Towards a Unified Neurobiological Theory of Creative Problem Solving

Sebastien Helie

Abstract—**Very little work in psychology has tried to simultaneously tackle the representational problem and the search problem in creative problem solving. Unifying these two fields is critical, because building and searching the problem space may be highly interactive activities that cannot be decoupled and studied in isolation. In addition. neuroscientific explanations of creative problem solving have been elusive. In this article, we propose an integrative biological theory called Neuro-EII to address these problems. Neuro-EII is based on the Explicit-Implicit Interaction (EII) theory of creative problem solving. Like EII, Neuro-EII is implemented using the CLARION cognitive architecture and can be used to simulate behavioral data. We present simulation results of a lexical decision task and use Neuro-EII to generate testable neuroscientific predictions. We conclude by discussing how the Neuro-EII theory can contribute to solving the seemingly intractable problem of identifying the brain circuit(s) supporting creative problem solving.**

I. Introduction

Many psychological theories have highlighted a role for implicit cognitive processes [1]-[3]. For instance, similarity has been shown to affect reasoning through processes that are mostly implicit [4]. In problem solving, implicit processes are often thought to generate hypotheses that are later explicitly tested [2], [5]-[6]. Yet, most theories of problem solving have focused on explicit processes that gradually bring the problem solver closer to the solution in a deliberative way [7]. However, when the initial state or the goal state of a problem can lead to many different interpretations, or when the solution paths are too complex to be fully grasped in working memory, the solution is often found by sudden 'insight', and regular problem solving theories are for the most part unable to account for this apparent absence of deliberative strategy [8].

A complementary line of research on creative problem solving has tried to tackle complex problem solving for many years. However, theories of creative problem solving tend to be fragmentary and usually focus only on a subset of phenomena, such as incubation (i.e., a period away from deliberative work on the problem) or insight (i.e., the sudden appearance of a solution). The lack of detailed computational models has resulted in their limited impact on the field of problem solving [9].

This separation between theories focusing on explicit and implicit processing leaves the field of problem solving research fragmented. On the one hand, proponents of explicit theories of problem solving typically suggest that every problem can be represented by a problem space including states and operators. Accordingly, problem solving research needs to account for how the problem solver navigates the problem space using available operators to move from the initial state to one of the goal states. This approach has received much attention but, for the most part, ignores the preliminary stage of building the problem space (i.e., finding a good problem representation). On the other hand, Gestalt psychologists instead focused their research effort on accounting for the stage of building problem spaces by generally interpreting problem solving as perception problems. However, less attention has been devoted to the process of navigating through the problem space once it has been built.

While both approaches have generated much research, very little work has tried to simultaneously tackle the representational problem and the search problem. Unifying these two fields is critical, because building and searching the problem space may be highly interactive activities that cannot be decoupled and studied in isolation [10]. In addition to unifying these two areas of problem solving, integrative theories of problem solving can help solve the seemingly intractable problem of identifying the brain circuit(s) supporting creative problem solving.

The remainder of this article is organized as follows. We begin by reviewing a classical stage decomposition of creative problem solving (Section II) and some of the evidence aimed at localizing the cognitive functions relevant for creative problem solving in the brain (Section III). Next, Section IV reviews an integrative theory of creative problem solving, namely the Explicit-Implicit Interaction (EII) theory. The EII theory is used to propose micro-cognitive processes that can account for the stage decomposition of creative problem solving and be implemented using a computational model (e.g., the CLARION cognitive architecture [11]; see Section V). Section VI builds on earlier findings of EII and proposes a neurobiological version of the theory (Neuro-EII). As an example application, Section VII uses Neuro-EII to account for and simulate (with CLARION) the effect of incubation in a lexical decision task and makes predictions about the brain areas that should be involved in this task. Finally, Section VIII concludes with a discussion of the possible impact that an integrative neurobiological theory of creative problem solving can have on the field.

Manuscript received May 1, 2013. S. Helie is with the Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907 USA (phone: 765-496-2692; e-mail: shelie@purdue.edu).

II. The Four stages of Problem Solving

According to Wallas [12], humans go through four different stages when trying to solve a problem: preparation, incubation, insight, and verification. The first stage, preparation, refers to an initial period of search in many directions using (essentially) logic and reasoning. If a solution is found at this stage, the remaining stages are not needed (this is equivalent to searching within a problem space and finding the solution). However, if the problem is ill-defined and/or too complex to be fully grasped (e.g., the problem space has too many branches), the preparation stage is unlikely to generate a satisfactory solution. When an impasse is reached, the problem solver stops attempting to solve the problem, which marks the beginning of the incubation phase. Incubation can last from a few minutes to many years, during which the attention of the problem solver is not devoted to the problem. The incubation period has been shown to increase the probability of eventually finding the correct solution [13]. In the proposed framework, incubation would be the quest to find a new (and hopefully simpler) problem space.

The following stage, insight, is the "spontaneous" manifestation of the problem and its solution in conscious thought (i.e., the "Eureka!" moment). This corresponds to finding a new problem space in which the search for the solution appears trivial. The fourth stage, verification, is used to ascertain the correctness of the insight solution (e.g., ensuring the equivalence of the initial and revised problem spaces). Verification is similar to preparation, because it also involves the use of deliberative thinking processes (with logic and reasoning). If the verification stage invalidates the solution, the problem solver usually goes back to the first or second stage and this process is repeated.

Even though the stage decomposition theory is difficult to test empirically, it has been used to guide much of Gestalt psychologists' early research program on problem solving [14]-[15]. According to Gestalt psychology, ill-defined problems are akin to perceptual illusions: they are problems that can be understood (perceived) in a number of different ways, some of which allow for an easier resolution. Hence, the preparation stage would be made up of unsuccessful efforts on an inadequate problem representation, incubation would be the search for a better problem representation, and insight would mark the discovery of a problem representation useful for solving the problem. The verification phase would verify that the new problem representation is equivalent to the initial problem representation [14]. This Gestalt theory of problem solving provides a sketchy high-level description of creative problem solving but no detailed psychological mechanism was proposed.

More recent research has focused on finding evidence supporting the existence of the individual stages of creative problem solving. Because the preparation and verification stages are thought to involve mostly regular reasoning processes [12], not much effort has been devoted to these two stages (relevant results can be borrowed from the existing literature; see, e.g., [4]). In contrast, incubation and insight have received more attention.

A. Incubation

A recent review of experimental research on incubation shows that most experiments have found a significant effect of incubation [13]. Those experiments found an effect of incubation length, preparatory activity, clue, and distracting activities on participants' performance. The review suggests that performance is positively related to incubation length and that preparatory activities can increase the effect of incubation. Presenting a clue during the incubation period also has a strong effect. If the clue is useful, the performance is improved; if the clue is misleading, the performance is decreased. Moreover, the effect of clues is stronger when the participants are explicitly instructed to look for clues [16].

The effect of distracting activities is not as clear. Helie, Sun, and Xiong [17] showed that distracting activities can have different effects on incubation depending on whether the distracting activities share cognitive resources/processing with the task used to assess the presence of incubation. Finally, incubation has also been linked to well-known cognitive effects such as reminiscence (i.e., the number of new words recalled in a second consecutive free recall test; [18]) and priming [19].

B. Insight

In a review of the different definitions used in psychology to characterize 'insight', Pols [20] found three main elements. First, insight does not constitute just another step forward in solving a problem: it is a *transition* that has a major impact on the problem solver's conception of the problem. Second, insight is *sudden*: It usually constitutes a quick transition from a state of 'not knowing' to a state of 'knowing'. Third, the new understanding is *more appropriate*: Even when insight does not directly point to the solution, it leads to grasping essential features of the problem that were not considered previously.

In experimental psychology, insight is often elicited using 'insight problems' [7]-[8], [20]. Such problems are diverse and characterized by the absence of direct, incremental algorithms allowing for their solutions. In many cases, they are selected because they have been shown to produce insight solutions in previous studies [8]. Empirically, insight is typically identified by a strong discontinuity in the subjective 'feeling of knowing' or the progress made in a verbal report [20].

III. The Cognitive Neuroscience of Creative Problem Solving

The search for a neuroscientific account of creative problem solving has been particularly elusive. While many authors have argued for a theory of right-hemisphere dominance [8], an extensive review of the literature conducted by Dietrich and Kanso [21] found that many studies did not support the right-hemisphere dominance hypothesis. The only brain area that consistently shows task-related activation is the prefrontal cortex and, even in this case, it is unclear what

kind of change or which specific structure is activated. For instance, many electroencephalography (EEG) studies reported task-related changes in the lower portion of the alpha band (8-12 Hz) in the prefrontal cortex [22]-[23]. However, many other studies did not find task-related changes in the alpha band measured in the prefrontal cortex [21]. A similarly confusing picture emerges in temporal and parietal cortices [24]. As such, Dietrich and Kanso concluded that "not a single currently circulating notion on the possible neural mechanisms underlying creative thinking survives close scrutiny." (p. 845).

One possible reason for this confusion may be that searching for the locus of creative thinking or creative problem solving is ill-defined and too big an endeavor, and that one should instead focus on sub-processes involved in creative problem solving [24]. For instance, insight research has focused on conflict resolution, and the results are more consistent: The anterior cingulate cortex plays an important role in insight, and the superior temporal gyrus plays a role at least in verbal problems [21]. Decomposing creative problem solving into a series of sub-processes to facilitate cognitive neuroscience explorations is the main motivation behind the current work.

IV. The Explicit-Implicit Interaction Theory

The Explicit–Implicit Interaction (EII) theory [6] constitutes an attempt at integrating and unifying existing theories of creative problem solving in two different senses. First, most theories of creative problem solving have focused on either a high-level stage decomposition [12] or on a process explanation of only one of the stages [25]. Second, the process theories of incubation and insight are often incomplete and sometimes mutually incompatible. EII attempts to integrate the existing theories to makes them more complete in order to provide a detailed description of the subprocesses involved in key stages of creative problem solving. EII starts from Wallas' stage decomposition of creative problem solving and provides a detailed process-based explanation sufficient for a coherent computational implementation. In this section, we present the core assumptions underlying the EII theory. Ample justification of the principles, and details on how EII can account for existing theories of incubation and insight, can be found in [6].

A. Principle #1: The co-existence of and the difference between explicit and implicit knowledge

The EII theory assumes the existence of explicit and implicit knowledge residing in two separate modules. Explicit knowledge is easier to access and verbalize and often said to be composed of symbols following hard constraints. Using explicit knowledge requires extensive attentional resources. In contrast, implicit knowledge is relatively inaccessible, harder to verbalize, often "subsymbolic", and follows soft constraints. Using implicit knowledge does not require much attentional resources. Because of these differences, explicit and implicit knowledge is processed differently. According to the EII theory, explicit processes perform some form of rule-based reasoning (in a very generalized sense) and represents relatively crisp and exact processing (often involving hard constraints), whereas implicit processing is 'associative' and represents soft-constraint satisfaction.

B. Principle #2: The simultaneous involvement of implicit and explicit processes in most tasks

Explicit and implicit processes are involved simultaneously in most tasks under most circumstances. This can be useful because different representations and processing are used to describe the two types of knowledge. As such, each type of processes can end up with similar or conflicting conclusions that contribute to the overall output.

C. Principle #3: The redundant representation of explicit and implicit knowledge

According to the EII theory, explicit and implicit knowledge is often redundant. In many cases, explicit and implicit knowledge can amount to re-descriptions of one another in different representational forms. For example, knowledge that is initially implicit is often later re-coded to form explicit knowledge (through "bottom-up learning"). Likewise, knowledge that is initially learned explicitly (e.g., through verbal instructions) is often later assimilated and re-coded into an implicit form, usually after extensive practice (top-down assimilation). There may also be other ways redundancy is created, e.g., through simultaneous learning of implicit and explicit knowledge. Redundancy often leads to interaction.

D. Principle #4: The integration of the results of explicit and implicit processing

Although explicit and implicit knowledge are often re-descriptions of one another, they involve different forms of representation and processing, which may produce similar or different conclusions. The integration of these conclusions can lead to synergy (e.g., overall better performance and faster learning). EII assumes that this synergy is an important component of creative problem solving.

E. Principle #5: The iterative (and possibly bidirectional) processing

Processing is often iterative and potentially bidirectional according to the EII theory. If the integrated outcome of explicit and implicit processing does not yield a definitive result (i.e., a result in which one is highly confident) and if there is no time constraint, another round of processing may occur, which uses the integrated outcome as part of the new input. Reversing the direction of reasoning may sometimes carry out this process (e.g., abductive reasoning). Alternating between forward and backward processing has been argued to happen also in everyday human reasoning.

F. Accounting for creative problem solving using EII

The preceding assumptions allow for a conceptual model that captures the four stages of Wallas' analysis of creative problem solving [12] (see Fig. 1). First, Wallas described the preparation stage as involving "the whole traditional art of logic" (p. 84). Hence, the preparation stage is mainly

captured by explicit processing in the EII theory. This is justified because explicit knowledge is usually rule-based (*Principle #1*), which includes logic-based reasoning as a special case. Also, the preparation stage has to be explicit in EII because peoples are responding to (explicit) verbal instructions, forming representations of the problem, and setting goals.

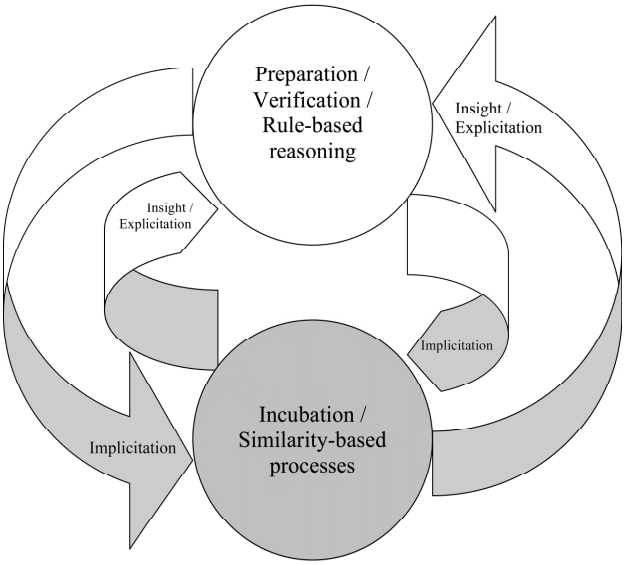

Fig. 1. Information flow in the EII theory. The grey sections are implicit while the white sections are explicit.

In contrast, incubation relies more heavily on implicit processes in EII. According to Wallas, incubation is the stage during which "we do not voluntarily or consciously think on a particular problem" (p. 86). This is consistent with EII's account of the difference in conscious accessibility between explicit and implicit knowledge (*Principle #1*). Moreover, incubation can persist implicitly for an extended period of time in Wallas' theory. This characteristic of incubation corresponds well with the above-mentioned hypothesis concerning the relative lack of attentional resource requirement in implicit processing.

The third stage, insight, is "the appearance of the 'happy idea' together with the psychological events which immediately preceded and accompanied that appearance" (p. 80). In EII, insight is obtained by the process of explicitation, which makes the output available for verbal report. It is worth noting that the intensity of insight is continuous [8], [20]. Correspondingly, explicitation is continuous in the EII theory (using an 'internal confidence level' or ICL) [6]. In particular, when the ICL of an output barely crosses the explicitation threshold, the output is produced but does not lead to an intense "Aha!" experience. In contrast, when the ICL of an output suddenly becomes very high and crosses the explicitation threshold, a very intense experience can result. According to the EII theory, intense insight experiences most likely follow the integration of implicit and explicit knowledge, as it can lead to a sudden large increase of the ICL caused by synergy (*Principle #4*).

Finally, the verification phase "closely resembles the first stage of preparation" (pp. 85-86): it should thus involve mainly explicit processing according to the EII theory. In addition, environmental feedback can be used in place of rule-based verification (when available). Regardless of how verification is accomplished, if it suggests that the insight solution might be incorrect, the whole process is repeated by going back to the preparation stage (*Principle #5*). In that case, EII predicts that the preparation stage can produce new information, because the knowledge state has been modified by the previous iteration of processing (e.g., some hypotheses may have been discarded as 'inadequate' or abductive reasoning might bring a new interpretation of the data).

V. A Connectionist Implementation of the Explicit-Implicit Interaction theory

The EII theory of creative problem solving has been implemented using the CLARION cognitive architecture [6]. The general structure of the model resulting from EII (implemented in the Non-Action-Centered Subsystem of CLARION) is shown in Fig. 2. The model is composed of two major modules, representing explicit and implicit knowledge respectively. These two modules are connected through bidirectional associative memories (i.e., the **E** and **F** weight matrices). In each trial, the task is simultaneously processed in both modules, and their outputs (response activations) are integrated in order to determine a response distribution. Once this distribution is specified, a response is stochastically chosen and the statistical mode of the distribution is used to estimate the internal confidence level (ICL). If this measure is higher than a predefined threshold, the chosen response is output; otherwise, another iteration of processing is done in both modules, using the chosen response as new the input.

In the model, explicit processing is captured using a two-layer linear connectionist network while implicit processing is captured using a non-linear attractor neural network [26]. The inaccessible nature of implicit knowledge may be captured by distributed representations in an attractor neural network, because units in a distributed representation are capable of accomplishing tasks but are less individually meaningful. This characteristic corresponds well with the relative inaccessibility of implicit knowledge. In contrast, explicit knowledge may be captured in computational modeling by localist representations, because each unit in a localist representation is more easily interpretable and has a clearer conceptual meaning. This characteristic captures the property of explicit knowledge being more accessible and manipulable. This difference in the representation of the two types of knowledge leads to a dual-representation, dual-process model. Below we present some key equations describing the implementation. The learning rules used to create the weight matrices can be found in the Appendix. Additional details can be found in [6].

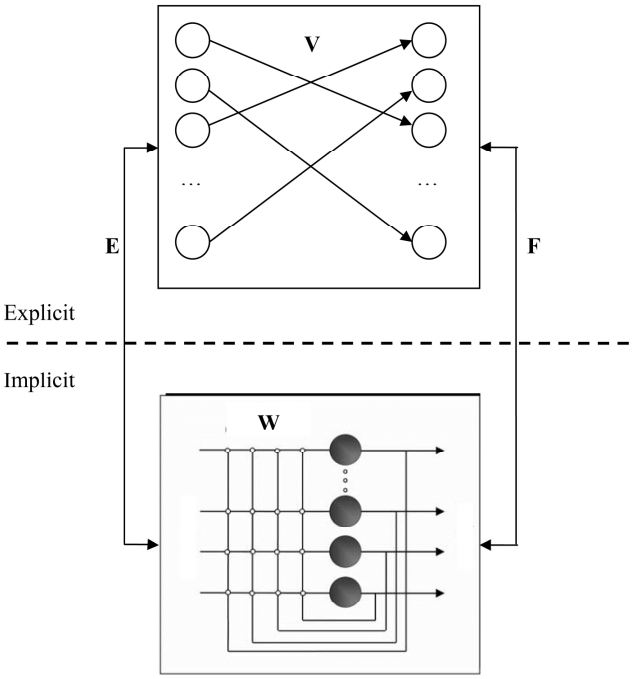

Fig. 2. General architecture of the connectionist model. The model is implemented in the Non-Action-Centered Subsystem of CLARION [11]. The upper-case letters correspond to the weight matrices.

The key process for explicit processing is described by a simple linear connectionist network:

$$\mathbf{y} = \mathbf{NVx} \tag{1}$$

where $\mathbf{y} = \{y_1, y_2, \ldots, y_m\}$ is a binary vector corresponding to the activation in the right layer in the top level of Fig. 2, $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ is a binary vector corresponding to the activation in the left layer in the top level of Fig. 2, **V** is a binary weight matrix connecting **x** and **y**, and **N** is a diagonal matrix normalizing the activation of **y**. The weight matrix in the top-level (i.e., **V**) encodes (mostly) pre-existing explicit rules/associations and is learned using a 'one-shot' Hebbian rule (Eq. A1).

In the bottom level, activation propagation is non-linear and implemented using the NDRAM network [26]:

$$\mathbf{z}_{[t+1]} = f(\mathbf{W}\mathbf{z}_{[t]}), \qquad f(z_i) = \begin{cases} +1 & , z_i > 1 \\ (\delta+1)z_i - \delta z_i^3, & -1 \le z_i \le 1 \\ -1 & , z_i < -1 \end{cases} \tag{2}$$

where $\mathbf{z}_{[t]} = \{z_1, z_2, \ldots, z_r\}$ is the bottom-level activation after t iterations in the network, **W** is the bottom-level weight matrix, and $0 < \delta < 0.5$ is the slope of the transmission function. The bottom-level weight matrix (i.e., **W**) represents implicit associations and is learned iteratively using a contrastive Hebbian learning rule (Eq. A2). The settling process described by Eq. 2 amounts to a search through a soft constraint satisfaction process, where each connection represents a constraint and the weights represent the importance of the constraints.

The last key equation is used for knowledge integration. Once the response activations have been computed in both levels, they are integrated using the *Max* function:

$$o_i = Max\left[y_i,\ \lambda(k_i)^{-1.1} \sum_{j=1}^{r} f_{ji} z_j \right] \tag{3}$$

where $\mathbf{o} = \{o_1, o_2, \ldots, o_m\}$ is the integrated response activation, **y** is the result of top-level processing (Eq. 1), z is the output of bottom-level processing (Eq. 2), **F** is a weight matrix connecting the bottom level to the top level, λ is a scaling parameter specifying the relative weight of bottom-level processing, and k_i is the number of nodes in the bottom level (in **z**) that are connected to y_i. The **F** weight matrix represents the associations between a conceptual (explicit) representation in the right layer of the top-level and its corresponding implicit representation in the bottom level. This association is pre-existing and learned using a 'one-shot' Hebbian rule (Eq. A4).

The result of Eq. 3 is normalized using a Boltzmann equation, and a response is stochastically selected. The statistical mode of the Boltzmann distribution is computed to estimate the ICL. This measure represents the relative support for the most likely response (which may or may not be the stochastically selected response). In the current model, the chosen response is output if the ICL is higher than threshold ψ. However, if the ICL is smaller than ψ, the search process continues with a new iteration using the chosen response to activate the left layer in the top level ($\mathbf{x} = \mathbf{V}^\mathbf{T}\mathbf{o}$ and $\mathbf{z} = \mathbf{Ex}$, where **E** is another weight matrix connecting the top level to the bottom level). More specifically, the **E** weight matrix represents the associations between a conceptual (explicit) representation in the left layer of the top-level and its corresponding implicit representation in the bottom level. This association is pre-existing and learned using a 'one-shot' Hebbian rule (Eq. A3). The algorithm specifying the complete process is summarized in Table I.

TABLE I: ALGORITHM OF THE CONNECTIONIST MODEL

1. Observe the current state of the environment;
2. Compute the response activations in each level;
3. Compute the integrated response activation and the resulting response distribution;
4. Stochastically choose a response and compute the statistical mode of the response distribution:
   a. If the mode is higher than ψ, output the response;
5. Else, if there is time remaining, go back to step 2.

VI. GROUNDING THE EXPLICIT-IMPLICIT INTERACTION THEORY IN NEUROSCIENCE

The EII theory, or its implementation in the CLARION cognitive architecture, does not specify any neurobiological process. However, they specify cognitive mechanisms that are involved in each one of the stages of creative problem solving. For instance, the preparation and verification stages involve hard constraints and rule following. They also involve explicit retrieval of memory for reasoning. As such,

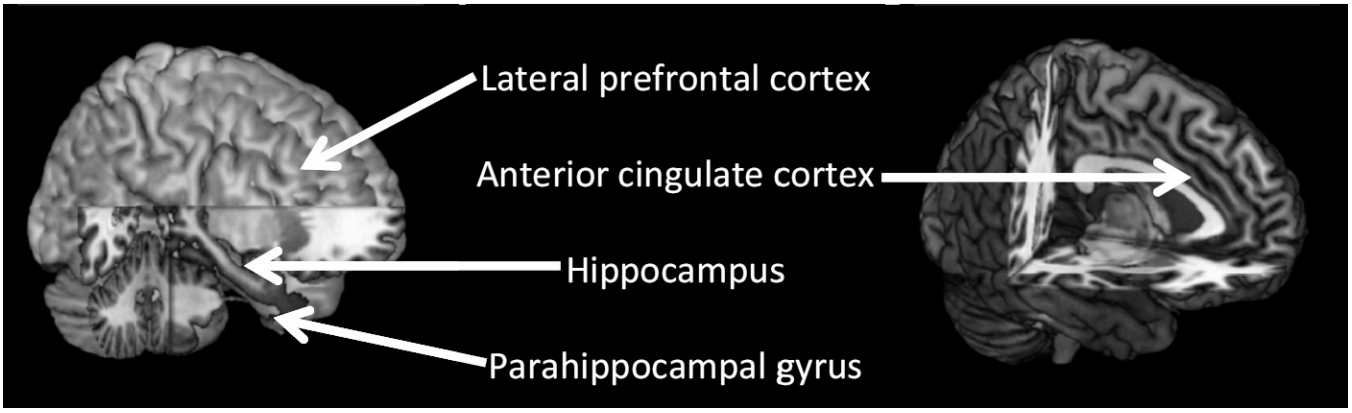

Fig. 3. The Neuro-EII theory of creative problem solving. The lateral prefrontal cortex and the hippocampus are involved in the preparation and verification stages, the parahippocampal gyrus is involved in the incubation stage, and the anterior cingulate cortex is involved in insight.

we would expect that these two stages involve mostly lateral prefrontal cortex (for rules) [27] and the hippocampus (for explicit memory retrieval) [28]-[29]. In contrast, the incubation stage involves mostly implicit memory retrieval, so it should rely on the parahippocampal gyrus (e.g., parahippocampal cortex, perirhinal cortex) [28], [30]-[31], and possibly the 'default' network [32]. Finally, the insight stage corresponds to crossing an uncertainty threshold, so it should rely on the anterior cingulate cortex [33]. This later assignment is interesting considering that some theories of insight have suggested that incubation makes us more sensitive to external cues, and Hashimoto and colleagues [31] have shown that the anterior cingulate cortex is active when implicit cues are available. The resulting biological model is shown in Fig. 3.

The Neurobiological version of EII (called Neuro-EII) is fully compatible with the psychological version of EII. First, different regions have been assigned to explicit and implicit memory retrieval (Principle #1). This separation is functional (i.e., based on the role of these two memory systems), but does not assume that there is a strict one-to-one mapping between the brain areas and the psychological functions. Second, the brain has no early gating mechanism so both explicit and implicit processes will compete to achieve the task in every trial (Principle #2). Third, the embodied theory of semantic representation [34] suggests that the same concepts are represented in different brain locations with different degrees of abstractness (Principle #3). This is fully consistent with the different memory and rule structures included in Neuro-EII. Fourth, activation from almost all of posterior cortex is integrated in the prefrontal cortex, and lateral prefrontal cortex (used to generate a decision) is highly interconnected with the anterior cingulate cortex (important for insight; Principle #4). Finally, most connections to the prefrontal cortex are both afferent and efferent, which allows for iterative processing using an elaborate feedback system (Principle #5). Hence, Neuro-EII is fully compatible with the principles underlying the psychological version of EII.

VII. Example Application

The CLARION implementation of the EII theory has been used to simulate many different tasks [6]. By looking at the conceptual explanation and relevant parameters, Neuro-EII allows for making predictions regarding what brain areas should be involved in the simulated tasks. For example, CLARION was used to simulate the effect of incubation in a lexical decision task.

Yaniv and Meyer [19] showed participants word definitions that were weakly associated with their definiendums (e.g., sextant). The participants had a limited time to find each definition's definiendum (i.e., the rare-word association task). If the participant found the definiendum, they were transferred to a lexical decision task where they had to classify briefly presented strings of letters as 'word' or 'non-word'. If the participant did not produce a definiendum, they were asked to rate their feeling of knowing (FOK) and then continued with the lexical decision task. The elapsed time between the rare-word association task and the lexical decisions task was interpreted as incubation [19]. A flow chart describing a trial is shown in Fig. 4. The results show that definitions that allowed for the retrieval of the correct definiendums or generated high FOKs produced priming (i.e., faster reaction times) for the target word in the lexical decision task.

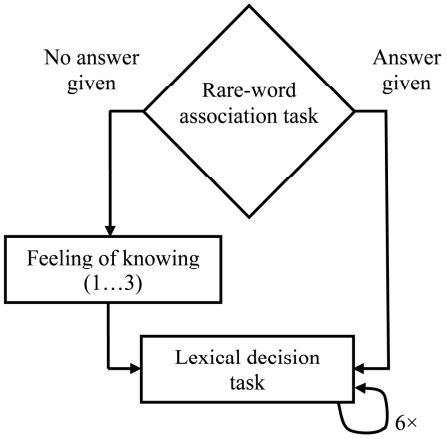

Fig. 4. Flow chart representing a trial in [19].

According to the EII theory, a rare-word association trial produces a simultaneous search in explicit and implicit memories (*Principle #2*). Because the target association is rare, explicit memory search is not likely to yield a satisfactory solution within the allotted time (i.e., the existing set of hard constraints do not necessarily lead to solutions). In contrast, implicit memory search is more likely to retrieve the desired association if given enough time, because soft constraint satisfaction can allow for a partial match that can be iteratively improved. However, implicit memory search is often cut short by the experimenter who then asks the participant to take part in lexical decision trials. At the beginning of the lexical decision trials, implicit knowledge is still in the same state as it was at the end of the corresponding rare-word association trial. Hence, if the association was retrieved or nearly retrieved during the rare-word association trial (i.e., with high FOK), the memory search is not wasted and the target word is primed for the lexical decision trials. In contrast, the correct recognition of unrelated words (distractors) is not affected by the previous state of implicit knowledge in the lexical decision trials, because the cognitive work during the corresponding rare-word association trial was irrelevant. This conceptual explanation by EII led to a detailed computational model that produced simulation in line with Yaniv and Meyer's results [19]. The results of 3,000 simulations with a CLARION-based model are shown in Fig. 5.

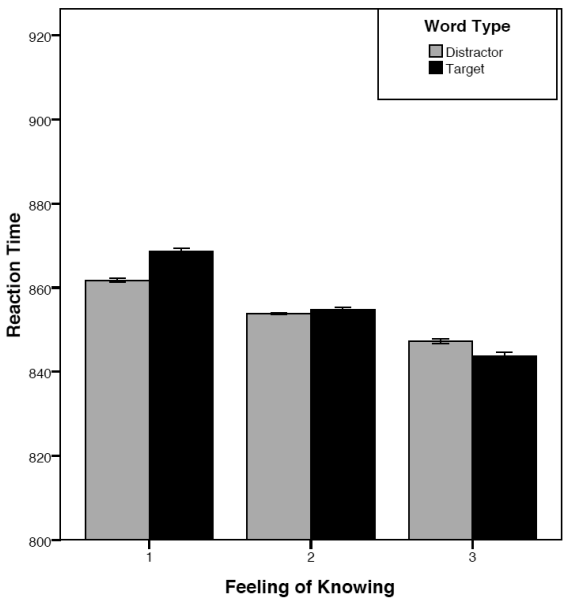

Fig. 5. Simulated response times in the lexical decision task for participants who did not produce a definiendum in the rare-word association task. Details can be found in [6].

The task explanation of EII suggests that this lexical decision task will involve activation in the lateral prefrontal cortex and the hippocampus (especially during the definition task), and the parahippocampal gyrus (especially during the lexical decision task). Importantly, activation in the anterior cingulate cortex should become predictive of the FOK as soon as the definition task ends, and this activation should continue while the lexical decision task is achieved. Future work should be devoted to collecting this data to verify Neuro-EII's predictions. However, note that there is no claim that Neuro-EII is a complete model of the task. For instance, activation in the occipital lobe is expected (since the participants are looking at words and definitions). Likewise, the superior temporal gyrus is likely to be active because of its role in language processing. Cognitive processes involved in vision and reading were not included in Neuro-EII, but they are certainly compatible. Importantly, the predictions related to creative problem solving were made possible by the unique integrative nature of Neuro-EII, and no other existing theory of creative problem solving can make these predictions *a priori*.

VIII. Conclusion

This article introduced a new integrative biological theory of creative problem solving. The Neuro-EII theory is built on the EII theory and its CLARION implementation and allows for a micro-process decomposition of creative problem solving and testable predictions of the brain circuits underlying the process. Future neuroimaging work is required to test the cognitive neuroscience data predicted by Neuro-EII, and a spiking version of the CLARION implementation should be created to simulate the new data [35]. It is important to note that, in its current state, Neuro-EII can predict brain areas that are involved in a given task, but not brain areas that are not involved in the task. As such, the theory shown in Fig. 3 should not be interpreted as a complete model of brain activity in creative problem solving. Yet, Neuro-EII constitutes an important first step towards a more complete understanding of the cognitive neuroscience of creative problem solving, and we hope that the predictions made by Neuro-EII will generate more consistent research on the cognitive neuroscience of creative problem solving.

References

[1] F. G. Ashby, L. A. Alfonso-Reese, A. U. Turken, and E. M. Waldron, "A neuropsychological theory of multiple systems in category learning," *Psychological Review*, vol. 105, pp. 442-481, 1998.

[2] J. B. T. Evans, "The heuristic-analytic theory of reasoning: Extension and evaluation," *Psychonomic Bulletin & Review*, vol. 13, pp. 378-395, 2006.

[3] A. S. Reber, "Implicit learning and tacit knowledge," *Journal of Experimental Psychology: General*, vol. 118, pp. 219-235, 1989

[4] R. Sun, *Integrating Rules and Connectionism for Robust Commonsense Reasoning*. New York: John Wiley and Sons, 1994.

[5] S. Helie, R. Proulx, and B. Lefebvre, "Bottom-up learning of explicit knowledge using a Bayesian algorithm and a new Hebbian learning rule," *Neural Networks*, vol. 24, pp. 219-232, 2011.

[6] S. Helie and R. Sun, "Incubation, insight, and creative problem solving: A unified theory and a connectionist model," *Psychological Review*, vol. 117, pp. 994-1024, 2010.

[7] J. Dorfman, V. A. Shames, and J. F. Kihlstrom, "Intuition, incubation, and insight: Implicit cognition in problem solving," in *Implicit Cognition*, G. Underwood, Ed. New York: Oxford University Press, 1996, pp. 257-296.

[8] E. M. Bowden, M. J. Beeman, J. Fleck, and J. Kounios, "New approaches to demystifying insight," *Trends in Cognitive Sciences*, vol. 8, pp. 322-328, 2005.

[9] W. Duch, "Computational creativity," in *Proceedings of the International Joint Conference on Neural Networks*. Vancouver, 2006, pp. 435-442.
[10] S. Helie (Ed.), *The Psychology of Problem Solving: An Interdisciplinary Approach*. New York: Nova Publishers, 2013.
[11] R. Sun, *Duality of the Mind: A Bottom-Up Approach Toward Cognition*. Mahwah: Lawrence Erlbaum Associates, 2002.
[12] G. Wallas, *The Art of Thought*. New York: Franklin Watts, 1926.
[13] R. A. Dodds, T. B. Ward, and S. M. Smith, "A review of experimental literature on incubation in problem solving and creativity," in *Creativity Research Handbook. Vol. 3*, M.A. Runco, Ed. Cresskill: Hampton Press, 2003.
[14] K. Duncker, "On problem solving. *Psychological Monographs*," vol. 58, pp. 1-113, 1945.
[15] W. Kohler, *The Mentality of Apes*. New York: Liveright, 1925.
[16] R. A. Dodds, S. M. Smith, and T. B. Ward, "The use of environmental clues during incubation," *Creativity Research Journal,* vol. 14, pp. 287-304, 2002.
[17] S. Helie, R. Sun, and L. Xiong, "Mixed effects of distractor tasks on incubation," in *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, B.C. Love, K. McRae, and V.M. Sloutsky, Eds. Austin: Cognitive Science Society, 2008, pp. 1251-1256.
[18] S. M. Smith, and E. Vela, "Incubated reminiscence effects," *Memory & Cognition*, vol. 19, pp. 168-176, 1991.
[19] I. Yaniv and D. E. Meyer, "Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 13, pp. 187-205, 1987.
[20] A. J. K. Pols, *Insight Problem Solving*. Doctoral Dissertation, Department of Psychology, University of Utrecht, Netherlands, 2002.
[21] A. Dietrich, and R. Kanso, "A review of EEG, ERP, and neuroimaging studies of creativity and insight," *Psychological Bulletin*, vol. 136, pp. 822-848, 2010.
[22] A. Fink, R. Grabner, M. Benedek, and A. Neubauer, "Divergent thinking training is related to frontal electroencephalogram alpha synchronization," *European Journal of Neuroscience*, vol. 23, pp. 2241–2246, 2006.
[23] O. M. Razumnikova, N. Volf, and I. V. Tarasova, "Strategy and results: Sex differences in electrographic correlates of verbal and figural creativity," *Human Physiology*, vol. 35, pp. 285–294, 2009.
[24] A. Fink, R. H. Grabner, M. Benedek, G. Reishofer, V. Hauswirth, M. Fally, C. Neuper, F. Ebner, and A. C. Neubauer, "The creative brain: investigation of brain activity during creative problem solving by means of EEG and FMRI," *Human brain mapping*, vol. 30, pp. 734-748, 2009.
[25] T. I. Lubart, "Models of the creative process: Past, present and future," *Creativity Research Journal*, vol. 13, pp. 295-308, 2001.
[26] S. Chartier and R. Proulx, "NDRAM: Nonlinear Dynamic Recurrent Associative Memory for learning bipolar and non-bipolar correlated patterns," *IEEE Transactions on Neural Networks*, vol. 16, pp. 1393-1400, 2005.
[27] S. A. Bunge, and J. D. Wallis, eds, *Neuroscience of Rule-Guided Behavior*. Oxford University Press, 2007.
[28] J. Yang, A. Meckingler, M. Xu, Y. Zhao, and X. Weng, "Decreased parahippocampal activity in associative priming: Evidence from an event-related fMRI study," *Learning & Memory*, vol. 15, pp. 703–710, 2008.
[29] M. G. Packard and B. J. Knowlton, "Learning and memory functions of the Basal Ganglia," *Annual Review of Neuroscience*, vol. 25, pp. 563–93, 2002.
[30] W.-C. Wang, M. M. Lazzara, C. Ranganath, R. T. Knight, and A. P. Yonelinas, "The medial temporal lobe supports conceptual implicit memory.," *Neuron*, vol. 68, no. 5, pp. 835–42, 2010.
[31] T. Hashimoto, S. Umeda, and S. Kojima, "Neural substrates of implicit cueing effect on prospective memory.," *NeuroImage*, vol. 54, no. 1, pp. 645–52, 2011.
[32] J. Yang, X. Weng, Y. Zang, M. Xu, and X. Xu, "Sustained activity within the default mode network during an implicit memory task.," *Cortex*, vol. 46, no. 3, pp. 354–66, 2010.
[33] W. H. Alexander and J. W. Brown, "Medial prefrontal cortex as an action-outcome predictor.," *Nature Neuroscience*, vol. 14, no. 10, pp. 1338–44, 2011.
[34] J. R. Binder and R. H. Desai, "The neurobiology of semantic memory.," *Trends in Cognitive Sciences*, vol. 15, no. 11, pp. 527–36, 2011.
[35] F. G. Ashby and S. Helie, "The Neurodynamics of Cognition: A Tutorial on Computational Cognitive Neuroscience," *Journal of Mathematical Psychology*, vol. 55, pp. 273–289, 2011.
[36] T. Kohonen, "Correlation matrix memories," *IEEE Transactions on Computers C*, vol. 21, pp. 353-359, 1972.

APPENDIX

In the top level, explicit knowledge is represented using weight matrix $\mathbf{V} = [v_{ij}]$, which was trained to encode the explicit rules using 'one-shot' Hebbian learning [36]:

$$\mathbf{V} = \sum_i \mathbf{y}_i \mathbf{x}_i^{\mathbf{T}} \tag{A1}$$

where $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k\}$ and $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_k\}$ are the sets containing the stimuli ($k \le n$ and $k \le m$), and $\mathbf{x}_i$ is associated to $\mathbf{y}_i$. The use of Hebbian learning to encode the rules ensures that $v_{ij} = 1$ if $\mathbf{x}_i$ is associated to $\mathbf{y}_j$ and zero otherwise (because the stimuli are binary; see Section V).

In the bottom level, implicit knowledge is represented by the **W** weight matrix, which is pre-trained to encode the implicit associations using a contrastive Hebbian learning rule [26]:

$$\mathbf{W}_{[t]} = \zeta \mathbf{W}_{[t-1]} + \eta\left(\mathbf{z}_{i[0]}\mathbf{z}_{i[0]}^{\mathbf{T}} - \mathbf{z}_{i[p]}\mathbf{z}_{i[p]}^{\mathbf{T}}\right) \tag{A2}$$

where $\mathbf{W}_{[t]}$ is the weight matrix at trial t, $0 << \zeta \le 1$ is a memory efficiency parameter, and $0 < \eta < \frac{1}{2(1-2\delta)r}$ is a general learning parameter (where r is the number of units in the bottom level; for a demonstration, see [26]). Note that Eq. A2 is the only iterative learning algorithm in this implementation of CLARION.

The associations between the top- and bottom-level representations are encoded using the **E** and **F** weight matrices. These matrices are trained using the same linear Hebbian rule as **V**:

$$\mathbf{E} = \sum_i \mathbf{t}_{\mathbf{1}i} \mathbf{x}_i^{\mathbf{T}} \tag{A3}$$

$$\mathbf{F} = \sum_j \mathbf{t}_{\mathbf{2}j} \mathbf{y}_j^{\mathbf{T}} \tag{A4}$$

where $\mathbf{T_1} = \{\mathbf{t}_{11}, \mathbf{t}_{12}, \ldots, \mathbf{t}_{1k}\}$ and $\mathbf{T_2} = \{\mathbf{t}_{21}, \mathbf{t}_{22}, \ldots, \mathbf{t}_{2k}\}$ are the sets containing the distributed representations of the top-level concepts located in **x** and **y** (respectively).