

Neurocognitive Spotlights: Configuring Domains for Ideation

Laxmi R. Iyer, Vaidehi Venkatesan and Ali A. Minai, *Senior Member, IEEE*

Abstract— Creativity is an important attribute of the human mind, and shows itself in all aspects of its function. However, its neural basis remains poorly understood. In this paper, we explore two issues with regard to creativity in the semantic domain: 1) What neural mechanism enable the brain to construct context-specific semantic spaces to facilitate the generation of relevant ideas? and 2) Can these mechanisms support greater creativity simply by exploring unusual semantic spaces? We use a variant of our previously developed neural model of ideation to show that a dynamical modular neural system can, indeed, learn to configure context-appropriate semantic domains based on experience, and that exploratory dynamics within this system can lead to the unmasking of novel emergent ideas.

I. INTRODUCTION

Most research on mental function has focused – for understandable reasons – on response-oriented processes, i.e., processes where specific responses can be defined, observed and quantified. These include object or face recognition, memory, attention, problem-solving, motor control, etc. However, this excludes many processes and functions that are most closely identified with higher cognition, including the processes of open-ended thought, idea generation, and creativity. A large part of the difficulty here lies in devising laboratory experiments that can, in some objective fashion, provide unambiguous data for these processes. As a result, processes such as creativity have been studied mainly by social and cognitive psychologists using behavioral experiments and human evaluation of data [1], [2], [3], [4], [5]. These experiments have produced a wealth of observations and several theories about the basis of creativity [6], [7], [8], [3], [5]. However, as the neural mechanisms underlying the more response-oriented functions have become better understood, it has become possible to infer similar mechanisms for internal processes such as ideation, and a number of computational models have been proposed [9], [10], [11], [12], including one by our research group [13], [14], [15], [16], [17], [18]. These models have been able to explain, albeit qualitatively, the phenomena observed during behavioral experiments in individual and group ideation (brainstorming), and have begun to suggest possible neural mechanisms for the ideation process.

In this paper, we consider two related issues: 1) Possible neural mechanisms that allow the search for ideas to focus

within a context-specific semantic space; and 2) Showing through computational modeling that searching in unusual semantic spaces can lead to the unmasking of emergent novel ideas.

II. BACKGROUND AND MOTIVATION

The model of mental function that we use is based on the view that mental responses such as percepts, thoughts and actions emerge from the interaction of synergistic dynamics between the processes of response generation, internal (cognitive) attention, and modulatory control driven by stimulus and evaluative feedback. Metastable coordinated patterns of activity – *synergies* – arise as the three processes interact, giving rise to perception, thought and action [19]. The metastability of the synergies can arise as a result of processes such as slow potassium currents, refractoriness, short-term synaptic plasticity [20], [21] and inhibitory modulation by larger-scale brain rhythms. This model is meant to be an abstraction, though it has been suggested that some brain regions may have specific roles in this regard [22], [23], [24]. The model is similar in spirit to the *agents of mind* model proposed by Houk [24], and to models by Grossberg and colleagues [25], [26], [27]. It is also related closely to coherence-based theories of cognition [28], [29], [30], [31], [32].

In this paper, our focus is on the problem of generating (possibly creative) ideas from the brain’s semantic substrate. Ideas are defined as structured combinations of concepts (see discussion below), but here we ignore the structural aspect, i.e., syntax and grammar, which are complicated issues in their own right, and look at ideas simply as combinations of simpler semantic elements termed *concepts*. In particular, we address two questions: 1: Since ideas are typically generated within a specific task or problem context, and the process can go on for an extended period, how can cognitive attention be kept focused on context-relevant information during the search without loss of flexibility? 2. How does variation in the pattern of cognitive attention affect the emergence of novel ideas during the ideation process?

III. CONCEPTUAL FORMULATION

Before describing our computational model, we look at some theoretical issues underlying our model of ideation and creativity, and describe the model in abstract conceptual terms.

A. Selective Internal Attention

Creativity can come in many forms, but all creative processes within an individual emerge from the substrate of knowledge encoded in that individual’s brain – possibly

Laxmi Iyer is with the Department of Computer Science, University of Cincinnati, Cincinnati OH 45221, Email: iyerlr@email.uc.edu, Vaidehi Venkatesan is with the Department of Computer Science, University of Cincinnati, Cincinnati, OH 45221-0030 email: venkatvi@email.uc.edu). Ali Minai is with the Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati OH 45221, Email: Ali.Minai@uc.edu

Acknowledgement: This work was supported in part by a National Science Foundation Human and Social Dynamics Program grant to Ali Minai (BCS-0728413), which includes support from the Deputy Director of National Intelligence, and by a National Science Foundation CreativeIT grant to Ali Minai (IIS-0855714)

in response to an external cue or an internal thought. As the brain's natural dynamics triggers synergistic patterns of activity within different regions, the effects of these patterns are manifested as percepts, thoughts and action, depending on the brain region involved. From this, it may be hypothesized that all these processes are equally capable of supporting creativity, and that creativity in all these cases is fundamentally similar. Viewing the brain as a multi-scale modular network, we suggest that creativity lies mainly in discovering novel ways of combining coordinated patterns of activity inherent in these modules into new higher-level synergies. In the case of semantic memory, this can take the form of combining concepts into ideas, whereas in the motor system, it corresponds to combining movements into actions. In a broad sense, creativity can be seen as a novel *response* to external and/or internal cues in a given context such as playing a board game, brainstorming on a problem, or composing a poem. An essential requirement in all these cases is that the search for novel combinations occur in a *domain* comprising a repertoire of knowledge and methods that makes sense within the given context. For example, creativity in coming up with culinary recipes should occur in the domain of food ingredients, and not in that of hardware or sports strategies. Thus, selectively configuring a context-appropriate response repertoire from the brain's representational substrate is an important aspect of creativity. This process must also have several other features, including the following:

- 1) The configuration of the domain for creativity must occur naturally through the dynamics of the system.
- 2) The dynamics must also provide mechanisms for controlling the scope of the domain, ranging from a very narrow focus suitable for recall of obvious responses to a broader one more amenable to creative responses.
- 3) Once it triggers a domain, the dynamics must be able to keep it active for an extended period, since many creative processes occur over long durations.
- 4) The process should generalize, so that novel contexts can trigger reasonably useful domains based on their similarity to known contexts.

The configuration of domain-specific response repertoires for a task or problem context is a general problem, but has not been addressed much in the literature – primarily because it is difficult to study experimentally. Most work on task-specific processing has focused on cognitive control [33], [34] and working memory (WM) [35], [36]. WM studies mainly consider maintenance of task-relevant information over periods of a few seconds at most [37], and it is well known that such *short-term working memory (ST-WM)* has extremely limited capacity [38]. Clearly, ST-WM clearly cannot be the whole story. Many real-world tasks require large amounts of domain-specific information over extended periods while a non-trivial behavior – such as a search for ideas – occurs. This is true in both narrow domains such as playing a game of checkers and in many everyday situations, such as behaving appropriately in a social setting, imagining

an extended episode, or solving a problem using previous experience. Theories such as long-term working memory [39] and global workspaces [40], [29] have been proposed to address this issue.

It has also been suggested (e.g., [9], [12]) that the novelty needed for creativity might arise – at least in part – from searching in *atypical domains*, which can lead to unusual combinations. This approach to creativity has been exploited widely by artists, notably in the surrealist movement. While accessing such atypical domains cannot be easy – as shown by the rarity of creativity – the dynamics of the system must allow for it. There is also some neurobiological evidence that the right hemisphere of the brain is involved in the creative process [41], [42].

IV. IDEAS AS CONCEPTUAL COMBINATIONS

The nature of ideas has long been an issue of interest to philosophers and scientists, who are primarily in the business of generating new ones [43], [44]. A recurring theme in this regard is to see ideas as *combinations* of existing ideas or concepts. For example, Simonton [7], quotes Poincare [45]: “Ideas arose in crowds; I felt them collide until pairs interlocked, so to speak, making a stable combination..... [T]he only combinations that have a chance of forming are those where at least one of the elements is one of those atoms freely chosen by our will. Now, it is evidently among these that is found what I called the good combination..... [A]mong the great numbers of combinations blindly formed ... almost all are without interest and without utility.” Simonton [7] (via Hadamard [43]) also quotes Einstein”: “... combinatory play seems to be the essential feature in productive thought”

From these quotes, one can infer three extremely important insights:

- Ideas are formed through combination of existing ideas.
- These combinations are triggered associatively by prior ideas.
- Most combinations explored in this process are not useful; good combinations arise rarely.

This led to the theory that creative ideas represent “remote associates”, i.e., unusual combinations of existing ideas, and the hypothesis that creative individuals had access to more unusual associations [46]. To describe the process by which ideas are generated, Campbell [6] proposed an evolution-like process of random variation and selective retention. Models along similar lines have recently been proposed by Simonton [7] and Duch [12]. An important feature of the search process in the space of combinations is associative triggering, i.e., currently active ideas activate others with which they are associated in memory. Models of ideation based on this have been proposed by Brown and Paulus [8] and Nijstad and Stroebe [5]. Recently, we have developed a larger model in which associative search plays a crucial role [13], [14], [17], [16], [18].

As discussed in the previous section, we postulate that ideas, like all mental responses, correspond to synergistic

patterns of activity in specific cortical areas – mainly the temporal lobe, where most semantic information seems to be represented [47], [48], [49]. We model this through a connectionist network whose nodes encode concepts. The nodes are abstractions for cortical cell assemblies, which are known to encode features of specific objects [50], [51]. The nodes are organized into a semantic network, with the link between two nodes representing the joint utility of the two concepts involved. This can be seen as the result of Hebbian learning based on previously experienced ideas. Thus, strongly connected groups of nodes in this network correspond to the ideas inherent in the system, and these are unmasked by the system’s dynamics during search as described below.

A very important feature of this representation is that the process of embedding ideas in the semantic network also creates novel ideas *emergently* through indirect correlations, and these too can be activated by the search process. We propose that such emergent ideas are a major manifestation of creativity, and processes that support their emergence are its prime enablers. We explore this by looking at the emergence of ideas under different patterns of cognitive focus – both familiar and unfamiliar.

V. OVERVIEW OF THE COMPUTATIONAL MODEL

Figure 1 shows the overall architecture of the computational model, comprising two major sub-systems: a *concept network (CN)*, and a *dynamic selector network (DSN)*. The concept network represents semantic memory as a connectionist semantic network, where nodes represent *concepts* and connections between these nodes encode the *compatibility* among concept pairs based on the prior experience of the system with ideas involving the two concepts. Thus, the CN represents the context-free semantic knowledge about inter-concept relationships that is available to the brain. Under normal conditions, the CN is largely inhibited, and there is no activity (possibly except small fluctuations due to noise). However, when all or part of the CN is disinhibited, conceptual combinations arise spontaneously through the dynamics of the system. The DSN is another recurrent neural system with a modular structure and embedded synergies (attractors) configured through a learning process. The DSN projects a bias, or *pattern of disinhibition*, onto the CN to configure a context-specific semantic space. Units in the DSN redundantly encode groups (modules) of concepts that share certain features. These are termed *categories*.

Given the concept network as the base, we hypothesize the following:

1. Context-specific semantic networks are induced in the CN by an internal attention process that projects a context-specific pattern of disinhibition onto the CN. This pattern selectively disinhibits a subset of nodes in the CN, leaving them free to interact through their existing connections, while the rest of the nodes remain inhibited. Thus, the DSN can be seen as projecting a context-specific “spotlight” onto the CN, configuring a context-specific semantic network termed

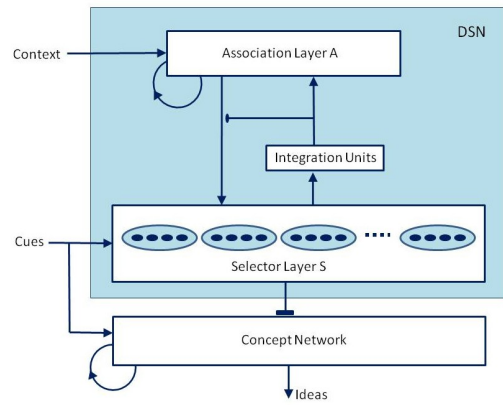


Fig. 1. Architecture of the computational model.

a *domain*, within which context-appropriate ideas can arise through a dynamic *search* process.

2. The function of internal attention is mediated by a modular, recurrent neural system – the dynamic selector network (DSN) – capable of sustaining patterns of activity over significant lengths of time. These patterns, which represent synergies (attractors) in the DSN, provide the gating bias to the CN.

3. The DSN attractors, called *graded attractors*, are highly flexible, with an adjustable level of activity. Each graded attractor is associated with a specific context, and is triggered when the system finds itself in that context. The result is the induction of a semantic network tuned to the context, within which the search for ideas can proceed.

4. The search process emerges from the natural dynamics of the CN under conditions of lowered inhibition. The network is always subject to competitive lateral inhibition, so that only an approximately constant number of units can be active at a time. This generates an *itinerant sequence* of activity across the CN (see Figure 2) where groups of concept nodes are co-activated. These conceptual combinations represent potential ideas, but are recognized as such only if they persist for a sufficient duration because of mutual compatibility between the activated concepts (see below). Otherwise, they are just considered transients.

5. When a context is presented to the system, only the concept nodes within that context’s domain (as encoded in the associated graded attractor) are available for activity, and the dynamics of conceptual combinations remains confined to this space – thus generating ideas that are relevant to the context. If the entire system is disinhibited, a context-free search results, generating ideas that “make sense” in general, including potentially novel or creative ideas.

6. Graded attractors can be modulated to limit or expand the number of active units, thus controlling the “width” of

the cognitive spotlight on the CN (see Figure 2). In more comprehensive versions of the model, graded attractors can also be switched and merged, and reshaped through long-term potentiation (LTP) mediated by dopamine-based reward signals [18]. However, we do not focus on these attributes, and use a simpler system to demonstrate the basic idea of a flexible cognitive spotlight to configure context-specific domains, and to study the role of this spotlight in finding novel combinations.

It should be noted that there are three separate inhibitory processes at work in the system: 1) Lateral inhibition in the CN, which keeps the number of co-active units small; 2) Lateral inhibition in the DSN, which permits only a limited number of DSN projection units to be co-active; and 3) The default top-down inhibition on the CN nodes, which is partially and context-specifically removed by the appearance of a graded attractor in the DSN.

The two sub-systems in the model have very different dynamics, though both are recurrent networks. The CN’s dynamics is characterized by flow [52], where patterns form and disintegrate on their own, with some persisting longer than others due to their internal cohesion. This is the dynamics of a process defining useful change, which is the essential attribute of all behavior – both thought and action. The DSN’s dynamics, in contrast, is characterized by extended periods of relative stability where a specific coordinated activity pattern – a graded attractor – is active, albeit modulated by feedback, punctuated by rapid switches to other stable patterns. Thus, the two systems represent two different time-scales – inner and outer cognitive loops, so to speak – each capturing an important aspect of the dynamics of cognition.

From this formulation, it is clear that ideas correspond to well-connected groups of nodes in the CN. Some of these are explicitly generated by the learning process, while others form emergently. The itinerant dynamics of the CN then unmask these as a result of associative triggering by other concepts. Of course, only ideas involving concepts within the subset disinhibited by the DSN can emerge. Thus, different patterns of disinhibition – different spotlight configurations – so to speak, unmask diverse sets of ideational possibilities, or *semantic affordances*, to borrow a useful idea from ecological cognition [53].

VI. DESCRIPTION OF THE MODEL

A. Semantic Data

To simulate the results of learning, the system is configured using a semantic database comprising the following components: 1) A *concept set*, $C = \{c^k\}$; 2) A *category set*, $Q = \{q^j\}$, where each category, q^j , is associated with a subset of concepts, $C_i \subset C$, which is termed the *member set* for category c^k ; 3) An *idea set*, $X = \{x^i\}$, where each idea, x^i , comprises a set of concepts drawn from C , i.e., $x^i = [c^{i_1}, \dots, c^{i_m}]$, where m is the *length* of the idea; and

4) A *context set*, $U = \{u^l\}$, where context u^l is associated with a subset of ideas, $X_l \subset X$, which is called the *sample set* for that context. We assume that the member sets for different categories may overlap, but the sample sets for different contexts do not.

As an example, some of the simulations reported below use a semantic data set derived from more than 5000 on-line culinary recipes. In this set, concepts correspond to ingredients such as “salt”, “olive oil” or “tomatoes”; categories define types of ingredients, e.g., “sweetener” or “green vegetable”; ideas correspond to the list of ingredients in a recipe; and contexts denote the type of recipe, e.g., “soup”, “appetizer”, “entree”, etc. From a much larger data set collected by web crawling and post-processing, we obtained a set of 489 concepts (ingredients), with 171 categories and 9 contexts. We used 100 ideas (ingredient lists) from each context, giving a total of 900 distinct ideas. In this case, the categories were assigned manually, though this can also be done through clustering. An informal analysis of the contexts indicates that they overlap significantly in their use of categories.

B. Inputs

The system has two different types of inputs: 1) Context input; and 2) Cues. The context input provide the system with information on the task, problem or situational context in which it finds itself. This information may be available persistently or transiently. For example, there may be a persistent reminder, such as a signboard, indicating the context explicitly, or a transient indicator, such as instructions before a brainstorming session. A key function for the DSN is to keep track of transiently indicated contexts after the indicator is gone, but retain flexibility at the same time.

The *Context Input*, G comes from of N_G *context feature units* (henceforth called *G-units*), and is denoted as $\{g_1, g_2, \dots, g_{N_G}\}$. The context layer has feedforward connections to the Association Layer (see below). Each context is represented by a real-valued vector of length N_G . Each unit k in the vector has a value between 0 and 1, encoding the degree to which the corresponding feature is present in the context.

The *Cue Input*, directly stimulates specific units in the concept network and the DSN, thus “suggesting” an idea comprising these concepts or categories. In the current simulations, only the input to the DSN is utilized to clamp cognitive attention (see the next section).

C. The Dynamic Selector Network

The DSN in the current model has two layers, an *association layer*, A , and a *selector layer*, S .

The *Selector Layer* (S) consists of N_S *selector units*, or *S-units*, organized into N_m *modules*. Each S-unit, i , is tuned to a particular category, q^{k_i} , and projects a *disinhibiting bias* onto all concept units in the member set of the category. This *projective field* for i is denoted as Γ_i . There are numerous S-units for each category, but each module contains at most one unit tuned to a given category. Each module, M^k , has units

tuned to different categories, so that the collection of units in M^k represents a category combination or *idea template*. In our general model, the units within a module are connected densely, while there are also sparse recurrent connections between modules. However, these recurrent connections are excluded from the simulations reported here. Each S-unit is driven by excitatory input from the A layer. Each module also has an *integration unit* whose activity at time t is proportional to the net activity of all units in the module. Projection units send recurrent connection to a sparse set of units in the A layer.

In the current system, modules are constructed by a heuristic clustering method that preferentially generates modules with category combinations seen frequently in the semantic data. This set of modules is then augmented by more modules generated through recombining the initial set, resulting in a very diverse population of broadly tuned modules. Thus, at the module level, S can be seen as a network of coarse-grained feature detectors in idea space.

The S layer is subject to competitive inhibition, so that only the $K^S(t)$ most activated non-refractory units can fire at time t . This inhibition is modulated by a signal from an evaluative system [18], so that $K^S(t)$ can be varied based on the performance of the system. Here, we implement the evaluating system as an external input to vary $K^S(t)$.

The *association layer*, A , consists of N_A *association units*, or *A-units*. It is stimulated by the context input, G , and sends feedforward connections to S . The units of the A are also recurrently connected, and embed a repertoire of fixed point attractors. The A system acts as a flexible map or *hidden layer* between the G input and the S layer, providing the opportunity for each component of the context input to link with any specific selector layer module. Maps of this sort are widely hypothesized to exist in the cortex [54].

The net synaptic input into the AN unit i at time t is given by:

$$y_i(t) = \gamma^{AG} \sum_{j \in G} w_{ij}^{AG} g_j(t) + \gamma^A \sum_{j \in A} w_{ij}^A z_j^A(t-1) \quad (1)$$

where y_i is the input into A-unit i , γ^{AG} is the gain from the stimulus layer to A , w_{ij}^{AG} is the weight from component j of the input G to A unit i , g_j is the output of G unit j , γ^A is the recurrent gain for A , w_{ij}^A is the weight from A-unit j to A-unit i , and z_j^A is the output of the A-unit j .

The state of A-unit i at time t is given by:

$$u_i(t) = (1 - \alpha^A) u_i(t-1) + \alpha^A y_i(t) \quad (2)$$

where $0 < \alpha^A \ll 1$ controls the update rate of the activity.

Activity in A is soft-competitive and approximately K^A units receiving the highest input are activated

The gain γ^{AG} is initially set to 1, but decays slowly as the episode proceeds. This represents the gradual removal of explicit context cues, requiring that the context be maintained internally by the system's dynamics.

During training, patterns in G and S jointly induce activity patterns in A , which are learned incrementally via autoassociative Hebbian learning. Over time, this results in the configuration of context-specific fixed point attractors in A . Contexts that involve ideas with very similar category combinations end up sharing an attractor, while contexts that require appreciably different category combinations induce separate attractors. Thus, attractors in layer A become the system's internal representations of the semantic domains that it must actually discriminate between. The attractors are also hetero-associated, via Hebbian learning in the A -to- S synapses, with the layer S activity patterns that give rise to them. Thus, if a given context can turn on the correct attractor in A , it, in turn, activates the correct activity pattern in S .

The system has one further notable feature: The synaptic plasticity in the A -to- S synapses is modulated by the net activity of the S Layer modules involved via the projection units. Connection to S-units in highly active modules are strengthened disproportionately, while those to S-units in less active modules are changed very little. As a result, the attractor in A develops a *graded projection* onto S , with the strongest weights to the modules most active during the presentation of the training idea. Since activity in S is competitive with modulatable inhibition, the S -layer activity pattern triggered by A in response to context G^r can be turned on to varying degree. When competitive inhibition in S is high, only the few most excited non-refractory units can turn on – and, due to the graded projection developed during learning, these are in precisely those modules most relevant to the current context (as represented by the attractor). As inhibition is lowered, more units turn on roughly in order of their relevance until, at low inhibition, all S-units excited by the attractor in A can turn on. This, of course, projects a *variable* pattern of disinhibition on the CN, ranging from a very narrow search space at high inhibition to a much broader one when the inhibition is low. Because of this flexible functionality, the attractors implemented in the DSN are termed *graded attractors* [17]. A key requirement for graded attractors is that the *order* in which the disinhibition targets new categories in the CN should correspond to the pattern of participation for categories within the given context. Categories with the highest participation should be disinhibited first, and the process should proceed gradually through categories with decreasing context-dependent utility. The results given below indicate that this does indeed happen.

D. The Concept Network

The concept network consists of N_C recurrently connected *concept units* or C-units, each representing a concept. The concept units can be seen as amodal representations of basic semantic elements (e.g., words). Each concept unit, i , receives input from two sources: 1) Excitatory input from other concept units, j , via recurrent weights, w_{ij}^{cc} ; and 2) A gating input, $\zeta_j = \bigvee_{j|i \in \Gamma_j} z_j^S(t)$ from all S-units, j , tuned to categories that concept i is a member of. The net synaptic input to C-unit i is given by:

$$y_i^C(t) = \zeta_j(t) \left[\gamma^C \sum_{j \in \text{CN}} w_{ij}^C z_j^C(t-1) + \gamma_{noise}^C \eta_i(t) \right] \quad (3)$$

where z_j^C is the output of C-unit j , w_{ij}^C are weights from other concept units, j , $\eta_i(t)$ is uniform white noise between 0 and 1, and γ^C , and γ_{noise}^C are fixed gain parameters.

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

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

where $0 < \alpha^C \ll 1$ is the update rate for i , and is set so that it is significantly lower than α^S , the update rate for the S layer. This leads to much faster dynamics in the CN.

Activity in the CN is soft-competitive, so that the K_C non-refractory units with the highest $u_i^C(t) > 0$ can fire at time t . The ability of units to fire depends on their recent activity, and a unit that has fired repetitively over a short interval becomes refractory for a time (see [18] for details). Additionally, the recurrent synapses in the CN also show *short-term synaptic plasticity* [20], [21], becoming temporarily less potent after persistent presynaptic activity.

The connectivity between pairs of C-units is set by Hebbian learning based on the co-occurrence of the corresponding concepts in the ideas included in the data set. Thus, concepts that co-occur frequently (independent of context) become more strongly connected.

Every conceptual combination represented by the set of K_C co-active C-units represent a potential idea, but is recognized consciously as such only if the group remains co-active continuously for τ_A time-steps, which is termed the *awareness threshold*. Since this is possible only when the C-units involved have strong mutual connectivity, which depends on their co-occurrence in known ideas, most ideas that reach the awareness threshold “make sense”. However, even these ideas are eventually destabilized by refractoriness of the C-units and the short-term depression of the synapses involved. In contrast, conceptual combinations that are insufficiently compatible never reach the awareness threshold at all. The result is an *itinerant dynamics* [52], with periods of transient activity punctuated by metastable equilibria representing ideas.

VII. SIMULATION AND RESULTS

As discussed above, we look at two issues in our simulations: The configuration of context-specific domains by the DSN, and the dynamics of ideation in familiar and unusual domains. Each of these is discussed below.

A. Domain Configuration by the DSN

To investigate the first issue systematically, we simulated a model of the DSN separately (i.e., without the CN) using an artificially constructed abstract dataset involving only contexts and categories. We define ten contexts, G^k , $k = 1, \dots, 10$, each with a *characteristic category distribution*, $\mu^k = [\mu_1^k \ \mu_2^k \ \dots \ \mu_{N_q}^k]$ over the set Q , where μ_j^k is the

probability that an idea in context G^k includes at least one concept from category Q^j . The CCDs for the contexts were defined such that contexts 1 through 4 and 5 through 7 form two distinct clusters and the other three are dissimilar. Ideally, the system should learn five graded attractors – one for each cluster – with some internal variation within each attractor based on the differences between the CCDs for the contexts within the cluster.

The system was trained using abstract ideas drawn randomly from the CCDs for each context. Each idea was represented as binary vector over the category set, Q , such that each idea had participation from 6-8 categories. The average pattern of category participation for each context was close to its CCD. After configuration, the network was probed by each of the 10 contexts and allowed to relax to the recalled graded attractor. The allowed level of activity in the S Layer, K_S , was then varied over the range 5 through 50, and the categories set of disinhibited categories at each level was obtained. This response was evaluated by two metrics:

1. **Average Attractor Quality**, A^k for context G^k measures whether the order in which categories are disinhibited as inhibition is lowered corresponds to the relative values of the categories in their context’s CCD. $A^k = 1$ indicates that categories are disinhibited in the order of their significance.

2. **Total Coverage**, B^k , for context G^k measures the degree to which the categories disinhibited at the lowest inhibitory level match those with significant value in the context’s CCD.

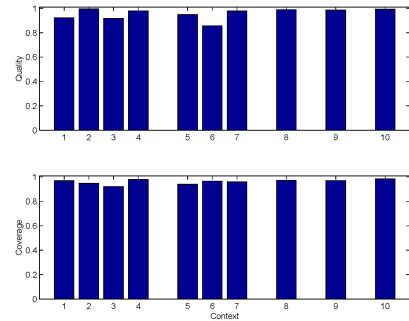


Fig. 2. Top: Quality for the learned graded attractors; Bottom: Coverage for the learned attractors.

Figure 2 shows the quality and coverage values obtained for the graded attractors learned for the 10 contexts. Clearly, appropriate graded attractors have been configured in all cases. Figure 3 shows a comparison of the CCD for each context with the category disinhibition pattern produced by the graded attractor for that context, averaged across all activity levels, K_S . The results show that the system indeed learned 5 graded attractors and, to a large degree, the correct pattern of disinhibition within each context.

B. Generating Novel Ideas in Novel Contexts

To explore the possibility for creative ideation, we applied the full model (DSN and CN) to the recipes dataset described earlier. The contexts for the dataset (except one) had high

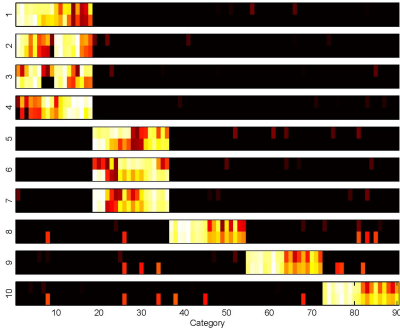


Fig. 3. Top: Category coverage profiles for each context. In each strip, the top half shows the CCD for a context and the lower half shows the category disinhibition pattern averaged over all values of K_S .

overlap in category space, presenting a significant challenge. However, the system was still able to learn reasonably good graded attractors for each context, though the anomalous context (context 5 - Drinks) posed problems. We see this last as mainly a small-size effect.

After configuring the system based on 900 ideas (ingredient lists) – 100 for each context, we tested the system under two protocols:

Protocol 1: Familiar Context Clamp: In this case, the system was stimulated with one of the contexts it was trained on and the ideas generated were analyzed for novelty. The awareness threshold for these simulations was set at 4 time-steps. To evaluate novelty, each generated idea, X , was compared with ideas in the training dataset, and *familiarity*, $\phi(X)$, was measured as the largest fraction of ingredients in the idea that occurred within any single idea in the dataset. Novelty $\nu(X)$, was then measured as $\nu(X) = 1 - \phi(X)$.

Protocol 2: Novel Context Clamp: In this protocol, an unusual combination of selector units was clamped on by a stable external cue to simulate the effect of “thinking outside the box”. No context input was applied. The novelty of ideas was evaluated as in Protocol 1. There has been significant interest within the ideation research community in studying brainstorming in novel contexts [1], [2], [3], [4], [5], and this experiment is a simple attempt to model this process.

In both cases, it is not sufficient to consider whether the ideas were novel, but also whether they were coherent. To get a graphical and numerical estimate of coherence for ideas, we use the following device. Since each generated idea is constrained to have a fixed number of ingredients, K , it can be visualized as a $K \times K$ matrix whose ij th entry is the synaptic weights, w_{ij}^C . Thus, highly coherent ideas, where all ingredients “support” each other, have large values in most cells (showing as light colors), while less coherent ideas have a sparser image. The sum of the weights in the idea image is used as a numerical measure of coherence. It should be noted that the *actual* weights used during idea generation are modulated by synaptic dynamics, and that some concept nodes occasionally become active due to noise or because most eligible C-units are temporarily refractory. These occasionally produce non-sensical ideas, but most

ideas generated do show interesting patterns of coherence.

Previous work in our group [55], [16], [18] has considered the effects of priming familiar or unusual categories and concepts on the generation of ideas. Protocol 2 (above) extends this by focusing cognitive attention on unusual semantic domains. Figure 4 shows a plot of the novelty values of unique ideas generated against their coherence values for both protocols. Each protocol was run for the same number of trials and duration per trial. Figure 5 shows a sampling of ideas images generated during both protocols.

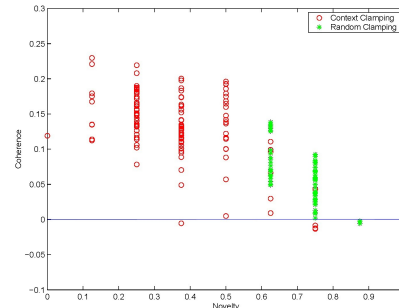


Fig. 4. Novelty vs. Coherence plot for ideas generated under Protocol 1 (red) and Protocol 2 (green).

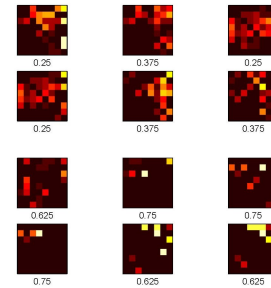


Fig. 5. Top set: Samples of ideas generated under Protocol 1. Bottom set: Samples of ideas generated under Protocol 2. The numerical labels denote novelty values for the ideas.

Comparing the results obtained from the two protocols show that: 1) The number of unique ideas is much higher in a familiar domain than in an unfamiliar one; 2) Ideas generated in the familiar domains span a broad range of novelty but a very large fraction have low to moderate novelty, while ideas generated in the unusual domains are all in the high novelty range; 3) Almost all ideas generated in the familiar and unfamiliar domains have positive coherence, but those in the former tend to have higher coherence values. On the last point, it is particularly notable that both cases lead to the generation of novel ideas, but the coherence values are comparable in the two cases. These data demonstrate both the advantages of familiarity obtained through prior learning and the potential for generating coherent emergent ideas in familiar and unfamiliar situations.

ACKNOWLEDGMENT

The authors would like to thank Simona Doboli, Vince Brown, Dan Levine and Paul Paulus for useful ideas.

REFERENCES

- [1] P. Paulus and M. Dzindolet, "Social influence processes in group brainstorming," *Journal of Personality and Social Psychology*, vol. 64, pp. 575–586, 1993.
- [2] H. Coskun, P. Paulus, V. Brown, and J. Sherwood, "Cognitive stimulation and problem presentation in idea generation groups," *Group Dynamics: Theory, Research, and Practice*, vol. 4, pp. 307–329, 2000.
- [3] P. Paulus and V. Brown, "Enhancing ideational creativity in groups: Lessons from research on brainstorming," in *Group Creativity*, P. Paulus and B. Nijstad, Eds. New York: Oxford University Press, 2003, pp. 110–136.
- [4] K. Dugosh and P. Paulus, "Cognitive and social comparison processes in brainstorming," *Journal of Experimental Social Psychology*, vol. 41, pp. 313–320, 2005.
- [5] B. Nijstad and W. Stroebe, "How the group affects the mind: A cognitive model of idea generation in groups," *Personality and Social Psychology Review*, vol. 3, pp. 186–213, 2006.
- [6] D. Campbell, "Blind variation and selective retention in creative thought as in other knowledge processes," *Psychol. Rev.*, vol. 67, pp. 380–400, 1960.
- [7] D. Simonton, "Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives," *Psychol. Bull.*, vol. 129, pp. 475–494, 2003.
- [8] V. Brown, M. Tumeo, T. Larey, and P. Paulus, "Modeling cognitive interactions during group brainstorming," *Small Group Research*, vol. 29, pp. 495–526, 1998.
- [9] M. Schilling, "A small-world network model of cognitive insight," *Creativity Res. J.*, vol. 17, pp. 131–154, 2005.
- [10] S. Diboldi and A. Minai, "A neural network model of creative idea generation in groups," in *Proceedings of the Ninth International Conference on Cognitive and Neural Systems*, Boston, 2005.
- [11] V. Brown and S. Diboldi, "A neural network simulation of interactive group brainstorming," in *Proceedings of the 18th Annual Convention of the Association for Psychological Science*, New York, 2006.
- [12] W. Duch, "Intuition, insight, imagination and creativity," *IEEE Comput. Intell.*, pp. 40–52, 2007.
- [13] S. Diboldi, A. Minai, and V. Brown, "Adaptive dynamic modularity in a connectionist model of context-dependent idea generation," in *Proceedings of the International Conference on Neural Networks, Orlando, Florida*, 2007, pp. 2183–2188.
- [14] L. Iyer, A. Minai, S. Diboldi, and V. Brown, "Modularity and self-organized functional architectures in the brain," in *Proceedings of the 7th International Conference on Complex Systems*, Boston, MA, 2007.
- [15] —, "Modeling exploration and exploitation in creative idea generation," in *In Proceedings of the 19th Midwest Artificial Intelligence and Cognitive Science Conference*, 2008.
- [16] L. Iyer, A. Minai, S. Diboldi, V. Brown, and P. Paulus, "Effects of relevant and irrelevant primes on idea generation: A computational model," in *Proceedings of IJCNN 2009*, 2009.
- [17] A. Minai, L. Iyer, D. Padur, and S. Diboldi, "A dynamic connectionist model of idea generation," in *Proceedings of IJCNN 2009*, 2009.
- [18] L. Iyer, S. Diboldi, A. Minai, V. Brown, D. Levine, and P. Paulus, "Neural dynamics of idea generation and the effects of priming," *Neural Networks*, vol. 22, pp. 674–686, 2009.
- [19] J. Kelso, *Synergies: atoms of brain and behavior*. Springer Verlag, 2009, p. 2007.
- [20] R. Zucker and W. Regehr, "Short-term synaptic plasticity," *Annual Review of Physiology*, vol. 64, pp. 355–405, 2002.
- [21] L. Abbott and W. Regehr, "Synaptic computation," *Nature*, vol. 431, pp. 796–803, 2004.
- [22] A. Graybiel, "Building action repertoires: memory and learning functions of the basal ganglia," *Current Opinion in Neurobiology*, vol. 5, pp. 733–741, 1995.
- [23] J. Fuster, *Cortex and Mind: Unifying Cognition*. Oxford University Press, 2003.
- [24] J. Houk, "Agents of the mind," *Biol. Cybern.*, vol. 92, pp. 427–437, 2005.
- [25] D. Bullock and S. Grossberg, "Neural dynamics of planned arm movements: emergent invariants and speed-accuracy properties during trajectory formation," *Psychological Review*, vol. 95, pp. 49–90, 1988.
- [26] G. Carpenter, S. Martens, and O. Ogas, "Information fusion and hierarchical knowledge discovery by ARTMAP neural networks," *Neural Networks*, vol. 18, pp. 287–295, 2005.
- [27] S. Grossberg and L. Pearson, "Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: toward a unified theory of how the cerebral cortex works," *Psychological Review*, vol. 115, pp. 677–732, 2008.
- [28] B. Baars and S. Franklin, "How conscious experience and working memory interact," *Trends in Cognitive Sciences*, vol. 7, pp. 166–172, 2003.
- [29] S. Dehaene and L. Naccache, "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework," *Cognition*, vol. 79, pp. 1–37, 2001.
- [30] G. Tononi, G. Edelman, and O. Sporns, "Complexity and coherency: Integrating information in the brain," *Trends in Cognitive Sciences*, vol. 2, pp. 474–484, 1998.
- [31] S. Bressler and J. Kelso, "Cortical coordination dynamics and cognition," *Trends in Cognitive Sciences*, vol. 5, pp. 26–36, 2001.
- [32] L. Perlovsky and R. e. Kozma, *Neurodynamics of Cognition and Consciousness*. Springer, 2007.
- [33] T. Braver, J. Reynolds, and D. Donaldson, "Neural mechanisms of transient and sustained cognitive control during task switching," *Neuron*, vol. 39, pp. 713–726, 2003.
- [34] N. Dosenbach, K. Visscher, E. Palmer, F. Miezin, K. Wenger, H. Kang, E. Burgund, A. Grimes, B. Schlagger, and S. Petersen, "A core system for the implementation of task sets," *Neuron*, vol. 50, pp. 799–812, 2006.
- [35] A. Baddeley, *Working Memory*. Oxford University Press, New York, 1986.
- [36] P. Goldman-Rakic, "Cellular basis of working memory," *Neuron*, vol. 14, pp. 477–485, 1995.
- [37] S. Funahashi, "Prefrontal cortex and working memory processes," *Neuroscience*, vol. 139, pp. 251–261, 2006.
- [38] G. Miller, "The magical number seven, plus or minus two: some limits of our capacity for processing information," *Psychological Review*, vol. 63, pp. 81–97, 1956.
- [39] K. Ericsson and W. Kintsch, "Long-term working memory," *Psychological Review*, vol. 102, pp. 211–245, 1995.
- [40] B. Baars, *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press, 2003.
- [41] K. Heilman, S. Nadeau, and D. Beversdorf, "Creative innovation: possible brain mechanisms," *Neurocase*, vol. 9, pp. 369–379, 2003.
- [42] E. Bowden, M. Jung-Beeman, J. Fleck, and J. Kounios, "New approaches to demystifying insight," *Trends in Cognitive Sciences*, vol. 9, pp. 322–328, 2005.
- [43] J. Hadamard, *The Psychology of Invention in the Mathematical Field*. Princeton University Press, 1945.
- [44] D. Simonton, *Scientific Genius: A Psychology of Science*. Cambridge University Press, 1988.
- [45] H. Poincaré, *The Foundations of Science*. Science Press, 1921.
- [46] S. Mednick, "The associative basis of the creative process," *Psychological Review*, vol. 69(3), pp. 220–232, 1962.
- [47] H. Damasio, T. Grabowski, D. Tranel, R. Hichwa, and A. Damasio, "A neural basis for lexical retrieval," *Nature*, vol. 380, pp. 499–505, 1996.
- [48] A. Martin, "The representation of object concepts in the brain," *Annual Review of Psychology*, vol. 58, pp. 25–45, 2007.
- [49] K. Patterson, P. Nestor, and T. Rogers, "Where do you know what you know? the representation of semantic knowledge in the human brain," *Nature Rev. Neurosci.*, vol. 8, pp. 976–987, 2007.
- [50] Y. Yamane, K. Tsunoda, M. Matsumoto, A. Phillips, and M. Tanifuji, "Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex," *Journal of Neurophysiology*, vol. 96, pp. 3147–3156, 2006.
- [51] L. Lin, R. Osan, and J. Tsien, "Organizing principles of real-time memory encoding: neural clique assemblies and universal neural codes," *Trends in Neurosciences*, vol. 29, p. 57, 2006.
- [52] I. Tsuda, "Towards an interpretation of dynamic neural activity in terms of chaotic dynamical systems," *Behav. Brain Sci.*, vol. 24, pp. 793–847, 2001.
- [53] J. Gibson, *The Ecological Approach to Visual Perception*. Psychology Press, 1979.
- [54] A. Pouget and L. Snyder, "Computational approaches to sensorimotor transformations," *Nature Neuroscience Supp.*, vol. 3, pp. 1192–1198, 2000.
- [55] S. Diboldi, V. Brown, and A. Minai, "A conceptual neural model of idea generation," in *Proceedings of IJCNN 2009*, 2009.