratcheting expansion and contraction of DSN activity was thus able to discover an efficient ESD for the context.

Upon closer analysis of the results in Fig. 6, it turned out that the search had discovered that, for this context (*Physically active outing*), the critic was perfectly satisfied with a combinations of concepts from two categories: *Sporting activities* and *Kids' fun and games*, and there was no need to include accommodations or long-distance transport. Thus, an idea such as `jet-skiing, skateboarding, beach volleyball, hula-hooping` scored fairly high. In contrast, the other unfamiliar context, *Nature-related vacation*, required both transport and accommodation,

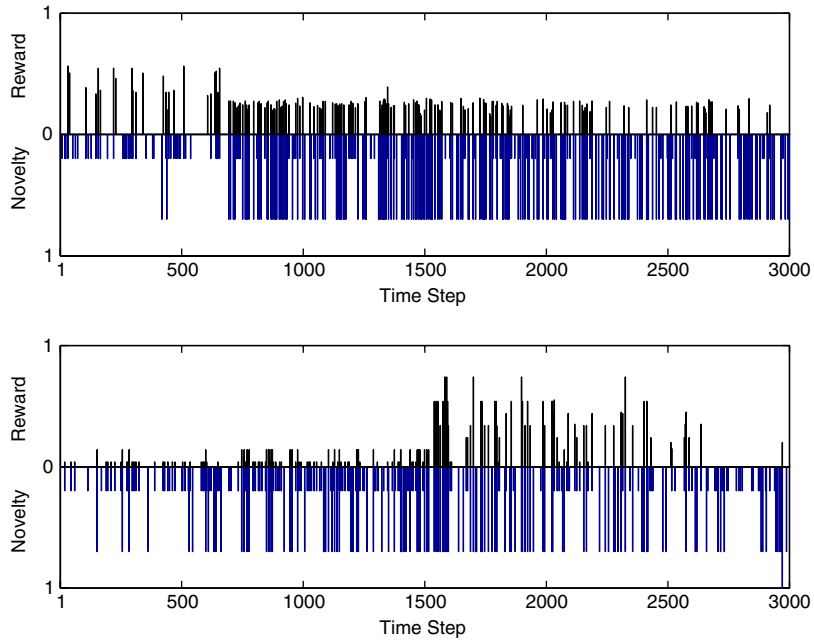

Fig. 5. Reward and Novelty time series for a somewhat unfamiliar context, *Nature-related vacation* (top panel), and a highly novel context, *Physically active outing* (bottom panel). Plotting conventions are as for Fig. 4. See text for discussion.

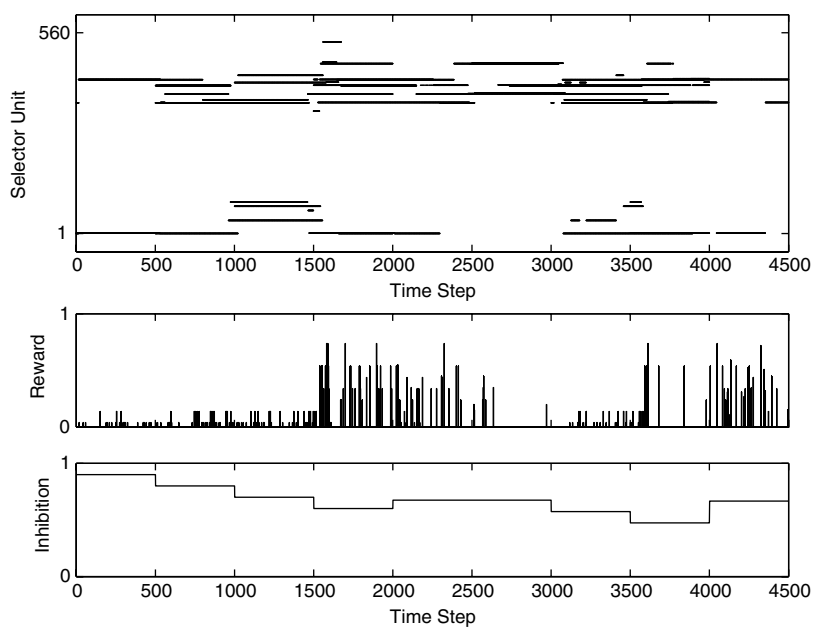

Fig. 6. Time series of selector unit activity (top panel), unique rewards (middle panel) and DSN inhibition (bottom panel) for the highly novel context, (*Physically active outing*). Each line in the top panel shows the activity of one selector unit over time. The number of units active expands as inhibition decreases and contracts as inhibition increases following higher rewards. See text for more discussion.

and a typical good idea was `air-travel, 5-star hotel, kayaking, canoeing`. Neither of these ideas – or these contexts – were part of the training set. These results demonstrate that the system can generalize to novel contexts in an appropriate way, exploit existing knowledge even in these novel contexts (if possible), and can also find completely novel ideas when necessary.

4. The conceptual dynamics model

As a complement to the work described in the previous section, we used an architecturally simpler continuous-time model of abstract conceptual dynamics alone to explore how priming categories with high- or low-accessibility affects the generation of ideas. This model can be seen as combining the functionality of the concept and dynamic selection networks into a modular oscillator network. The model is inspired by the theory of semantic networks (Mednick, 1962) and the modular organization of the cortex with long range excitatory connections and local inhibition (Constantinidis et al., 2001; Goldman-Rakic, 1995; Mountcastle, 1997).

This model retains the essential features of the more detailed model above: (1) Ideas are represented explicitly as combinations of individual *concepts* grouped together into *categories*; (2) Ideas are generated through a dynamic, adaptive neural system using biologically inspired mechanisms; (3) Useful, novel ideas are generated as conceptual combinations from one or more categories; and (4) The dynamics of idea generation is determined by: (a) the

connectivity pattern and strength of existing semantic relations between concepts, (b) the sequence of previously generated ideas, and (c) inhibitory modulation of activity controlling the semantic distance between co-active concepts.

The system consists of a single layer of N_c interconnected *concept units* representing single concepts. Each concept unit is a neural oscillator comprising a coupled pair of elements, one excitatory and one inhibitory, each of which can be seen as populations or assemblies of neurons. The ith excitatory and inhibitory units follow the equations:

$$\frac{1}{\tau_E}\frac{\mathrm{d}E_i(t)}{\mathrm{d}t} = -E_i(t) + f\left(a_iE_i(t) + \gamma_E \sum_{j=1,j\neq i}^{N_c} w_{ij}(t)E_j(t) - b_iI_i(t) - \gamma_I I_{global}(t) + S_i(t) + \eta_i(t)\right) \quad (8)$$

$$\frac{1}{\tau_I}\mathrm{d}I_i(t)/\mathrm{d}t = -I_i(t) + f(c_iE_i(t)) \quad (9)$$

where E_i and I_i are, respectively, the levels of activity in the ith excitatory and inhibitory units, τ_E and τ_I are time constants, w_{ij} is the strength of the connection from excitatory unit j to excitatory unit i, a_i, b_i and c_i are the strengths of self-excitation, cross-inhibition and cross-excitation, respectively, in the ith oscillator, $S_i(t)$ is an external, excitatory input, η_i is Gaussian noise, I_{global} is a global inhibitory input, representing lateral inhibition and defined as: $I_{global} = \sum_{j=1}^{N_c} E_j(t)$, and γ_E, γ_I are, respectively, the excitatory and inhibitory gains. The activation function $f(\ )$ is the sigmoid $f(x) = 1/(1 + \exp(-\varrho_\diamond(x - \vartheta_\diamond)))$, where $\varrho_\diamond$ and $\vartheta_\diamond$ control the sigmoid's shape and $\diamond$ is E or I (the parameters have different values for the excitatory and inhibitory units.) This model is similar to Wilson and Cowan's model of interacting excitatory and inhibitory neural populations (Wilson & Cowan, 1972). The values chosen here are such that a concept unit exhibits limit cycle oscillations for a range of non-zero external input values and stable fixed point behavior for smaller or larger values. All parameters are the same for all concept units, except a_i, b_i and c_i for which a small random deviation around the fixed value is added.

Since this model is abstract (i.e., not based on linguistic elements), categories are encoded implicitly by the pattern of connectivity between concept units. The connection from concept j to concept i is a measure of how often the two concepts are active together in a good idea compared with the number of times j is active alone. This induces an inhomogeneous connectivity structure that makes specific categories more or less accessible from others. Given the currently active category, the remaining categories can be assigned to subsets comprising those with low, medium and high accessibility (see below).

The dynamics of the model, given a transient initial activation within one category, depends on the balance between inter-concept excitation and inhibition (γ_E and γ_I) and is as follows: With no global inhibition ($\gamma_I = 0$) and with γ_E larger than a threshold, all concept units become active and stay active; lowering the inter-concept excitation causes the system to exhibit synchronized oscillations, while lowering γ_E causes the activity to die out further. When the global inhibition is non-zero and higher than a threshold, the dynamics is more interesting. The activity remains confined to a small number of concepts, and switches from one set of active concepts to another (i.e. from one idea to another) itinerantly. The small, non-zero mean Gaussian noise added to each excitatory unit ensures that the activity does not die out.

The model has short-term dynamic connection strengths between different excitatory units: Connections between concepts active at small positive time differences slowly decay to 0 with a time constant τ_d, while all weights lower than the fixed, initial value – the steady state value – go back up with a time constant τ_r. This depletion-and-recovery dynamics has an inhibitory short-term effect that lowers the likelihood of ideas being repeated over short time spans. It corresponds to slowly lowering the probability of staying in the same category while generating ideas from it. This is also a feature of the associative model of Brown et al. (1998). In contrast, the SIAM model (Nijstad & Stroebe, 2006) increases the likelihood of repeating ideas. The ratio between τ_r and τ_d controls the speed with which connections between concepts in previously active ideas go down and up.

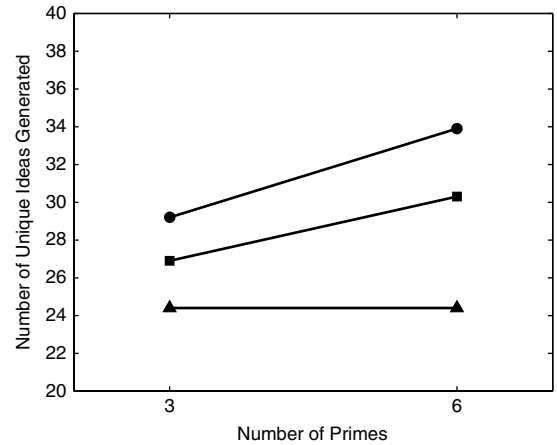

Fig. 7. Number of ideas generated as a function of number and accessibility of primes. Priming low-accessible categories (solid circles) produced more unique ideas than priming high-accessible categories (solid squares). The solid triangles indicate the control (no priming) case. Data are re-plotted from Leggett (1997).

Ideas are represented by the activity of excitatory units: all units with an activity over a threshold θ_{on} at time t comprise an idea. The novelty of an idea is judged by whether its component concepts were already strongly connected (indicating prior co-activation). More novel ideas are likely to be spread over concepts in more than one category.

4.1. Modeling the effect of category priming

Experimental results (Coskun et al., 2000; Dugosh & Paulus, 2005; Dugosh et al., 2000; Nijstad et al., 2002) show that providing primes in the form of relevant ideas or category labels can have a positive effect on brainstorming performance. Memory representations differ in the strength of their associations with each other in an individual's semantic network, allowing some to be accessed more readily than others. From a brainstorming perspective, this means that some ideas may be difficult for an individual to generate, but may become more accessible when activated by priming input. Thus, category *accessibility* is an important factor in understanding brainstorming performance (Brown et al., 1998).

To study the effects of external priming, Leggett (1997) provided individual brainstormers with ideas from high frequency categories which individuals use to generate most of their ideas and from low frequency categories that are used less often. A high frequency category is considered more accessible than a low frequency one. Leggett (1997) also primed individuals with either a high or low number of exemplars from each category. A high number of primes indeed resulted in more non-repetitive ideas compared to a low number of primes or to the control case when no primes were presented, and primes from low-accessible categories led to the production of more unique ideas than the same number of primes from high-accessible categories, as shown in Fig. 7. These results are reproduced by simulations of the model (see Fig. 10).

The neural model described above was used to study: (1) The effect ofpriming low-accessible categories versus high-accessible

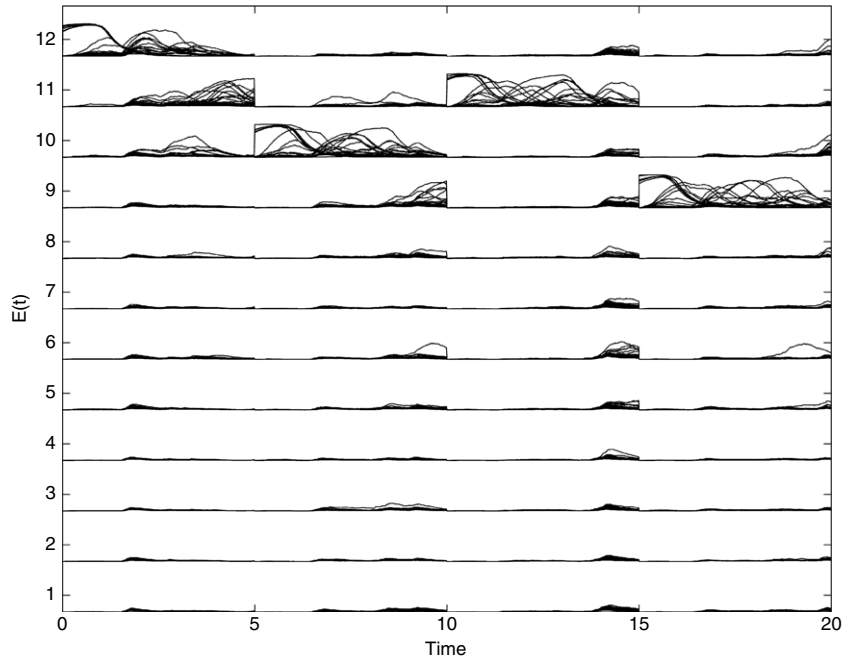

Fig. 8. Performance of the model for primes from high-accessible categories. The figure shows the activity of excitatory units over time in each category, beginning with a high-accessible category. Plots numbered 1–4 represent low-accessible categories, 5–8 medium-accessible categories and 9–12 high-accessible categories. The results show that activity is confined mainly to high-accessible categories.

categories during ideation, and (2) The effect of priming with a low number versus a high number of hints during a brainstorming session. Here, the N_c concept units are divided equally into N_q categories, with no overlap. Also, the N_q categories are divided equally into three sets: low-, medium- and high-accessible categories. We interpret the accessibility of a category as directly related to the number of incoming connections from other categories. To avoid the effect of connectivity versus overlap between concepts, the overlap between concepts in different categories is set to zero in the current version of the model. All connections are chosen randomly, with the same probability of connecting two concepts in one category, independent of category type, while connections between concepts in different categories have distinct probabilities: lower if the destination concept is from a low-accessible category, and appropriately higher if it is from a medium- or high-accessible category. Also, all connection strengths come from one of two Gaussian distributions, both with the same variance, but one with a larger mean (i.e. strong connections) and another with lower mean (i.e. weak connections). The choice of random connections between concepts was made to serve as a baseline case. The current model is purely abstract, with its structure reflecting the hypothesized structure of semantic representations in the brain. In the future, we plan to simulate the model with more realistic connectivity patterns similar to those observed in semantic networks (Steyvers & Tenenbaum, 2005).

The model simulated here has $N_q = 12$ categories, each with $N_c = 25$ unique concepts. The first four are designated low-accessible, the middle four medium-accessible and last four high-accessible categories. They differ by the probability of an incoming connection from another category: for low-accessible categories, this value is 0.005, for medium-accessible ones, 0.015 and for high-accessible ones, 0.03. The intra-category probability of connection is the same for all categories and is 0.3. For non-zero connections between concepts in the same category, the strong ones (i.e. Gaussian mean 0.6 and variance 0.04) are selected with probability 0.6, while the rest have mean 0.3 and variance 0.04. Non-zero connections between concepts in different categories have a probability of 0.4 to be strong (mean $= 0.6$, variance $= 0.04$), while the rest are weak (mean $= 0.1$, variance 0.04).

The mean values of the coupling parameters for a concept unit, i, are: $a_i = 1$, $b_i = 1$, $c_i = 2$, with a standard deviation of 0.01. The time constants are: $\tau_E = 0.4$, $\tau_I = 2$. The parameters of the sigmoidal activation function are: $\varrho_E = 12$, $\vartheta_E = 0.3$ for excitatory units and $\varrho_I = 18$, $\vartheta_I = 0.65$ for inhibitory units. Other parameters are: $\gamma_E = 0.14$, $\gamma_I = 0.05$, noise $\sim N(0.1, 0.1)$. The model equations were simulated with Runge–Kutta numerical integration using a constant integration step of 0.01.

Ideas were extracted from the activity of the excitatory units ($E_i(t)$) as follows: The active concept units were determined at each time step as those with $E_i(t)$ greater than a threshold (0.7). If the set of active concepts were different from the previous set in one or more bits, it was recorded as an idea.

Simulations with the model showed that low-accessible categories are indeed harder to reach, unless primed. Fig. 8 shows the activity of excitatory units in all 12 categories when the initial state is already in a high-accessible category (top third of the graph). There were four prime presentations placed at equal intervals during the simulation. A prime was encoded as the activation of either 3 (low number of primes) or 6 (high-number of primes) excitatory units chosen randomly from a high-accessible category. It can be seen that bumps of activity are spread mostly over the high-accessible categories, due to their higher-number of incoming connections and the priming of these categories. The activity rarely spreads to the middle and low thirds of the graph corresponding to medium- and low-accessible categories, respectively. Thus, the model reproduces the same behavior as seen in experiments for high- and low-accessible categories: high-accessible categories are sampled more frequently than low-accessible ones.

Fig. 9 illustrates the activity of excitatory units when low-accessible categories are primed. The primes are chosen as in the experiment for Fig. 8, but from low-acessible categories. In contrast to Fig. 8, low-accessible categories are now also activated, and there is an overall increase in the total number of unique ideas. Together, Figs. 8 and 9 simulate all four priming contingencies: high-/low-accessibility categories in combination with low or high number of primes. The control case was also simulated (not shown).

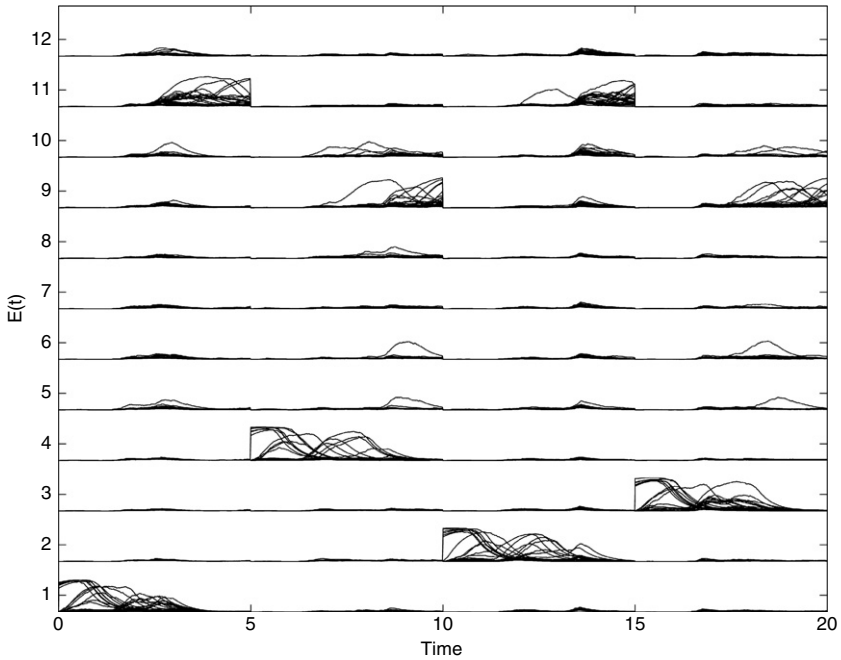

Fig. 9. Modeling the effect of low-accessible primes. The figure shows the activity of excitatory units over time in each category. Plots numbered 1–4 represent low-accessible categories, 5–8 medium-accessible categories and 9–12 high-accessible categories. Concepts from all categories are now activated (see text).

Fig. 10 shows the number of unique or non-repetitive ideas in all five conditions: low/high number of primes, low-/high-accessible categories and the control condition, when no primes where presented. In this last case, the activity starts from a small background noise. Results show that more primes increase the number of unique ideas, independent of the type, and that primes from low-accessible categories are more stimulating than those from high-accessible categories. These results were also predicted by the Brown et al. (1998) associative memory model and are qualitatively consistent with the experimental results shown in Fig. 7 (Leggett, 1997).

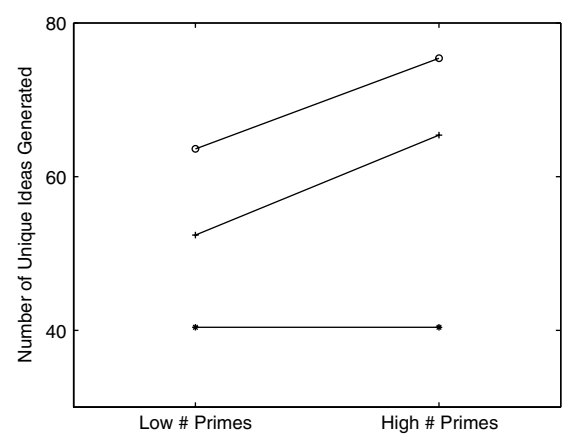

Fig. 10. Number of unique ideas generated by the model under different priming conditions. A low number of primes or a high number of primes are presented four times during the simulation at equal intervals from either low-accessible categories (circles) or high-accessible categories (pluses). The no priming case is shown by the flat curve at the bottom (asterisks). The simulation matches the experimental results in Fig. 7.

5. Biological connections and future directions

We have constructed a neural network model of the creative process and begun to simulate behavioral data on brainstorming. The model provides a framework that is amenable to further extension in several directions. For example, it can be applied to explaining other behavioral data showing beneficial effects of breaks on the number and diversity of ideas generated in a brainstorming session (Paulus et al., 2006). So far, we have drawn some rough analogies between key components of the model and various interconnected brain regions, as shown in Fig. 1. This provisional mapping can be refined through fMRI studies of brainstorming paradigms.

The critic in the network of Fig. 2 rates potential ideas on five different criteria: quality, admissibility, coherence, novelty, and efficiency (Iyer et al., 2009). The last four of these criteria are largely intrinsic properties of the ideas and the tasks for which the ideas are intended. The quality criterion is more subjective and can be influenced by social and affective inputs as well as by cognitive priming. The influences on quality are subserved by inputs to the basal ganglia (or the internal critic) from other brain systems, including the amygdala (emotional salience), hippocampus (task relevance), and prefrontal cortex (plans) (Newman & Grace, 1999). Such influences on this gating can be utilized to extend the model to account for differences in idea generation between individuals acting alone and the same individuals in groups. Social psychologists have long noted that interactive brainstorming groups, in the absence of interventions such as priming or hints, tend to produce fewer ideas than the same individuals acting alone (see Paulus and Brown (2003) for review). This productivity loss in groups has been explained both by cognitive factors such as attentional interference and by emotional factors such as each individual's apprehension about the evaluation of their ideas by others in the group. Attentional interference can be modeled by mutual inhibition between cortical representations of ideas. Evaluation apprehension can be explained by an inhibitory connection from the amygdala to the basal ganglia that reduces the perceived quality of every idea that the brainstormer considers (i.e., makes the internal critic "harsher"), thereby decreasing the probability that any given idea will be articulated. Our network model allows for considerable flexibility, including context and

mood dependence, in the external influences on the critic's quality judgments.

Finally, it is also possible to extend the model to include differences in semantic organization in the left and right hemispheres, which have been hypothesized to play a significant role in creative thinking (Bowden et al., 2005; Duch, 2007; Heilman et al., 2003; Schilling, 2005). We are currently working on such models for both individual and group scenarios.

Acknowledgments

This work was supported by collaborative National Science Foundation Human and Social Dynamics Program grants to Simona Doboli (BCS-0729470), Ali Minai (BCS-0728413) and Paul Paulus (BCS-0729305), which includes support from the Deputy Director of National Intelligence. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Helpful suggestions by Tom Griffiths, Shane Mueller, Jim Houk and Steve Bressler are gratefully acknowledged.

References

Andrews, M., Vigliocco, G., & Vinson, D. Integrating experiential and distributional data to learn semantic representations. *Psychological Review* (in press).

Apicella, P. (2007). Leading tonically active neurons of the striatum from reward detection to context recognition. *Trends in Neuroscience*, *30*, 299–306.

Aston-Jones, C., & Cohen, J. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450.

Bechara, A., Damasio, H., & Damasio, A. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, *10*, 295–307.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Boden, M. (1995). Creativity and unpredictability. *Stanford Humanities Review*, *4*(2).

Botvinick, M., Cohen, J., & Carter, C. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539–546.

Bowden, E., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences*, *9*, 322–328.

Bressler, S., & Kelso, J. (2001). Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, *5*, 26–36.

Brown, V., & Doboli, S. (2006). A neural network simulation of interactive group brainstorming. In *Proceedings of the 18th annual convention of the association for psychological science*.

Brown, V., Tumeo, M., Larey, T., & Paulus, P. (1998). Modeling cognitive interactions during group brainstorming. *Small Group Research*, *29*, 495–526.

Burgess, C., & Lund, K. (1997). Modeling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, *12*, 177–210.

Caramazza, A., & Mahon, B. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, *7*, 354–361.

Carpenter, G., & Grossberg, S. (1987). ART2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, *26*, 4919–4930.

Carpenter, G., Martens, S., & Ogas, O. (2005). Information fusion and hierarchical knowledge discovery by ARTMAP neural networks. *Neural Networks*, *18*, 287–295.

Constantinidis, C., Franowicz, M., & Goldman-Rakic, P. (2001). Coding specificity in cortical microcircuits: A multiple-electrode analysis of primate prefrontal cortex. *Journal of Neuroscience*, *21*, 3646–3655.

Coskun, H., Paulus, P. B., Brown, V., & Sherwood, J. (2000). Cognitive stimulation and problem presentation in idea generation groups. *Group Dynamics: Theory, Research, and Practice*, *4*, 307–329.

Damasio, A. (1989). The brain binds entities and events by multiregional activations from convergence zones. *Neural Computation*, *1*, 123–132.

Damasio, H., Grabowski, T., Tranel, D., Hichwa, R., & Damasio, A. (1996). A neural basis for lexical retrieval. *Nature*, *380*, 499–505.

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*, 179–229.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 14529–14534.

Doboli, S., Brown, V., & Minai, A. (2009). A conceptual neural model of idea generation. In *Proceedings of the 2009 International Joint Conference on Neural Networks* (pp. 723–729).

Doboli, S., & Minai, A. (2005). A neural network model of creative idea generation in groups. In *Proceedings of the ninth international conference on cognitive and neural systems*.

Doboli, S., Minai, A., & Brown, V. (2007). Adaptive dynamic modularity in a connectionist model of context-dependent idea generation. In *Proceedings of the international conference on neural networks* (pp. 2183–2188).

Duch, W. (2007). Intuition, insight, imagination and creativity. *IEEE Computer Intelligence*, 40–52.

Dugosh, K., & Paulus, P. (2005). Cognitive and social comparison processes in brainstorming. *Journal of Experimental Social Psychology*, *41*, 313–320.

Dugosh, K., Paulus, P., Roland, E., & Yang, H.-C. (2000). Cognitive stimulation in brainstorming. *Journal of Personality and Social Psychology*, *79*, 722–735.

Durstewitz, D., Kelc, M., & Güntürkün, (1999). A neurocomputational theory of dopaminergic modulation of working memory functions. *Journal of Neuroscience*, *19*, 2807–2822.

Durstewitz, D., Seamans, J., & Sejnowski, T. (2000). Neurocomputational models of working memory. *Nature Neurosciences Supplement*, *3*, 1184–1191.

Frank, M., & Claus, E. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review*, *113*, 300–326.

Freeman, W. (2000). *How brains make up their minds*. Columbia University Press.

Fuster, J. (2000). The prefrontal cortex of the primate: A synopsis. *Psychobiology*, *28*, 125–131.

Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron*, *14*, 477–485.

Graybiel, A. (1995). Building action repertoires: Memory and learning functions of the basal ganglia. *Current Opinion in Neurobiology*, *5*, 733–741.

Griffiths, T., Steyvers, M., & Tenenbaum, J. (2007). Topics in semantic representations. *Psychological Review*, *114*, 211–244.

Heilman, K., Nadeau, S., & Beversdorf, D. (2003). Creative innovation: Possible brain mechanisms. *Neurocase*, *9*, 369–379.

Houk, J. (2005). Agents of the mind. *Biological Cybernetics*, *92*, 427–437.

Iyer, L., Minai, A., Doboli, S., & Brown, V. (2007). Modularity and self-organized functional architectures in the brain. In *Proceedings of the 7th international conference on complex systems*.

Iyer, L., Minai, A., Doboli, S., & Brown, V. (2008). Modeling exploration and exploitation in creative idea generation. In *Proceedings of the 19th midwest artificial intelligence and cognitive science conference*.

Iyer, L., Minai, A., Doboli, S., Brown, V., & Paulus, P. (2009). Effects of relevant and irrelevant primes on idea generation: A computational model. In *Proceedings of the 2009 International Joint Conference on Neural Networks* (pp. 1380–1387).

Kellenbach, M., Brett, M., & Patterson, K. (2001). Large, colourful or noisy? attribute- and modality-specific activations during retrieval of perceptual attribute knowledge. *Cognitive, Affective and Behavioral Neuroscience*, *1*, 207–221.

Kruschke, J. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Landauer, T., & Dumais, S. (1997). A solution to plato's problems: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.

Leggett, K. (1997). The effectiveness of categorical priming in brainstorming. *Master's thesis*. University of Texas at Arlington.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.

McClelland, J., & Rogers, T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*, 310–322.

McRae, K., de Sa, V., & Seidenberg, M. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, *69*(3), 220–232.

Minai, A., Iyer, L., Padur, D., & Doboli, S. (2009). A dynamic connectionist model of idea generation. In *Proceedings of the 2009 International Joint Conference on Neural Networks* (pp. 2109–2116).

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, *319*, 1543–1546.

Moss, H., Hare, M., Day, P., & Tyler, L. (1994). A distributed memory model of the associative boost in semantic priming. *Connection Science*, *6*, 413–427.

Mountcastle, V. (1997). The columnar organization of the neocortex. *Brain*, *120*, 701–722.

Mueller, S.T., & Shiffrin, R. (2006). Rem-II: A model of the developmental co-evolution of episodic memory and semantic knowledge. In *Proc. international conference on development and learning*.

Newman, J., & Grace, A. (1999). Binding across time: The selective gating of frontal and hippocampal systems modulating working memory and attentional states. *Consciousness and Cognition*, *8*, 196–212.

Nijstad, B., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review*, *3*, 186–213.

Nijstad, B., Stroebe, W., & Lodewijkx, H. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generating task. *Journal of Experimental Social Psychology*, *38*, 535–544.

Osborn, A. (1957). *Applied imagination*. New York: Scribner's.

Patterson, K., Nestor, P., & Rogers, T. (2007). Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987.

Paulus, P., & Brown, V. (2003). Enhancing ideational creativity in groups: Lessons from research on brainstorming. In P. Paulus, & B. Nijstad (Eds.), *Group creativity* (pp. 110–136). New York: Oxford University Press.

Paulus, P., & Dzindolet, M. (1993). Social influence processes in group brainstorming. *Journal of Personality and Social Psychology*, *64*, 575–586.

Paulus, P., Nakui, T., Brown, V., & Putman, V. (2006). Effects of past instructions and brief breaks on brainstorming. *Group Dynamics: Theory, Research and Practice*, *10*, 206–219.

Raaijmakers, J., & Shiffrin, R. (1981). Search of associative memory. *Psychological Review*, *88*.

Schilling, M. (2005). A small-world network model of cognitive insight. *Creativity Research Journal*, *17*, 131–154.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience*, *1*, 199–207.

Singer, W., Engel, A., Kreiter, A., Munk, M., Neuenschwander, S., & Roelfsema, P. (1997). Neuronal assemblies: Necessity, signature and detectability. *Trends in Cognitive Sciences*, *1*, 252–260.

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

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tononi, G., Edelman, G., & Sporns, O. (1998). Complexity and coherency: Integrating information in the brain. *Trends in Cognitive Sciences*, *2*, 474–484.

Tsuda, I. (2001). Towards an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behavioral and Brain Science*, *24*, 793–847.

Verguts, T., Ameel, E., & Storms, G. (2004). Measures of similarity in models of categorization. *Memory & Cognition*, *32*, 379–389.

Vigliocco, G., Vinson, D., Lewis, W., & Garrett, M. (2004). Representing the meaning of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422–488.

Warrington, E., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*, 829–854.

Wilson, H., & Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1–24.

Wu, S., Hamaguchi, K., & Amari, S.-I. (2008). Dynamics and computation of continuous attractors. *Neural Computation*, *20*, 994–1025.