

## A Dynamical Connectionist Model of Idea Generation

Ali A. Minai, Laxmi R. Iyer, Divyachapan Padur, and Simona Doboli

*Abstract:* In this paper, we present a model for the generation of ideas within a creative thinking/brainstorming context. In the model, ideas emerge as conceptual combinations from the interaction of complex dynamics at several semantic levels: Features, concepts, categories, and previously generated ideas. This dynamics is shaped by external information on task context, constraints and goals, and is modulated by evaluative feedback from an internal critic working through reinforcement. While the model is abstract, it attempts to capture the interplay between semantic representations in the temporal, frontal and parietal cortices, working memory in the prefrontal cortex, attentional selection by the basal ganglia, and modulation from the dopaminergic reward system. We show that a context-specific itinerant search for novel but meaningful conceptual combinations (ideas) emerges naturally from the dynamics of this system. We also briefly describe a computational model for ideation in groups using a multi-agent formalism. The initial focus of this model is on studying the potential benefits of cognitive diversity in agent groups, e.g., the presence of convergent and divergent thinkers, or agents with different semantic organizations.

### I. INTRODUCTION

Creative thinking is a distinctive feature of human cognition, and is critical to the success of individuals, organizations and societies. However, the biological and social factors underlying this process remain poorly understood, in part because they are difficult to measure directly through experiments [1]. While there has been extensive neurobiological study of individual semantic cognition [2]-[13], the neurobiology of creative ideation has been much less studied [14]-[16]. Creativity has been studied mainly through behavioral experiments with individuals [17]-[19] and groups [20]-[24]. A crucial link between the

psychobiology of individual cognition and social cognition is still missing, and computational modeling represents a promising approach in this regard. In this paper, we present a dynamic connectionist model of individual idea generation, and discuss how it can be embedded in social models of group ideation. The approach builds on earlier models using associative neural networks [21], [25]-[27]], and draws inspiration from alternative models of creativity [28].

The proposed model for individual ideation comprises a connectionist dynamical system that embodies semantic information at many levels. Ideas – defined as conceptual combinations – arise emergently as *metastable resonant states* from the natural dynamics of this system. This dynamics is shaped by several factors, including task/goal specification, prior learning, and modulation by an internal critic that evaluates ideas as they are generated. While abstract, the model attempts to capture the interplay between semantic representations in the temporal and parietal cortex, working memory in the prefrontal cortex, attentional selection by the basal ganglia, and modulation from the dopaminergic reward system. In this respect – and in its overall conception – the model shares many features with adaptive resonance theory (ART) [29],[30] and the mental agents model of Houk and Wise [31],[32]. It also draws upon ideas from dynamical models of cognitive function [33]-[37]. Group ideation is modeled by a system of interacting agents [38] generating ideas through the interaction of their internal cognitive dynamics and ideas received from other agents.

### II. BACKGROUND

Ideas are fundamentally semantic constructs, and modeling the ideation process requires an understanding of two issues, which are briefly discussed in this section.

#### A. The Organization of Semantic Information

Over the last three decades, extensive research has shown that specific regions of the cortex are preferentially activated during semantic tasks [2]-[13], suggesting the existence of one or more semantic processing networks in the brain [4],[7],[2]. These regions include the left temporal lobe [2], parts of the frontal and parietal cortices, and the premotor cortex, as well as the right inferotemporal cortex [10]. A consensus has emerged [2],[12] that conceptual representations of concrete objects are based on activating representations of their sensorimotor features [3]. The differential distribution of specific features in semantic categories then produces an emergent differentiation in categorical representation as well [11]. In addition to these modality-specific (i.e., feature-based or use-based)

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 (0729305), which includes support from the Deputy Director of National Intelligence.

Ali A. Minai (corresponding author) is with the Department of Electrical & Computer Engineering, University of Cincinnati, Cincinnati, OH 45221-0030, USA (phone: 513-556-4783; fax: 513-556-7326; e-mail: Ali.Minai@uc.edu).

Laxmi R. Iyer is with the Department of Computer Science, University of Cincinnati, Cincinnati, OH 45221-0030, USA (email: iyerlr@email.uc.edu).

Divyachapan Padur is with the Department of Electrical & Computer Engineering, University of Cincinnati, Cincinnati, OH 45221-0030, USA (e-mail: [divyachapan@gmail.com](mailto:divyachapan@gmail.com)).

Simona Doboli is with the Department of Computer Science, Hofstra University, Hempstead, NY 11549, USA (e-mail: Simona.Doboli@hofstra.edu).

representations, there is also evidence for amodal representations in the temporal pole and nearby areas [10],[13]. Top-down control of semantic information selection and retrieval is mediated by the ventro-lateral prefrontal cortex (VL-PFC) [2].

### B. Control Mechanisms

In any given context, the ideation process must focus selectively on a relevant part of the overall semantic space, i.e., a *domain of thought*. This is similar to attentional focus in perception or the selective triggering of motor programs in behavior. Indeed, core cognitive processes such as ideation or imagery should be seen as “internal behaviors” and should be subject to the same control mechanisms [32]. Selective activation of cortico-thalamic modules by the basal ganglia (BG) is thought to be the primary mechanism mediating such control [31],[32],[39]-[41], with important roles played by the anterior cingulate cortex (ACC), the orbitofrontal cortex (OFC) and the locus coeruleus (LC) [42]. In this framework, the prefrontal cortex (PFC) instantiates a task-relevant working memory that is used as the basis for selectively activating functional response networks (e.g., motor programs) in other parts of the cortex [43],[44] through the basal ganglia [32],[46], with the ACC-OFC-LC system modulating the perception of task utility through noradrenergic projections [42]. Useful mappings from working memory to behavioral programs are seen as being configured via reinforcement learning mediated by dopamine signals [45],[46]. Dopamine has also been proposed as subserving a stabilizing function [47],[48] and context-dependent switching [49].

## III. A CONCEPTUAL FRAMEWORK FOR IDEATION

Based on the above observations, we model the semantic-conceptual system as a network of distributed and differentiated feature-based systems linked together by a higher-level system of amodal and categorical “pointer” representations. This is described below in greater detail.

### A. Basic Assumptions

The fundamental assumptions underlying the proposed framework are:

1. Semantic knowledge is organized through modular representational units termed *concepts*. The set of all concepts defines the *conceptual space* for the semantic system. Concepts are defined in terms of *features*, which can be sensory, motor, functional or abstract, as suggested by the studies cited above.
2. Salient combinations of features implicitly cluster concepts that share these features into overlapping partial *categories*. These clusters thus provide a semantic space more coarse-grained than the concept space, and its elements can serve as building blocks of useful search spaces for ideas – much like the semantic fields postulated in [27].

3. Ideas are structured combinations of concepts bound by the coordinated (e.g., simultaneous or synchronized) activation. The space of all conceptual combinations defines the *idea space* for the system.

4. The flow of thoughts/ideas arises from the natural dynamics of the networks in which they are embedded, involving the formation, brief persistence and dissolution of coordinated patterns of activity in these networks. The dynamics of these activity patterns arises from the continual, bidirectional interaction of external stimulus, internal attention, working memory, associational activation of conceptual information, and modulatory feedback, leading to itinerant dynamics [34] with metastable resonant states [29],[30].

5. During productive ideation, the flow of ideas does not range over the entire concept space, but is confined a subspace termed the *effective search domain (ESD)*. The ESD is also continually reconfigured by the natural dynamics of the system based on the given task context, previously accumulated semantic knowledge, external cues, real-time feedback from an *internal evaluation system* [42],[45].

6. The reconfiguration of the ESD is mediated by an *internal attention/working memory (IAWM) system* and exploits the anatomical and functional modularity of this system [31],[32]. The mechanisms underlying the dynamic configuration of the ESD for ideation involve selective combination and switching of category modules.

7. The dynamics of conceptual flow and ESD reconfiguration slowly re-organize the whole system over an extended period to embed useful ideas into the semantic repertoire and to make future ideation more productive.

The main functional feature of the model is continuous bidirectional interaction between three subsystems – 1) the Feature Subsystem; 2) the Concept Subsystem; and 3) the Internal Attention/Working Memory (IAWM) subsystem – each comprising several components with pervasively recurrent connectivity. The system produces responses through a process of *emergent resonance*, i.e., the appearance and temporary persistence of activity patterns as a result of the dynamic interactions. Conceptually, this can be seen as a dynamic constraint satisfaction process, where each subsystem tries to influence the activity in the other two and the only activity patterns that *can* persist are those that are consistent across all three, i.e., make sufficient “sense” within the system in the given context. Associative modification in the synapses within and between subsystems ensure that such self-consistent patterns of activity correspond to functionally useful – albeit implicit – internal representations. Resonance-based models of this type have a long and successful history in computational neuroscience [29],[30], and resolve the problematic issue of causality by

embedding choice as an emergent phenomenon within a circular process.

### B. Creativity and Novelty

Consistent with our unified view of mental response, we see ideation as spanning the same spectrum from “automatic” to “effortful” as all other response processes [43]. Familiar contexts trigger familiar ideas automatically based on previous learning, while novel contexts require effortful search through idea space. Another dimension of creativity lies in discovering novel ideas *within* familiar contexts, which requires the emergence of new conceptual combinations. Based on these considerations, we identify four broad functional modes for ideation: 1) Automatically finding familiar ideas in familiar contexts; 2) Finding useful familiar ideas in novel contexts; 3) Finding novel ideas in familiar contexts; and 4) Finding novel ideas in novel contexts. Of these, the last three would generally be considered creative, but it is important to see that the same mechanism – itinerant conceptual combinations in dynamically configured search domains – is hypothesized to underlie all four modes, with the difference in effort depending on the information pre-configured by previous learning. Over time, ideation requiring creativity can, and does, become mundane through learning – just as skilled action becomes automatic with practice.

## IV. COMPUTATIONAL MODEL

The broad architecture of the model is shown in Figure 1, and is discussed below in greater detail.

### A. System Components

**The Feature Subsystem:** The feature subsystem encodes generic unimodal and polymodal attributes of concepts. Each unit in this subsystem corresponds to a specific attribute. A particular object/concept activates a distributed pattern of activity across the feature subsystem, depending on its attributes. Damasio [4] and others [50] have suggested that these features are bound into coherent representations by synchronization against a backdrop of random activity.

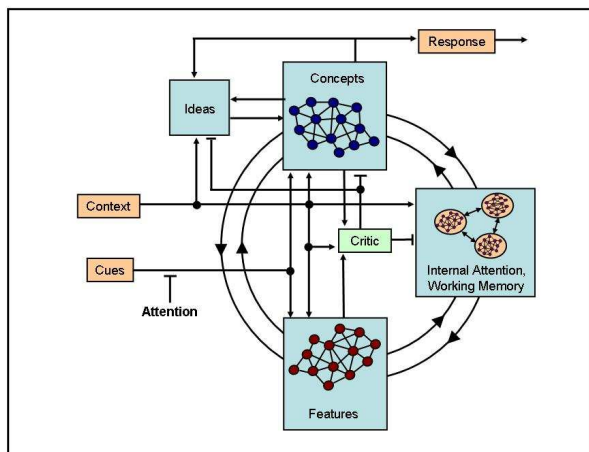


Figure 1: Architecture of the computational model.

Based on experimental data, we tentatively identify the Feature Subsystem with a highly interconnected network comprising the primary sensory cortices, parts of the bilateral inferotemporal cortex (IT), the fusiform gyrus, the premotor cortex and supplementary motor area, the posterior parietal cortex, parts of the frontal lobe and several other widely scattered areas. These are areas activated by the perception and recognition of objects [2],[7],[11].

**The Concept Subsystem:** The Concept Subsystem corresponds to amodal and lexical representations that have been proposed to function essentially as pointers to the feature-based direct representation in the Feature Subsystem [2]-[13]. These provide a “vocabulary” for referencing complex concepts, and mediate lexical and phonological representations. Concepts are represented in the subsystem by the activity of specific units which are associated reciprocally with the corresponding feature representations in the Feature Subsystem. The lateral weights between two concept units represents their joint utility, i.e., their propensity to participate jointly in useful ideas under various contexts. Thus, the connectivity of the Concept Subsystem is symmetric, and encodes the context-independent semantic structure of the concept space.

The units in the concept subsystem have competitive  $k$ -of- $N$  activation, with the  $k$  most active units firing at each step, provided a threshold condition is met. Each unit can only stay on for a finite time (randomly chosen between 4 and 10 time steps) followed by a refractory period of random duration. This can be seen as modeling bursty dynamics among the neural population. The competitive activation means that only a few concepts can be active simultaneously, comprising a potential idea. If the co-active concepts have strongly excitatory mutual connections (based on previous learning), they form an emergent attractor and remain persistently active until switched off by the constraint on burst duration. Such persistently active conceptual combinations are recognized as ideas, while those that do not persist are simply transients. Thus, the ideas that arise from this itinerant dynamics typically group concepts that “go together”, i.e., the ideas make sense. However, novel ideas (unusual conceptual combinations) can be allowed to emerge (and persist) by modulation of the neurons’ excitability while maintaining the level of competition.

Based on experimental evidence, we identify the Concept Subsystem with lexical areas [5] – and possibly with the temporal pole of the left hemisphere that appears to provide amodal representations [13]. The recurrent connectivity may also loop through the cerebellum, which seems to play a role in sequencing both thought and action [32].

**The Internal Attention/Working Memory Subsystem:** The IAWM Subsystem is seen as a rich reservoir of higher order combinations – of features and concepts – providing

the “raw material” for constructing productive idea spaces. Each unit of this subsystem – termed a *category unit* for convenience – groups together concepts that share a *salient subset* of features, and is excitatorily connected to these concepts and features (Figure 2). Essentially, the category units cluster partially similar concepts, and serve as basis functions for constructing subspaces within the overall space of ideas just as the units of the Concept Subsystem serve as basis functions for the ideas themselves. Some category units may group together unusual feature combinations, and can serve as “short cuts” to link together normally disparate concepts into novel ideas [51] – perhaps in an analog to the proposed function of the right hemisphere in mediating insight [27].

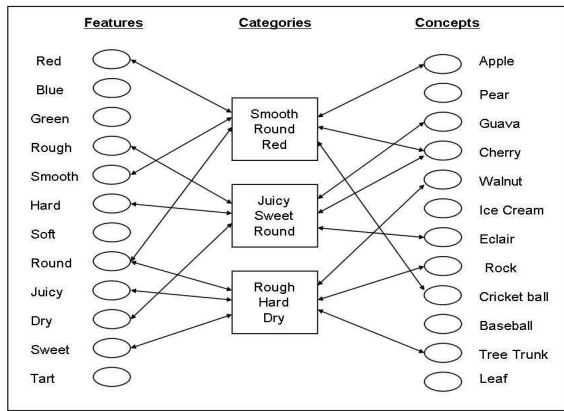


Figure 2: IAWM connectivity with features and concepts.

Because of their structured projections onto the concept units, each combination of active IAWM units “illuminates” a part of idea space from a specific perspective defined by the features associated with these active units. This defines a *bias mask* on the concepts. Due to its competitive nature, the itinerant activity in the Concept Subsystem remains confined in and around this biased region – thus defining the ESD for the ideation process. This is discussed below in more detail.

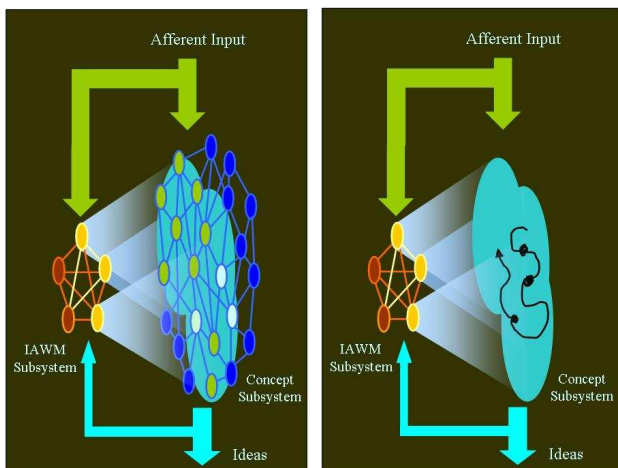


Figure 3: Configuration of the ESD in the Concept Subsystem by biasing from the IAWM Subsystem.

Category units with similar features have strong connections to each other, resulting in a modular organization. These modules are sparsely connected by excitatory and inhibitory connections, as well as indirectly through a set of *gating units*, which are essentially pattern recognition devices. Each module has a set of gating units (GUs), and can respond only if these gating units are sufficiently active. In addition to recurrent feedback from category units, the GUs receive input representing task context and goal information. Over time, each module’s GUs learn the external and internal contingencies (i.e., the activity of other modules) under which it is useful to enable their module, given what the module encodes through its connections with the Concept and Feature Subsystems. The system is thus able to dynamically configure modules into stimulus- and context-specific functional networks representing a working memory for ideation. These can be expanded and contracted through modulation of global inhibition by evaluative feedback, and modules can be switched in and out by varying the activity of the gating units. The gating system is a simplified model of the BG system – with striatal spiny neurons performing the pattern recognition function – which is assumed to mediate selective activation of cortical response modules in many cognitive processes [31],[32],[39]-[41] and of the thalamic targets of this system. The gating units and the category units are also subject to modulation by a reward signal from the Critic (see below), which corresponds to a dopamine input that encodes reward in the brain [45]. Figure 4 shows the architecture of the IAWM system.

Because of the structure described above, the IAWM subsystem has a very rich repertoire of switching and modulatory possibilities, and useful patterns of activity can be quickly configured, stabilized, expanded and contracted, merged, dissolved and reconfigured through the dynamics of the system. This, in turn, projects a “dynamic searchlight” onto the concept space, creating the ESD in which ideas can be searched for. As activity circulates in the great loop of the system, the pattern of activation in the IAWM subsystem serves as a self-organized dynamic adaptive filter to focus the flow in productive directions. The activity of category units changes much more slowly than that of Concept units, with individual category units staying tonically active for extended periods during search. Such context-latching tonically active units have, in fact, been observed in the prefrontal cortex during the performance of extended tasks [40], and we speculate that similar PFC units exist for representing internal “tasks” such as searching for a good idea. The latching in such units can be triggered by BG units that become tuned to specific contexts [41].

We identify the IAWM subsystem with the frontal and left temporal lobe regions encoding categorical information [6],[4],[2], prefrontal cortex (PFC), basal ganglia (BG), anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), locus coeruleus (LC), and possibly right hemisphere regions that encode broader, possibly non-verbal, semantic fields [27]. The context input may also involve the

prefrontal cortex, serving as a working memory of the task situation. The projection from PFC to the striatum is known to be modulated by dopaminergic signals from the substantia nigra, as is the connectivity within the PFC [47],[48].

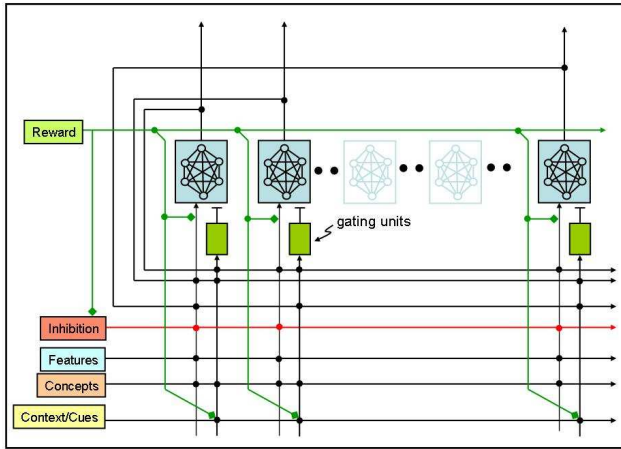


Figure 4: Architecture of the IAWM system. Connections between category units can be excitatory or inhibitory. Global inhibition is modulated by cumulative reward.

**The Critic:** The critic is an internal evaluation system providing reinforcement to the other systems in order to facilitate both working memory and permanent associations. Dopamine has been identified as the primary mediator of reinforcement in both cortical and subcortical regions [45], reflecting the direct value judgments of subcortical nuclei. In the current model, the critic is implemented as a rule-based oracle without an explicitly neural architecture. The rules embedded in the critic represent reasonable context-specific semantic evaluations over the space of ideas. The goal in the present model is for the neural system to infer these ideational preferences of the critic in each context – including those not explicitly encoded in the system – purely through evaluative feedback. The critic thus serves as a convenient way to ground the model. Eventually, our aim is to train critics based on reinforcement provided by human evaluators or other real-world oracles. It should be noted, however, that creative ideation represents a somewhat different situation than the reinforcement-based learning of sequential responses in embodied systems [52]. Even in the real world, ideas typically do not elicit unambiguous evaluations – either by the thinker or by his/her interlocutors, whereas it is usually clear whether a behavioral response accomplished its goal.

**The Idea Subsystem:** This subsystem builds representations of generated ideas found to be useful in specific contexts, and associates them with these contexts. These ideas then help shape the EDS when similar contexts occur in the future.

### B. Setting Initial Synaptic Weights

In the current version of the model, we embed a repertoire of features, concepts and categories according to simple Hebb-like rules and the specification of each semantic level as

described above. A set of familiar contexts is defined, and a large number of sensible ideas are constructed for each of these using the available concepts. These ideas are used to determine the initial connectivity between concepts and, very importantly, between category units. Appropriate category modules are also associated synaptically with each of these contexts. Thus, the system is pre-disposed to retrieve relevant ideas in these *familiar contexts*, which corresponds to automatic recall [43]. However, the system can also produce ideas in unfamiliar contexts, or novel ideas in familiar contexts, by supplementing the basic dynamics of the system with short-term synaptic modification mediated by the reward signal as described in the next section.

### C. Shaping the Search for Ideas

As described above, activity patterns in each subsystem interact dynamically to generate resonant patterns across the whole system. Essentially, the Feature and Concept Subsystems generate a trajectory through the space of ideas, while the IAWM Subsystem biases this trajectory by illuminating specific regions of this space. This “shaping” process draws upon previous experience as well as current successes or failures, and is perhaps the most novel aspect of the model. Its function is described below in greater detail.

Activity in the IAWM Subsystem is competitive so only a limited number of units can be on at a time. This number depends on the level of global inhibition, which is itself adaptive. The feature and concept subspaces represented by these active units thus constitute the current space for ideation (the ESD), and the dynamics of this selective focus is the primary control mechanism for the process. This is done through three interacting mechanisms: 1) By modulating the global inhibition across the subsystem, changing the number of simultaneously active units; 2) By gating the inhibitory and excitatory connections between IAWM modules; and 3) By temporarily modifying the connections between category units based on reinforcement from the critic.

When a context is first presented to the system, it activates some category units based on previous associations (e.g., if the given context is identical or similar to a familiar one). This creates an initial ESD in the concept space, and the itinerant flow in the Concept Subsystem begins to generate candidate ideas. If the context is familiar, sensible ideas will appear automatically because of prior learning. If the context is not familiar, or if novel ideas are required, the initial process is fruitless, indicating that an effortful response is needed [43]. The building level of frustration lowers the level of global inhibition and allows the ESD to expand by activating other IAWM modules of likely relevance. This is termed the *trial stage* of the search, where the system is groping for a good ESD. Once the ESD includes concepts that can combine usefully into good ideas, these ideas begin to appear, generating reward feedback. The system now enters the *discovery stage* of the search. The rewards have

two consequences. First, every time an idea is rewarded, the connections between category units whose concepts participated in the rewarded idea are strengthened. This corresponds to the dopamine reward signal making all category units more excitable [49] and the feedback from the currently active concepts, causing their parent category units to fire strong bursts and temporarily potentiate their mutual connections. Such a potentiation mechanism has recently been proposed as a substrate for working memory [53]. The connections with and between the rewarded concept units are also potentiated. Second, as this reward accumulates, the inhibition is raised again, causing some IAWM modules to drop out of the ESD. However, the modules with recently rewarded category units tend to survive this, leaving a more parsimonious and productive ESD for the given context.

If reward dries up again, inhibition is lowered again and the process repeats. Several such repetitions essentially comprise an annealing process, with which the system gradually discovers an appropriate conceptual subspace for the given context. Once ideation is over, this discovered ESD is associated with the context input for future use (the context becomes familiar), while the other weights decay back to their earlier values with a small residual change from the recent learning. If reinforced further by repetition, this change can reconfigure the overall semantic space.

The process described above tries to balance the two imperatives of exploitation, i.e., using prior knowledge, and exploration, i.e., seeking new possibilities [52]. The search first focuses on ideas that have worked previously in similar situations. If that does not prove fruitful, the search is widened, but still tries to exploit previously learned semantic relationships. This is facilitated by the encoding of these relationships into the modularity pattern of the IAWM Subsystem. The system is thus inherently configured for *experience-guided exploration*.

#### D. Modeling Styles of Thought

It is well known that people show diverse styles of thinking during ideation [1]. In particular, a distinction is made between *convergent (exploitative) thinkers*, who select a narrow search domain (ESD) and tend to remain focused within it, and *divergent (exploratory) thinkers*, who select broad domains and/or switch easily between domains. These styles of thinking can be captured with the model described above. We hypothesize that the difference lies primarily in the mechanisms of the IAWM Subsystem, with three specific dimensions of variation:

Difference in connectivity patterns, leading some individuals to have more strongly partitioned idea spaces while others have more linkages between diverse ideas.

Differences in inhibitory modulation, leading to greater or lesser propensity for switching and merging categories.

Differences in response to dopaminergic modulation, with convergent thinkers being less sensitive to such modulation.

Interestingly, many pathologies of cognitive or behavioral control (e.g., OCD, Tourette's syndrome, Parkinsonism, Huntington's disease, etc.) have been ascribed to variation in connectivity or deficits of modulation in the basal ganglia, orbitofrontal cortex and the anterior cingulate cortex [47],[54], which are part of our IAWM Subsystem. Convergent and divergent thinking may just be more moderate, non-pathological examples of such variation. Other factors not captured by our model may also be involved. Only experimental data can settle the issue, but the model can suggest hypotheses and make predictions. An especially interesting area in this regard is the anecdotally established connection between chemical intoxicants (alcohol, peyote, laudanum, LSD, etc.) and creativity.

#### E. Metrics for Ideation

The greatest problem in studying ideation – especially creativity – is to quantify it [1]. Within our limited framework, we have defined four metrics to evaluate ideas:

**Admissibility**: Does the idea have concepts satisfying the minimum necessary requirements for the given context/goal?

**Quality**: Does the idea include concepts/features are especially desirable for the given context/goal?

**Parsimony**: Is the idea as compact as possible, e.g., does not include redundant concepts, while including all necessary components?

**Novelty**: Does the idea combine concepts that, a priori, had a low probability of combination (as reflected in the system's current synaptic weights)?

A situation-specific metric is constructed by a weighted combination of these, e.g., if the goal is to seek novel ideas, Novelty is given a higher weight. Admissibility is a binary metric, and is always included with a maximal weight.

### V. SIMULATION RESULTS FOR THE INDIVIDUAL MODEL

At this preliminary stage, we have implemented the model using connectionist networks for each layer, and simulated it with a few embedded “toy” contexts, covering a few hundred concepts and features. The results reported in this section mainly illustrate the dynamics of ESD configuration. Figure 5 shows a case where ideas are needed in an unfamiliar context (Romantic Vacation) which resembles a familiar one (Relaxing Vacation). The latter is thus primed by the former, and the initial search produces ideas that would be suitable for the familiar context. However, they are of relatively low quality in the actual context – which the system does not know explicitly. However, through the dynamic search mechanism described above, the system

eventually discovers good ideas in the new context, as indicated by the quality and reward time series (lower two panels). It is worth noting that discovering good ideas

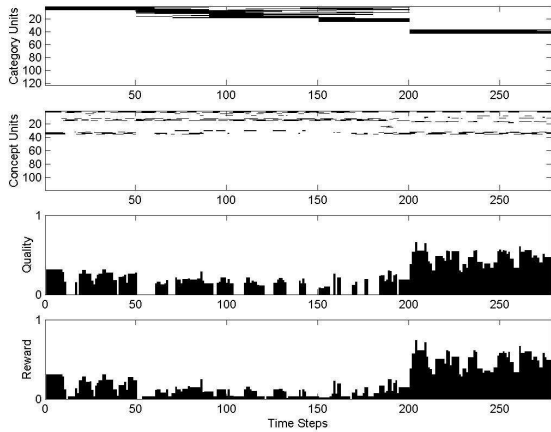


Figure 5: Idea generation in a novel context. The novel context is “Romantic Vacation”, which is similar to the familiar context “Relaxing Vacation”.

required finding a completely different category module than the one used for the similar familiar context (top panel), but the two shared many concepts (second panel), as would be expected from their similarity. Examples of ideas generated early in the search included (*motel, visit to scenic location, picnic, board game*) and (*motel, staying with relatives, picnic, board game*), which were not very good. However, eventually, the system found ideas like (*flight, five star hotel, island, picnic*) and (*flight, island, chalet, visit to a scenic location, picnic*), which were much better. Figure 6 shows an instance where the unfamiliar context (Healthy Meal) has relatively less similarity to the primed familiar context (Fast Food). However, the limited resemblance eventually does allow the system to find good ideas by switching to a very different set of categories. The model has also been used to replicate results from behavioral experiments where an individual generating ideas is primed with relevant and irrelevant cues. These results are reported elsewhere [55].

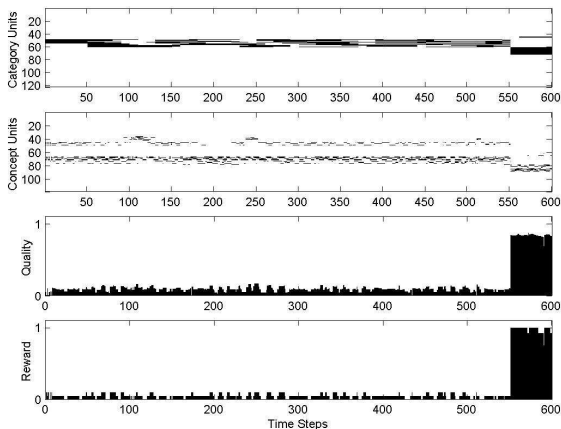


Figure 6: Idea generation in a more novel context. The novel context is “Healthy Meal”, which is slightly similar to the familiar context “Fast Food”.

## VI. GROUP IDEATION MODEL

In group situations, ideas expressed by individuals can modify the internal ideation processes in other group members. This does not always lead to increased productivity, but has a distinctive phenomenology [20]-[24]. We model a group as a social network of agents [38],[56], each with its own ideation process, represented by a dynamical neural network with units representing ideas clustered into multiple domains – corresponding roughly to the ESDs of the individual model. The system’s dynamics tends to remain within its current domain unless switched out of it by noise, interference, or by an external cue (e.g., an idea from another agent). Each domain covers a subset of potential ideas, and ideas can be shared between domains. Each given context (represented by a fixed external input) defines a value distribution over the set of ideas, and the goal of the system dynamics is to find a domain that is dense in high-value ideas. The multi-agent system is a social network with a *parameterized interaction protocol (PIP)* that models the agents’ social behavior. We also consider different patterns of connectivity between agents and different patterns of attention to cues among agents.

We investigate several specific issues with this model, but the focus is on looking at the effects of agent diversity on the productivity of ideation. In particular, we model convergent agents as having multiple domains separated by high potential barriers, while divergent agents have fewer, broader domains with smaller barriers. Agents can also differ in how they respond to cues from other agents, and how ideas age within their own working memory. These factors are major determinants of productivity in group ideation [20]-[24].

## VII. CONCLUSION

In this paper, we have described a dynamical, distributed computational model for ideation based loosely on results from neurobiology and cognitive science. The model is still somewhat speculative, but is a first step towards a more detailed and realistic model. The current version is nevertheless able to capture important aspects of the phenomenology of ideation.

## ACKNOWLEDGMENT

This work is part of a larger collaboration with Vince Brown, Paul Paulus and Dan Levine, and has benefited greatly from their input. A.A. Minai acknowledges helpful suggestions from Jim Houk and Shane Muller.

## REFERENCES

- [1] J.P. Guilford, *The Nature of Human Intelligence*, New York: McGraw-Hill, 1967.
- [2] A. Martin, “The representation of object concepts in the brain,” *Ann. Rev. of Psychol.*, vol. 58, pp. 25-45.
- [3] E.K. Warrington and T. Shallice, “Category-specific semantic impairment,” *Brain*, vol. 107, pp. 829-854, 1984.

- [4] A.R. Damasio (1989) The brain binds entities and events by multiregional activations from convergence zones," *Neural Computation*, vol. 1, pp. 123-132, 1989.
- [5] H. Damasio et al. (1996) A neural basis for lexical retrieval. *Nature*, 380, 499-505.
- [6] A. Martin, C.L. Wiggs, L.G. Ungerleider, and J.V. Haxby, "Neural correlates of category-specific knowledge," *Nature*, vol. 379, pp. 649-652, 1996.
- [7] D. Tranel, H. Damasio and A.R. Damasio, "A neural basis for the retrieval of conceptual knowledge," *Neuropsychologia*, vol. 35, pp. 1319-1327, 1997.
- [8] C.J. Mummery, T. Shallice and C.J. Price, "Dual process model in semantic priming: a functional imaging perspective," *Neuroimage*, vol. 9, pp. 516-525, 1999.
- [9] H. Damasio et al., "Neural systems behind word and concept retrieval," *Cognition*, vol. 92, pp. 179-229, 2004.
- [10] M.L. Kellenbach, M. Brett and K. Patterson, "Large, colourful or noisy? Attribute- and modality-specific activations during retrieval of perceptual attribute knowledge," *Cognit. Affect. Behav. Neurosci.*, vol. 1, pp. 207-221, 2001.
- [11] A. Caramazza, A. and B.Z. Mahon, "The organization of conceptual knowledge: the evidence from category-specific semantic deficits," *Trends in Cog. Sci.*, vol. 7, 354-pp. 361, 2003.
- [12] R.F. Goldberg, C.A. Perfetti and W. Schneider, "Distinct and common cortical activations for multimodal semantic categories," *Cognit. Affect. Behav. Neurosci.*, vol. 6, pp. 214-222, 2006.
- [13] K. Patterson, P.J. Nestor and T.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.
- [14] K.M. Heilman, S.E. Nadeau and D.O. Beversdorf, "Creative innovation: possible brain mechanisms," *Neurocase*, vol. 9, pp. 369-379, 2003.
- [15] K. Christoff, J.M. Ream J.D.E. Gabrieli, "Neural basis of spontaneous thought processes," *Cortex*, vol. 40, pp. 623-630, 2004.
- [16] E.M. Bowden, M. Jung-Beeman, J. Fleck and J. Kounios, "New approaches to demystifying insight," *Trends. In Cog. Sci.*, vol. 9, pp. 322-328, 2005.
- [17] M.D. Mumford and S.B. Gustafson, "Creativity syndrome: integration, application, and innovation," *Psychol. Bull.*, vol. 103, pp. 27-43, 1988.
- [18] M.I. Mobley, L.M. Doares and M.D. Mumford, "Process analytic models of creative capacities: Evidence for the combination and reorganization process," *Creativity Res. J.*, vol. 5, pp 125-155, 1992.
- [19] D.K. Simonton, "Foresight and insight? A Darwinian perspective," in R.J. Sternberg and J.E. Davidson, Eds., *The Nature of Insight*, pp. 465-494, Cambridge, MA: MIT Press.
- [20] P.B. Paulus and M.T. Dzindolet, "Social influence processes in group brainstorming," *J. Personal. Soc. Psychol.* vol. 64, pp. 575-586, 1993.
- [21] V. Brown, M. Tumeo, T.S. Larey and P.B. Paulus, "Modeling cognitive interactions during group brainstorming," *Small Group Res.*, vol. 29, pp. 495-526, 1998.
- [22] H. Coskun, P.B. Paulus, V. Brown and J.J. Sherwood, "Cognitive simulation and problem presentation in idea generation groups," *Group Dynamics*, vol. 4, pp. 307-329, 2000.
- [23] K.L. Dugosh, P.B. Paulus, E.J. Roland and H.C. Yang, "Cognitive simulation in brainstorming," *J. Personality and Soc. Psychol.*, vol. 79, pp. 722-735, 2000.
- [24] B.A. Nijstad and W. Stroebe, "How the group affects the mind: a cognitive model of idea generation in groups," *Personality and Soc. Psychol. Rev.*, vol. 10, pp. 186-213, 2006.
- [25] S. Diboldi, A.A. Minai and V.R. Brown, "Adaptive dynamic modularity in a connectionist model of context-dependent idea generation," in *Proc. IJCNN*, Orlando, FL, August 2007.
- [26] L.R. Iyer, A.A. Minai, S. Diboldi and V.R. Brown, "Modularity and self-organized functional architectures in the brain," *Proc. 7th Int. Conf. on Complex Systems*, Boston, MA, October 2007.
- [27] W. Duch, "Intuition, insight, imagination and creativity," *IEEE Comput. Intell.*, Aug. 2007, pp 40-52.
- [28] D.R. Hofstadter, *Fluid Concepts and Creative Analogies*. Basic Books, 1995.
- [29] G.A. Carpenter. and S. Grossberg, "ART2: self-organization of stable category recognition codes for analog input patterns," *Appl. Optics*, vol. 26, pp. 4919-4930, 1987.
- [30] G.A. Carpenter. and S. Grossberg, S., "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Comp. Vision, Graphics, and Image Proc.*, vol. 37, pp. 54-115, 1987.
- [31] J.C. Houk and S.P. Wise, "Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action," *Cerebral Cortex*, vol. 5, pp. 95-110, 1995
- [32] J.C. Houk, "Agents of the mind," *Biol. Cybern.*, vol. 92, pp. 427-437, 2005.
- [33] W.J. Freeman, *Neurodynamics: An Exploration of Mesoscopic Brain Dynamics*, London: Springer, 2000.
- [34] I. Tsuda, "Towards an interpretation of dynamic neural activity in terms of chaotic dynamical systems," *Behav. Brain Sci.*, vol. 24, pp. 793-847, 2001.
- [35] S.L. Bressler and E. Tognoli, "Operational principles of neurocognitive networks," *Int. J. Psychophysiol.*, vol. 60, pp. 139-148, 2006.
- [36] V.K. Jirsa, "Connectivity and dynamics of neural information processing," *Neuroinformatics*, vol. 2, pp. 183-204, 2004.
- [37] G. Tononi, G.M. Edelman and O. Sporns, "Complexity and Coherency: integrating information in the brain," *Trends in Cog. Sci.*, vol. 2, pp. 474-484, 1998.
- [38] R.L. Goldstone, M.E. Roberts and T.M. Guereckis, "Emergent processes in group behavior," *Curr. Dir. Psychol. Sci.*, vol. 17, pp. 10-15, 2008.
- [39] A.M. Graybiel, "The basal ganglia and chunking of action repertoires," *Neurobiology of Learning and Memory*, vol. 70, 119-136, 1998.
- [40] N. Fujii and A.M. Graybiel, "Representation of action sequence boundaries by macaque prefrontal cortical neurons," *Science*, Vol. 301, pp. 1246-1249, 2003.
- [41] M.S. Jog, Y. Kubota, C.I. Connolly, V. Hillgaart, A.M. Graybiel, "Building neural representations of habits," *Science*, Vol. 286, pp. 1745-1749, 1999.
- [42] G. Aston-Jones and J.D. Cohen, "An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance," *Ann. Rev. Neurosci.*, vol. 28, pp. 403-450, 2005.
- [43] S. Dehaene, M. Kerszberg and J.-P. Changeux, "A neuronal model of a global workspace in effortful cognitive tasks," *Proc. Nat. Acad. Sci. USA*, vol. 95, pp. 14529-14534, 1998.
- [44] A.P. Shimamura, "The role of the prefrontal cortex in dynamic filtering," *Psychobiology*, vol. 28, pp. 207-218, 2000.
- [45] W. Schulz, "Multiple reward signals in the brain," *Nature Rev. Neurosci.*, vol. 1, pp. 199-207, 2000.
- [46] R.C. O'Reilly and M.J. Frank "Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia," *Neural Computation*, vol.18, pp. 283-328, 2006.
- [47] J.D. Cohen and D. Servan-Schreiber, "Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia," *Psychological Review*, vol. 99, pp. 45-77, 1992.
- [48] D. Durstewitz, "Neurocomputational models of working memory," *Nature Neuroscience Supp.*, vol. 3, pp. 1184-1191, 2000.
- [49] A.J. Gruber, P. Dayan, B.S. Gutkin and S.A. Solla, "Dopamine modulation in the basal ganglia locks the gate to working memory," *J. Comput. Neurosci.*, vol. 20, pp. 153-166, 2006.
- [50] C.M. Gray and W. Singer, "Stimulus-specific neuronal oscillations in the orientation columns of the cat visual cortex," *Proc. Nat. Acad. Sci USA*, vol. 86, pp. 1689-1702, 1989.
- [51] M.A. Schilling, "A small-world network model of cognitive insight," *Creativity Res. J.*, vol. 17, pp. 131-154, 2005.
- [52] R.S. Sutton A.G. Barto, *Reinforcement Learning*, Cambridge, MA: MIT Press, 1998
- [53] G. Mongillo, O. Barak and M. Tsodyks, "Synaptic theory of working memory," *Science*, vol. 319, pp. 1543-1546, 2008..
- [54] A.M. Graybiel and S.L. Rauch, "Towards a neurobiology of obsessive-compulsive disorder," *Neuron*, vol. 28, pp. 343-347, 2000.
- [55] L.R. Iyer, A.A. Minai, V.R. Brown and S. Diboldi, "Effects of relevant and irrelevant cues on idea generation: a computational model," submitted to IJCNN 2009.
- [56] R. Sun, "Cognitive social simulation incorporating cognitive architectures," *IEEE Intell. Sys.*, Sep-Oct 2007, pp. 33-39, 2007.