Creative Problem Solving: A CLARION theory

Sebastien Helie, Ron Sun

Abstract—**Psychological theories of problem solving have largely focused on explicit processes that gradually bring the solver closer to the solution step-by-step in a mostly explicit and deliberative way. This approach to problem solving is typically inefficient or ineffective when the problem is too complex, ill-understood, or ambiguous. In such a case, a 'creative' approach to problem solving might be more appropriate. We propose a computational psychological model implementing the Explicit-Implicit Interaction theory of creative problem solving (i.e., the CLARION theory of creative problem solving) that centers on the interaction of implicit and explicit processing. The model based on the CLARION theory has been used to simulate a variety of empirical psychological data sets.**

I. Introduction

According to Wallas, a "single achievement of thought" is a four stage process [1]. The first stage, preparation, is the accumulation of knowledge that allows for directly solving the problem at hand (i.e., in a sequential, logical way). However, this approach to problem solving sometimes fails to provide a satisfying solution and an impasse is reached. Surprisingly, if the problem solver puts the problem aside and directs his/her attention to some other task, s/he is more likely to later find the correct solution [2]-[3]. This phenomenon, which has been the topic of countless anecdotes in the history of science, is thought to result from Wallas' second stage: incubation. Because the solution often comes as a surprise to the solver, the stage following the incubation period was called illumination (or insight), during which the problem solver has the impression that the solution completely elucidates the problem. However, the quality of the solution remains to be evaluated by the fourth stage: the validation process.

This paper focuses on the incubation and insight stages. A connectionist model of creative problem solving based on the *Explicit-Implicit Interaction* (EII) theory [4]-[5] (resulting from the CLARION cognitive architecture [6]-[10]) is sketched, and example psychological simulations are presented to show the ability of the model to simulate incubation and insight data.

Manuscript received May 2, 2010. This research was supported by a postdoctoral research fellowship from *Le Fonds Québecois de la Recherche sur la Nature et les Technologies* to the first author and the research grants provided by *Army Research Institute* W74V8H-04-K-0002 and *Office of Naval Research* N00014-08-1-0068 to the second author.

S. Helie is with Psychology Department, University of California, Santa Barbara, CA 93106-9660 USA (phone: 805-893-7909; fax: 805-893-4303; e-mail: helie@psych.ucsb.edu).

R. Sun is with Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180-3590 USA (e-mail: rsun@rpi.edu).

II. The CLARION Cognitive Architecture

CLARION is a cognitive architecture that is, in part, based on two basic assumptions: representational differences and learning differences of two different types of knowledge: implicit versus explicit [5]-[8]. These two types of knowledge differ as to their accessibility and attentional requirements. The top level of CLARION (as in Fig. 1) contains explicit knowledge (easily accessible, requiring more attentional resources) whereas the bottom level contains implicit knowledge (harder to access, mostly automatic). Because knowledge in the top and bottom levels is different, Sun et al. have shown that integrating the results of top- and bottom-level processing captures the interaction of implicit and explicit processing in humans [5]-[8].

CLARION is further divided into two different subsystems (see Fig. 1): the *Action-Centered Subsystem* and the *Non-Action-Centered Subsystem*. The Action-Centered Subsystem (with both levels) contains procedural knowledge concerning actions and procedures (i.e., it serves as the long-term procedural memory), while the Non-Action-Centered Subsystem (with both levels) contains declarative knowledge [6], [8]. The Non-Action-Centered Subsystem is controlled by the Action-Centered Subsystem and constitutes another long-term memory (semantic/episodic). The Non-Action-Centered Subsystem is also used for reasoning [5], [9].

The second assumption in CLARION concerns the existence of different learning processes in the top and bottom levels [6]-[8]. In the bottom level, implicit associations are learned through gradual trial-and-error learning. In contrast, learning of explicit rules is often "one-shot" and represents the abrupt availability of explicit knowledge following "explicitation" or newly acquired linguistic information in the top level. The inclusion and emphasis on bottom-up learning (i.e., the transformation of implicit knowledge into explicit knowledge) is, in part, what distinguishes CLARION from other cognitive models.

III. The Explicit-Implicit Interaction (EII) Theory

The Explicit-Implicit Interaction theory [4]-[5] is a psychological theory of creative problem solving derived from CLARION that relies mainly on the following principles: (a) there are two types of knowledge, namely explicit and implicit (a well established distinction in psychology: see, e.g., [8], [11]-[12]), that are simultaneously involved in most tasks [6]. (b) Explicit and implicit knowledge are often "redundant" [7], and the results of explicit and implicit processing are often integrated to provide a response [6]. In psychology, implicit knowledge refers to knowledge that affects behavior without the

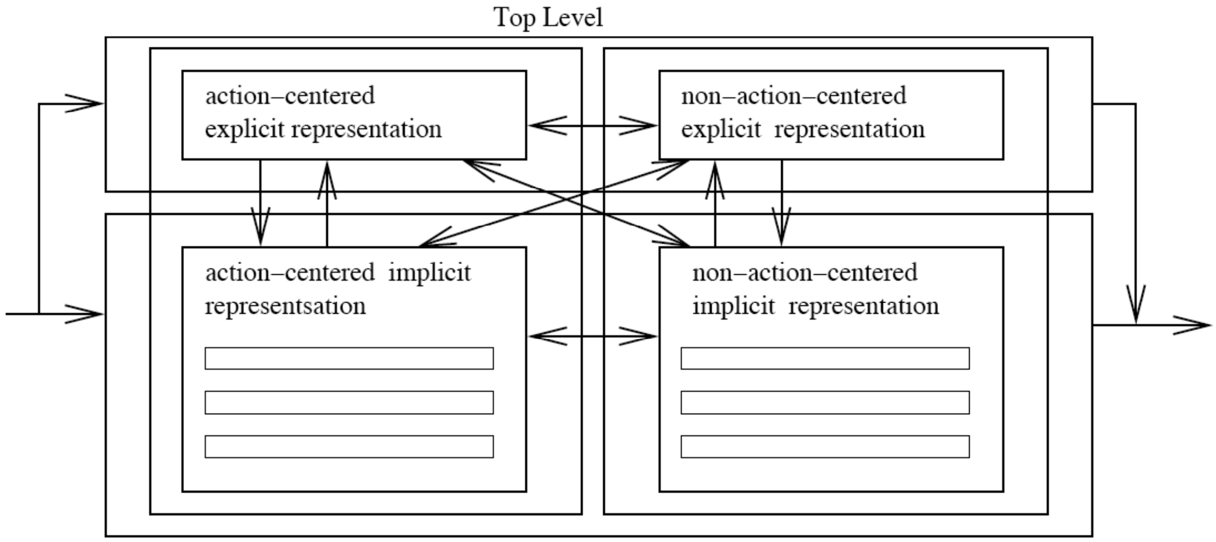

Fig. 1. A high-level representation of CLARION.

awareness of the participant while explicit knowledge may (potentially) come with the accompanying awareness [6], [10]-[11]. (For details and justifications of the principles, see [5]-[6].)

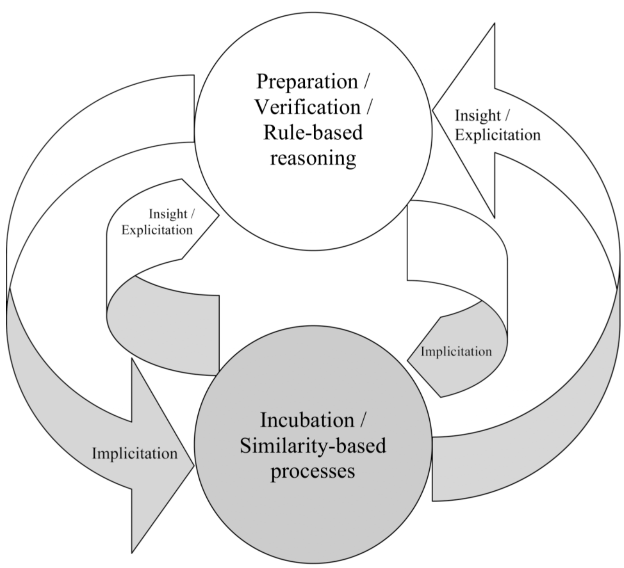

Fig. 2. Information flow in the EII theory. The grey sections are implicit while the white sections are explicit.

In the Explicit-Implicit Interaction theory, the preparation and verification stages of problem solving are mostly captured by explicit processing, while incubation corresponds mostly to implicit processing (see Fig. 2). According to this theory, Wallas' third stage, insight, involves an "explicitation" process of implicit processes and consequently produces a solution in an explicit form (with the integration of the two kinds of information). In addition, the Explicit-Implicit Interaction theory assumes the presence of activation thresholds that control the *Internal Confidence Level* (ICL) necessary to come up with a solution to the problem. If a threshold is not crossed after the results of explicit and implicit processing have been integrated, another iteration of processing is performed. This simple yet powerful iterative theory has been formalized into a connectionist model that was used to simulate a variety of tasks [5]. The model, implemented in the CLARION cognitive architecture, is described in *Section IV*. Simulations examples are presented in *Sections V* and *VI:* One for incubation and the other for insight.

IV. A Connectionist Model For Creative Problem Solving

The general structure of the model resulting from EII (implemented in the Non-Action-Centered Subsystem of CLARION) is shown in Fig. 3. 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; [13]). 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 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 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 (*NDRAM*: [14]). 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 [8]. This characteristic corresponds well with the relative inaccessibility of implicit knowledge [11]-[12]. 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 [6]. This difference in the representation of the two types of knowledge leads to a dual-representation, dual-process model. Theoretical arguments for such models are presented in, e.g., [8].[1]

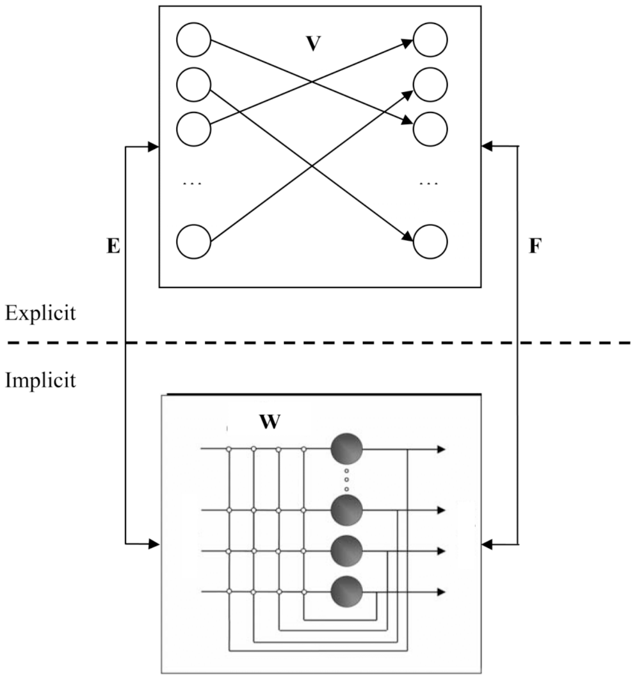

Fig. 3. General architecture of the connectionist model. The model is implemented in the Non-Action-Centered Subsystem of CLARION [5], [8]-[9].

Specifically, explicit knowledge is localistically represented in the top level using binary activation. The left layer in Fig. 3 (denoted **x**) is composed of n units while the right layer (denoted **y**) is composed of m units. These layers are connected using the binary weight matrix **V**, and the information is transmitted using the standard weighted sum (dot product, i.e., $\mathbf{y} = \mathbf{NVx}$, where **N** is a diagonal matrix normalizing the activation of **y**).[2]

In the bottom level, implicit knowledge is represented using r bipolar units (denoted **z**). Specifically, $\mathbf{z} = \mathbf{t_1} + \mathbf{t_2}$, where $\mathbf{t_1}$ represents the first s units in **z**, which are connected to the left layer in the top level using the **E** weight matrix. Meanwhile, $\mathbf{t_2}$ represents the remaining $r - s$ units in **z**, which are connected to the right layer in the top level using weight matrix **F**. In words, the **E** and **F** weight matrices are used to 'translate' explicit knowledge into implicit knowledge (i.e., 'implicitation') and vice-versa (i.e., 'explicitation').

[1] Note that *Section IV* proposes an algorithmic model [15]. Hence, the units do not represent biological neurons. The model serves as a constructive proof of the network/representation properties.

[2] In the model, all the weight matrices are learned using Hebbian learning. This type of learning has the advantage of psychological and biological plausibility. The **V**, **E**, and **F** weight matrices are learned using regular Hebbian learning (i.e., the outer matrix product). The bottom-level weight matrix (**W**) is learned using a contrastive Hebbian learning rule [14]. More details can be found in the appendix of [5].

Bottom-level activation (**z**) is modified through a settling process using the NDRAM transmission rule [14]:

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

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. This settling process 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 [16]. Note that it was estimated psychologically that each iteration in this network takes roughly 350 ms of psychological time (for justifications, see [5], [8]).

Once the response activations have been computed in both levels, they are integrated using the *Max* function:

$$o_i = Max\left[y_i,\ \lambda(k_i)^{-1.1}\sum_{j=1}^{r} f_{ji} z_j\right] \tag{2}$$

where $\mathbf{o} = \{o_1, o_2, \ldots, o_m\}$ is the integrated response activation, $\mathbf{y} = \{y_1, y_2, \ldots, y_m\}$ is the result of top-level processing, λ is a scaling parameter specifying the relative weight of bottom-level processing, k_i is the number of nodes in the bottom level (in **z**) that are connected to y_i ($k_i \leq r - s$), and $\mathbf{F} = [f_{ij}]$ is a weight matrix. The integrated response activation is then transformed into the Boltzmann response distribution:

$$P(o_i) = e^{o_i/\alpha}\left(\sum_j e^{o_j/\alpha}\right)^{-1} \tag{3}$$

where α is a noise parameter (i.e., the temperature). Note that low noise levels tend to exaggerate the probability differences, which lead to a narrow search of possible responses and favors stereotypical responses. In contrast, high noise levels tend to minimize the probability differences, which leads to a more complete search of the response space.

A response is stochastically chosen based on the response distribution (3) and the statistical mode of the 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 ($\mathbf{x} = \mathbf{V}^{\mathbf{T}}\mathbf{o}$; $\mathbf{z} = \mathbf{Ex}$).

The algorithm specifying the complete process is summarized in Table I. More details of the computational model can be found in [5]. *Section V* presents an example simulation of incubation while *Section VI* presents an example simulation of insight. Note that in all the simulations, the parameters were set to reasonable values to qualitatively reproduce the human data. No optimization technique was used.

TABLE I: ALGORITHM OF THE CONNECTIONIST MODEL

1.	Observe the current state of the environment;
2.	Compute the response activations;
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.

V. INCUBATION IN A FREE RECALL TASK

A. Experiment

Smith and Vela [17] studied the effect of incubation on the number of new words recalled during a second free recall test in a two-phased free recall experiment. This measure is referred to as 'reminiscence' in experimental psychology.

The participants had five minutes to memorize 50 line drawings. Following this study phase, the participants took part in the first free recall test, which lasted 1, 2, or 4 minutes. Once the first free recall test was completed, the participants had a 0-, 1-, 5-, or 10-minute break (which constituted the incubation phase). After the incubation phase, all the participants took part in a second free recall test. The length of the second free recall test was the same as the first (and based on the same set of line drawings seen earlier, without re-studying them). 221 participants were tested in this 3 × 4 design and the dependant variable was the reminiscence score.

B. Empirical Results

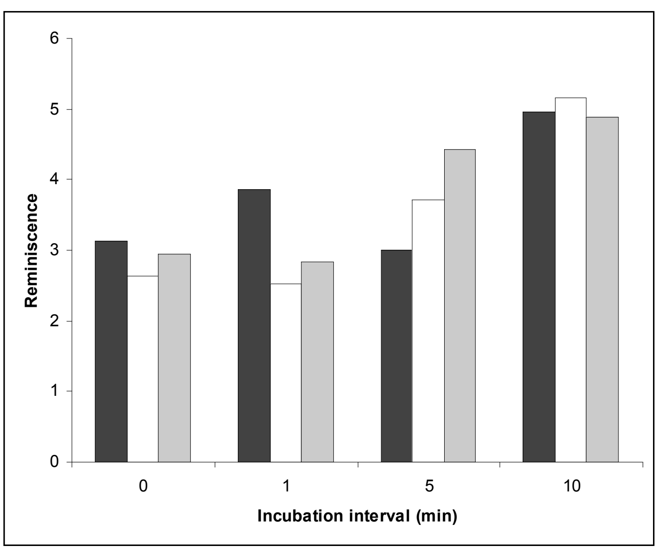

Fig. 4. Reminiscence effect found in [17]. The black bars represent 1-minute tests, the white bars represent 2-minute tests, and the grey bars represent 4-minute tests.

A Test Duration × Incubation Interval ANOVA was performed on reminiscence (see Fig. 4). There was no effect of Test Duration, $F(2, 209) = .27$, $p > .7$, but Incubation Interval had a significant effect on reminiscence, $F(3, 209) = 9.4$, $p < .01$. The mean reminiscence scores for each incubation interval were 2.90, 3.15, 3.72, and 5.00. *Post hoc* tests ($\alpha = .05$) showed that the first two incubation intervals (0 and 1 minute) yielded similar reminiscence scores, and that these scores were smaller than those obtained for longer incubation intervals (5 and 10 minutes respectively, which did not differed statistically). Subsequent experiments showed that the effect of Incubation Interval on reminiscence was significant only during the first minute of the second free recall test [17].

C. Simulation

To simulate this task with EII, only the right layer was used in the top level of the CLARION model (see Fig. 3) and each node represented a different line drawing (word).[3] In the bottom level, all the concepts (each represented by a top-level node) were encoded with a common pool of nodes using distributed representations ($r = 500$ and $s = 0$), and a different bottom-level distributed representation was randomly generated for each line drawing (word).[4]

To simulate the first free recall test, a random pattern activated the bottom level, and the activations were propagated within the bottom level until convergence of the attractor neural network. The resulting stable activation state activated the top-level right-layer nodes through bottom-up "explicitation", and the top-level activations were transformed into a distribution as explained before (as activations of the corresponding top-level right-layer nodes; $\alpha = 0.06$). The statistical mode of the distribution was used to estimate the ICL, which was compared with the threshold value ($\psi = 0.896$). If the ICL was sufficiently high, a word (node) was stochastically chosen for recall. Otherwise, no response was output, and a new random pattern was used to activate the bottom level and start the process again. Each spin in the bottom level was estimated to take 350 ms of psychological time. Hence, the durations of recall were 171, 343, and 686 spins (for 1, 2, and 4 minutes respectively).

Simulation of the incubation period was basically the same as that of the first recall test except for the following: (a) recall was noisier ($\alpha_{inc} = 0.085$). (b) If an item was recalled, it was stored in a buffer memory [5], [8]. The incubation intervals were 0, 171, 857, and 1714 spins (for 0, 1, 5, and 10 minutes respectively).

The second free recall test was identical to the first, except that items in the buffer memory were output at the beginning of this period. This represented the fact that in [17], most words were recalled during the first minute of the second test (see *Section IV.B*).

[3] This can be accomplished by setting the number of nodes in the left layer to zero (i.e., $n = 0$ and $m = 50$) and the knowledge integration parameter to a large value (causing top-level rules to be ignored; e.g., $\lambda = 50$).

[4] The bottom level was pre-trained to encode the random representations for 15 epochs with: $\eta = 0.001$, $\xi = 0.9999$, $\delta = 0.49$, and $p = 10$. For an explanation of the parameters, see [14].

D. Simulation Results

Twelve thousand simulations were run (1,000 in each of the 12 conditions). A Test Duration × Incubation Interval ANOVA was performed on the reminiscence scores (see Fig. 5). As in the analysis of the human data, the incubation length had a significant effect on reminiscence $F(3, 11988) = 1661.34$, $p < .0001$. The mean reminiscence scores were 1.73, 2.03, 3.04, and 3.80 for the 0-, 1-, 5-, and 10-minute incubation intervals respectively, which is similar to the human data. *Post hoc* Tukey analyses showed that all incubation levels were statistically different ($p < .05$). The interaction between the factors also reached statistical significance $F(6, 11988) = 57.70$, $p < .0001$). This interaction indicates that the effect of test duration is different for short and long incubation intervals. Yet, the statistical significance of the difference is mainly due to the large number of degrees of freedom in the statistical analysis (i.e., number of simulations), because the biggest group difference within each incubation level is smaller than 1 word (i.e., the biggest difference between any pair of test lengths within the same incubation length is 0.739). Finally, the main effect of Test Duration did not reach statistical significance $F(2, 11988) = 0.78$, $p = 0.46$, as in the human data.

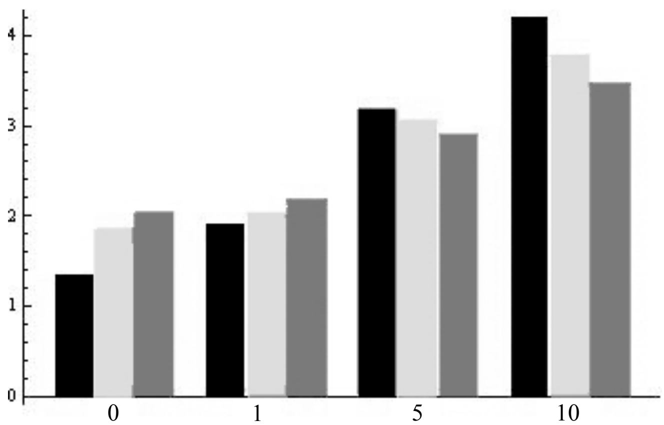

Fig. 5. Simulated reminiscence effect. The x-axis is the incubation length and the y-axis is the mean reminiscence score. The black bars represent 1-minute tests, the light gray bars represent 2-minute tests, and the dark gray bars represent 4-minute tests.

E. Discussion

The model was able to reproduce the effect of incubation on reminiscence reported in [17]. The main difference between the human data and the simulation data was the interaction between test duration and incubation interval. This difference could have been fixed by modeling more accurately the working memory capacity (e.g., by using an upper limit on memory capacity). However, the focus of this simulation was not to capture the minute details of the free recall data but to show that a general integrative model could account for the qualitative patterns in the human data. Overall, the model was successful in achieving this goal of capturing the data concerning the effect of incubation in a free recall experiment.

VI. Insight in a Discovery Task

A. Experiment

In their Experiment 3A, Bowers and his colleagues [20] tried to assess the continuous nature of the implicit processes leading to insight. To test this hypothesis, fifteen clue-words were presented sequentially to the participants, and their task was to find a word (the target) that was associated with all of them. First, a single clue-word was displayed, and the participants had 15 s to generate a potential solution (i.e., a word associated with the clue-word). Following this 15 s period, a second clue-word was added and the participants had 14 s to generate a word associated with the two clue-words. Each time a new clue-word was added, all the previous clue-words remained on the screen, and the time allowed to generate a potential solution (i.e., a word associated with all the available clue-words) was shortened by one second until it reached the sixth clue-word. From that point on, the participants had ten seconds to generate a potential solution after every additional clue. Note that clue-words thirteen to fifteen were on average thirteen times more frequently associated to the target than the other clue-words according to the Kent-Rosanoff word association test [21]. At any moment, if the participant thought that s/he might have generated the correct solution, s/he was to mark it as a ‘hunch’; if s/he was convinced that s/he had generated the correct associate, s/he was to mark it as a ‘solution’. 100 participants solved sixteen different problems. The dependant variables were the number of clue-words needed to find a hunch and the number of additional clue-words required to turn a hunch into a solution.

B. Empirical Results

The results showed that the mean number of clue-words required to generate a hunch was 10.12 ($SD = 4.55$) and that only 1.79 ($SD = 0.96$) additional clue-words were required to produce a solution (no statistical test was provided). These results were interpreted as showing that continuous processes are involved in this discovery task, and that the suddenness of insight simply reflects the reach of a “conscious” threshold. Also, the authors further speculated that hunches were implicitly generated, and explicitly tested. Hence, hunches should be the result of implicit processing while solutions should be found explicitly.

C. Simulation

In the top level, the left layer contained the clue-words while the right layer contained the targets (see Fig. 3). Each clue-word and each target word was represented using a different node ($n = 240$, $m = 16$). Also, the top-level associations between clue-words thirteen through fifteen and their associated target were strengthened to reproduce the experimental manipulation of word associations.[5] In the bottom level, each clue word in the sequence was associated to a different subset of units (using distributed representation); all the target words also shared a common subset of units ($r = 160$, $s = 150$). The increased associations

[5] In the pre-training, these clue-words were presented thirteen times each, whereas other clue-words were presented only once.

of clue-words thirteen through fifteen with the target words was (redundantly) encoded in the bottom level by adjusting relevant frequencies during the training of the bottom-level network (just like in the top level).[6]

To simulate this task, a stimulus (clue-word) first activated the top-level's left layer and the corresponding part of the bottom level. The activation was simultaneously used to seek explicit and implicit associations. (When multiple clue-words were presented simultaneously, the representations were simply superposed using summation, emulating the simultaneous presentation of several clue-words.) Because each iteration in the bottom level was assumed to takes 350 ms of psychological time (as in the previous simulation), 43 iterations were allowed for the first clue-word (15 s), and there was a decrement of 3 iterations for each subsequent clue-word until the fifth (a decrement of 1 s). From the sixth to the fifteenth, 28 iterations were allowed (10 s). Once the processing in the bottom level was completed, this information was sent back to the top level (through "explicitation") and integrated with the top-level activation (λ = 1). This integrated activation was further transformed into a Boltzmann distribution and the mode of the distribution was used to estimated the ICL. If the ICL was higher than a lower threshold (ψ_1 = 0.959), the output was marked as a hunch. If a hunch had already been identified and the ICL was higher than a higher threshold (ψ_2 = 0.961), the output was identified as a solution.[7]

D. Simulation Results

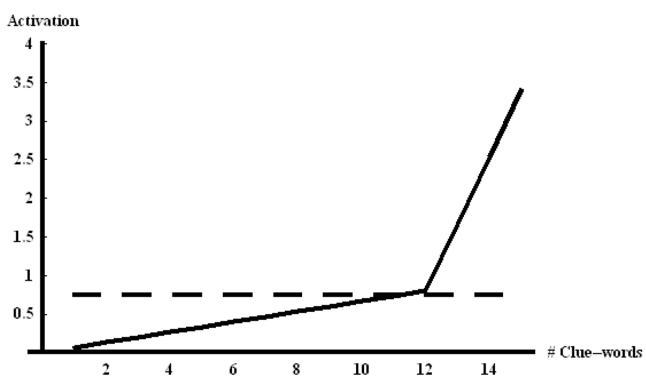

Fig. 6. Mean raw activation in the top and bottom levels after the presentation of each clue-word. The dashed line represents the bottom-level activation while the full line represents the top-level activation. The model response results from the integration of top- and bottom-level activation using a *Max* function.

One hundred simulations were run, each representing a different human participant (the same number as in the human experiment).[8] The mean number of clue-words needed to identify a hunch was 9.8 (SD = 4.4), and only 2.0 (SD = 3.0) additional clue-words were needed to correctly identify the solution, which is well inside the confidence interval of the human data (10.12±0.46 and 1.79±0.1, respectively). Fig. 6 shows the activation present in the top (the full line) and bottom (the dashed line) levels of the model. Because the model uses *Max* in integrating the results from the two levels after the presentation of each clue-word (2), hunches were on average generated by implicit processing (i.e., the dashed line is above the full line after 9.8 clue-words), while solutions were the results of explicit associative retrieval (i.e., the full line is above the dashed line after 11.8 clue-words). All these simulation results are in line with Bowers and his colleagues' data and predictions [20].

E. Discussion

CLARION successfully reproduced the empirical data. The activation in the model increased continuously, and "hunches" were generated mostly by the bottom level. 'Solutions' were mostly produced by the top level. This process is in line with (a) Bowers and his colleagues' speculations [20], (b) Evans' heuristic-analytic theory of reasoning [22], and (c) the EII theory [4]-[5]. Hence, CLARION captures data related to insight in the discovery task.

VII. Conclusion

In this paper, a connectionist model of creative problem solving has been proposed. The connectionist architecture is simple and yet powerful enough to capture psychological data related to incubation in free recall [17] and insight in the discovery task [20]. These simulations suggest that, in line with existing psychological theories and human data, the performance in different psychological tasks is affected by implicit processing even when attention is diverted, and that insight might be the result of a continuous process that *appears* to be sudden.

Other simulations supporting the presence of other stages of Wallas' analysis of creative problem solving [1] have been run using CLARION-based models (e.g., [4]-[5]). These additional simulations suggest that performance in insight problem solving (e.g., [23]-[24]) rely on the same basic processes as the tasks simulated in this paper (e.g., implicit processing and knowledge interaction [5]). Also, analysis of creative performances suggest that the Explicit-Implicit Interaction theory may provide an integrative framework relating all of these phenomena [5]. Future work should be devoted to the simulation of many more such tasks/problems, as well as the simulation of regular problem solving to further substantiate CLARION [8]-[9].

References

[1] G. Wallas, *The Art of Thought*. New York: Franklin Watts, 1926.
[2] 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.
[3] 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.
[4] S. Helie and R. Sun, "Knowledge integration in creative problem solving," in *Proceedings of the 30th Annual Meeting of the Cognitive*

[6] The bottom-level network was trained for 10 epochs with: η = 0.001, ξ = 0.9999, δ = 0.4, and p = 3. For an explanation of the parameters, see [14].

[7] A response was always output, and the ICL was only used to differentiate hunches from solutions. The solution that produced the ICL was always chosen (α took a very small value, e.g., α = 0.0001).

[8] This simulation is simpler than the Smith and Vela simulation. Hence, 100 simulations were sufficient to produce statistically stable results.

Science Society, B. C. Love, K. McRae, and V. M. Sloutsky, Eds. Austin, TX: Cognitive Science Society, 2008, pp. 1681-1686.

[5] S. Helie and R. Sun, “Incubation, insight, and creative problem solving: A unified theory and a connectionist model,” *Psychological Review*, in press..

[6] R. Sun, P. Slusarz, and C. Terry, “The interaction of the explicit and the implicit in skill learning: A dual-process approach,” *Psychological Review*, vol. 112, pp. 159-192, 2005.

[7] R. Sun, E. Merrill, and T. Peterson, “From implicit skills to explicit knowledge: A bottom-up model of skill learning,” *Cognitive Science*, vol. 25, pp. 203-244, 2001.

[8] R. Sun, *Duality of the Mind: A Bottom-Up Approach Toward Cognition*. Mahwah: Lawrence Erlbaum Associates, 2002.

[9] R. Sun and X. Zhang, “Accounting for a variety of reasoning data within a cognitive architecture,” *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 18, pp. 169-191, 2006.

[10] R. Sun and T. Peterson, “Autonomous learning of sequential tasks: Experiments and analyses,” *IEEE Transactions on Neural Networks*, vol. 9, pp. 1217-1234, 1998.

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

[12] C. Seger, “Implicit learning,” *Psychological Bulletin*, vol. 115, pp. 163-196, 1994.

[13] B. Kosko, “Bidirectional associative memories”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, pp. 49-60, 1988.

[14] 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.

[15] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman, 1982.

[16] J. J. Hopfield and D. W. Tank, “”Neural” computation of decisions in optimization problems”, *Biological Cybernetics*, vol. 52, pp. 141-152, 1985.

[17] S. M. Smith and E. Vela, “Incubated reminiscence effects,” *Memory & Cognition*, vol. 19, pp. 168-176, 1991

[18] B. H. Cohen, “Recall of categorized word lists,” *Journal of Experimental Psychology*, vol. 66, pp. 227-234, 1963.

[19] J. R. Anderson and R. Milson, “Human memory: An adaptive perspective,” *Psychological Review*, vol. 96, pp. 703-719, 1989.

[20] K. S. Bower, G. Regehr, C. Balthazard, and K. Parker, “Intuition in the context of discovery,” *Cognitive Psychology*, vol. 22, pp. 72-110, 1990.

[21] G. Kent and A. Rosanoff, “A study of association in insanity,” *American Journal of Insanity*, vol. 67, pp. 37-96, 1910.

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

[23] C. A. Kaplan and H. A. Simon, “In search of insight,” *Cognitive Psychology*, vol. 22, pp. 374-419, 1990.

[24] G. Knoblich, S. Ohlsson, H. Haider, and D. Rhenius, “Constraint relaxation and chunk decomposition in insight problem solving,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 25, pp. 1534-1555, 1999.