Psychological Review
2010, Vol. 117, No. 3, 994–1024

© 2010 American Psychological Association
0033-295X/10/$12.00 DOI: 10.1037/a0019532

Incubation, Insight, and Creative Problem Solving: A Unified Theory and a Connectionist Model

Sébastien Hélie
University of California, Santa Barbara

Ron Sun
Rensselaer Polytechnic Institute

This article proposes a unified framework for understanding creative problem solving, namely, the explicit–implicit interaction theory. This new theory of creative problem solving constitutes an attempt at providing a more unified explanation of relevant phenomena (in part by reinterpreting/integrating various fragmentary existing theories of incubation and insight). The explicit–implicit interaction theory relies mainly on 5 basic principles, namely, (a) the coexistence of and the difference between explicit and implicit knowledge, (b) the simultaneous involvement of implicit and explicit processes in most tasks, (c) the redundant representation of explicit and implicit knowledge, (d) the integration of the results of explicit and implicit processing, and (e) the iterative (and possibly bidirectional) processing. A computational implementation of the theory is developed based on the CLARION cognitive architecture and applied to the simulation of relevant human data. This work represents an initial step in the development of process-based theories of creativity encompassing incubation, insight, and various other related phenomena.

Keywords: creative problem solving, implicit processing, computational modeling, creativity, cognitive architecture

Many psychological theories of problem solving and reasoning have highlighted a role for implicit cognitive processes (e.g., Evans, 2006; Reber, 1989; Sun, 1994; Sun & Zhang 2004). For instance, implicit processes are often thought to generate hypotheses that are later explicitly tested (Evans, 2006). Also, similarity has been shown to affect reasoning through processes that are mostly implicit (Sun, 1994; Sun & Zhang, 2006). 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 (Dorfman, Shames, & Kihlstrom, 1996). However, when an ill-defined or complex problem has to be solved (e.g., when the initial state or the goal state can lead to many different interpretations or when the solution paths are highly complex), the solution is often found by sudden insight (Pols, 2002; Reber, 1989; Schooler & Melcher, 1995; Schooler, Ohlsson, & Brooks, 1993), and regular problem-solving theories are for the most part unable to account for this apparent absence of deliberative strategy (Bowden, Jung-Beeman, Fleck, & Kounios, 2005).

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 concentrate only on a subset of phenomena, such as focusing only on incubation (i.e., a period away from deliberative work on the problem; for a review, see S. M. Smith & Dodds, 1999) or insight (i.e., the sudden appearance of a solution; for a review, see Pols, 2002). The lack of detailed computational models has resulted in their limited impact on the field of problem solving (Duch, 2006).

In the present work, a general theory, the *explicit–implicit interaction* (EII) theory, is proposed. The EII theory integrates Wallas's (1926) high-level stage decomposition of creative problem solving with more detailed, process-based theories of incubation and insight (as detailed and implemented computationally later). Furthermore, the present article shows how EII can be used to provide a natural, intuitively appealing reinterpretation of several existing theories of incubation (e.g., unconscious work theory, conscious work theory, recovery from fatigue, forgetting of inappropriate mental sets, remote association, opportunistic assimilation; as reviewed in S. M. Smith & Dodds, 1999), several existing theories of insight (e.g., constraint theory, fixation theory, associationistic theory, evolutionary theory; as reviewed in Mayer, 1995; Ohlsson, 1992; Pols, 2002; Schilling, 2005; Schooler & Melcher, 1995; Simonton, 1995; S. M. Smith, 1995), and several existing theories of creativity (e.g., Geneplore, evolutionary theory of creativity; as reviewed in Campbell, 1960; Finke, Ward, & Smith, 1992). While the EII theory cannot account for all instances of creative problem solving, it is more integrative and more complete than the above-mentioned theories.

Another important characteristic of the EII theory is that the processes are specified with sufficient precision to allow their implementations into a quantitative, process-based, computational

Sébastien Hélie, Department of Psychology, University of California, Santa Barbara; Ron Sun, Cognitive Science Department, Rensselaer Polytechnic Institute.

This research was supported by a postdoctoral research fellowship from Le Fonds Québecois de la Recherche sur la Nature et les Technologies granted to Sébastien Hélie and by Army Research Institute Contract W74V8H-05-K-0002 and Office of Naval Research Grant N00014-08-1-0068 to Ron Sun. The code used for the simulations can be found at http://www.psych.ucsb.edu/~helie/papers/pr-eii/ and http://www.cogsci.rpi.edu/~rsun/pr-eii.

Correspondence concerning this article should be addressed to Sébastien Hélie, Department of Psychology, University of California, Santa Barbara, Santa Barbara, CA 93106-9660, or to Ron Sun, Cognitive Science Department, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180. E-mail: helie@psych.ucsb.edu or rsun@rpi.edu

model using the CLARION cognitive architecture (Sun, 2002; Sun, Merrill, & Peterson, 2001; Sun, Slusarz, & Terry, 2005). In the present work, CLARION is used to simulate, capture, and explain human data related to incubation and insight in tasks such as lexical decision (Yaniv & Meyer, 1987), free recall (S. M. Smith & Vela, 1991), and problem solving (Durso, Rea, & Dayton, 1994; Schooler et al., 1993). CLARION is also applied to conceptually capturing and explaining several computational models of creativity from artificial intelligence (see, e.g., Boden, 2004; Hofstadter & Mitchell, 1994; Langley & Jones, 1988; Rowe & Partridge, 1993; Schank & Cleary, 1995).

The remainder of this article is organized as follow. First, Wallas's (1926) ubiquitous stage decomposition is introduced. Second, the existence of the two stages on which this article focuses, namely, incubation and insight, is justified through reviewing relevant experimental psychology literature. This discussion is followed by a review of the existing theories of incubation and insight, which serves to motivate the EII theory of creative problem solving. The new theory is then explained and the previously reviewed theories are reinterpreted in the new framework. This is followed by the implementation of EII using the CLARION cognitive architecture. Four tasks previously used to experimentally justify incubation and insight are then simulated with the proposed computational model. The article concludes by discussing the implications of the EII theory and the CLARION model for psychological research on creativity as well as for computational models of creativity in artificial intelligence.

Creative Problem Solving: Four Stages

The role of creativity in problem solving has been acknowledged since Wallas's (1926) seminal work. According to Wallas, humans go through four different stages when trying to solve a problem: preparation, incubation, illumination (i.e., 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. However, if the problem is ill defined and/or complex, 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 empirically shown to increase the probability of eventually finding the correct solution (e.g., Dodds, Ward, & Smith, in press; S. M. Smith & Dodds, 1999). The following stage, illumination, is the spontaneous manifestation of the problem and its solution in conscious thought.[1] The fourth stage, verification, is used to ascertain the correctness of the insight solution. 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 general and not easily testable, it has been used to guide much of Gestalt psychologists' early research program on problem solving (e.g., Duncker, 1945; Kohler, 1925; Maier, 1931). 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 (Pols, 2002). 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 (Duncker, 1945). This Gestalt theory of problem solving provides a sketchy high-level description of creative problem solving, but no detailed psychological mechanism (especially process-based or computational mechanism) has been proposed.

More recent research has turned to 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 (Wallas, 1926), not much effort has been devoted to these two stages (relevant results can be borrowed from the existing literature; see, e.g., Johnson-Laird, 1999; Simon, 1966; Sun, 1994; Zadeh, 1988). In contrast, incubation and insight have received much attention; some of the most relevant results on incubation and insight are reviewed below.

Incubation

A recent review of experimental research on incubation showed that most experiments have found significant effects of incubation (Dodds et al., in press). Those experiments investigated the effects of incubation length, preparatory activity, clue, distracting activity, expertise, and gender on participants' performance. The review suggested 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 (Dodds, Smith, & Ward, 2002). (The other three factors, distracting activity, expertise, and gender, have not been studied enough to yield a stable pattern of results, but see Hélie, Sun, & Xiong, 2008, for a discussion of the effect of distracting activity.)

In addition to being correlated to such factors (Dodds et al., in press), incubation has 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; S. M. Smith & Vela, 1991) and priming (Yaniv & Meyer, 1987). For example, S. M. Smith and Vela (1991) showed that reminiscence in a free-recall task was increased by longer intertest interval (i.e., the length of incubation). Furthermore, in relation to priming, Yaniv and Meyer (1987) showed that participants who rated their feeling of knowing (FOK) as high in a rare-word association task (which is suggestive of more efficient incubation) were primed for a related solution in a

[1] This moment is often referred to as the "Aha!" experience or the "Eureka!" (of Archimedes). In modern literature, illumination has been called insight (Bowden et al., 2005; Pols, 2002; Schooler & Melcher, 1995).

subsequent lexical decision task. In contrast, other participants who rated their FOK as either medium or low (which is suggestive of less efficient incubation) were not primed in the subsequent lexical decision task. Overall, the review by Dodds et al. (in press) and the results presented above (S. M. Smith & Vela, 1991; Yaniv & Meyer, 1987) support the existence of incubation in problem solving (and in other psychological tasks).

Insight (Illumination)

In a recent review of the different definitions used in psychology to characterize insight, Pols (2002) 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 (e.g., Bowden et al., 2005; Dorfman et al., 1996; Isaak & Just, 1996; Mayer, 1995; Pols, 2002). Such problems are diverse and characterized by the absence of direct, incremental algorithms allowing their solutions. In many cases, they are selected because they have been shown to produce insight solutions in previous studies (Bowden et al., 2005). Empirically, insight is identified by a strong discontinuity in the FOK, the feeling of warmth, or the progress made in a verbal report (Pols, 2002). Some research has even shown a sudden increase of heart rate just before insight is reached (whereas regular problem solving is accompanied by a steady increase in heart rate; see Jausovec & Bakracevic, 1995). Overall, the existing data (e.g., Duncker, 1945; Durso et al., 1994; Jausovec & Bakracevic, 1995; Maier, 1931; Ohlsson, 1992) and the observed phenomena (Duncker, 1945; Durso et al., 1994; Kohler, 1925; Maier, 1931; Schooler et al., 1993; Schooler & Melcher, 1995) support the existence of insight in creative problem solving (see, e.g., Mayer, 1995; Ohlsson, in press; Pols, 2002, for further reviews).

Existing Theories

Many process theories have been proposed to explain incubation and insight (as reviewed below). However, it should be noted that each of these theories can be used to explain only certain limited aspects of the data in the literature (S. M. Smith & Dodds, 1999). Furthermore, most existing theories do not attempt to explain insight and incubation simultaneously. Below, some of the better known theories of incubation and insight are reviewed to provide the necessary background for the EII theory.[2]

Review of existing theories of incubation.

Unconscious work. The most natural process theory of incubation, stemming directly from Wallas's (1926) intuition, is known as the *unconscious work theory* (Dorfman et al., 1996; S. M. Smith & Dodds, 1999). According to this theory, the problem solver continues to work unconsciously on the problem after abandoning conscious work. A creative solution to a problem is developed unconsciously and reaches consciousness as a whole. The unconscious work theory has the advantage of being consistent with most anecdotes in the history of science. However, the presence of unconscious work is difficult to assess experimentally (S. M. Smith & Dodds, 1999).

Conscious work. The *conscious work theory* was proposed in light of the difficulties with the experimental assessment of unconscious processes (S. M. Smith & Dodds, 1999). According to the conscious work theory, a creative solution is found by working intermittently on the problem while attending to mundane activities (e.g., taking a shower, driving, etc.). Because attention switching from the mundane activity to the incubated problem is very fast, the short episodes of work on the incubated problem are forgotten, and only the final step is remembered.

Recovery from fatigue. The preparation phase in real-world situations can be very long and tiring. The problem solver might be cognitively drained and therefore unable to solve the problem (S. M. Smith & Dodds, 1999). According to this theory, the stage of incubation is a cognitive respite period, which allows rejuvenation of the problem-solving skills.

Forgetting of inappropriate mental sets. False assumptions are sometimes made during the preparation phase. These false assumptions erroneously constrain the possible solution space and prevent the solver from producing certain solutions (S. M. Smith & Dodds, 1999); the false assumptions must be forgotten to allow the problem solver to be creative (and access productive solutions). The incubation period serves this purpose.

Remote association. Solutions to already-solved problems are often stored in long-term memory. When a new problem is encountered, the (previously stored) solutions to similar problems are automatically retrieved. However, these solutions might be inappropriate and block the correct solution from being discovered. Less likely solutions are discovered only when the most likely solutions have all been investigated. The incubation phase is thus used to eliminate stereotypical solutions.

Opportunistic assimilation. Unsolved problems are often encoded in long-term memory. As long as the problem remains unsolved, the resulting memory structure is primed, and environmental clues that may be useful in solving the problem can easily activate the appropriate structure (S. M. Smith & Dodds, 1999). Incubation is the period in which environmental clues are assimilated. Such incubation makes the problem solver sensitive to details and hints that would have gone unnoticed without the priming from the unsolved problem in long-term memory (Langley & Jones, 1988).

Review of existing theories of insight.

Constraint theory. The first theory assumes that insight problems involve the satisfaction of a large number of constraints (Mayer, 1995). Limits related to cognitive resources make it difficult to simultaneously satisfy a large set of constraints (Simon, 1972), which explains the intense experience associated with insight. This *constraint theory* of insight has been used mainly to describe the problem-solving process via schema completion (e.g., Schank & Cleary, 1995). In such a case, the problem solver

[2] While there have been other theories proposed over the years, they have had more limited impact compared with the theories reviewed here (e.g., they were not included in the *Encyclopedia of Creativity*; see S. M. Smith & Dodds, 1999). Hence, we have chosen to limit the discussion to these high-impact theories.

mentally constructs a structure that includes the initial condition (problem) and the goal state (solution) and fills in the gap between the initial condition and the goal state that may exist.

Fixation theory. Although the constraint theory has been useful in explaining historical anecdotes and verbal reports of problem solving (Mayer, 1995), it is limited to cases where only the path between the initial state and the solution is missing. Unfortunately, this is not always the case. For example, the solution state is often unknown in advance. As such, the *fixation theory* (Mayer, 1995; Ohlsson, 1992, in press; Pols, 2002; Schilling, 2005; Schooler & Melcher, 1995; S. M. Smith, 1995) also assumes that insight problems involve the satisfaction of constraints, but it does not assume that all the constraints are stated in the initial problem: Problem solvers sometimes wrongly assume constraints that are not part of the problem, which limits the search process to a portion of the solution space (Isaak & Just, 1996). According to this theory, insight is experienced when these unwarranted constraints are relaxed and a new portion of the solution space becomes available for exploration. The rejection of these constraints is usually achieved by restructuring the problem (S. M. Smith, 1995).

Associationistic theory. In the preceding theories, insight has been interpreted as successfully satisfying a set of constraints (so as to break an impasse). However, not all theories assume that an impasse has to be reached or that constraints must be satisfied. The *associationistic theory* assumes that knowledge is encoded using a knowledge graph (Pols, 2002; Schilling, 2005). Accordingly, problems are solved by retrieving the correct association (path) using parallel search processes. Insight is nothing special (Mayer, 1995; Schooler & Melcher, 1995): The only difference between insight and noninsight solutions is the strength of the associations. Insight is experienced when an unlikely association solving the problem is retrieved.

Evolutionary theory. The *evolutionary theory* of insight (Campbell, 1960; Pols, 2002; Schilling, 2005; Simonton, 1995) is based on the three principles of Darwin's theory of evolution: (a) blind variation/generation of solutions, (b) evaluation/selection of a solution, and (c) retention of the selected solution (Simonton, 1995; see also Hadamard, 1954). According to the evolutionary theory of insight, knowledge is represented by nodes in a graph, and associations (links) are formed using an evolutionary selection principle. Solution generation (i.e., the formation of associations) and selection are performed unconsciously, and only the selected solution (association) reaches consciousness. If the solution adequately solves the problem, insight is experienced.

EII: An Integrative Theory of Creative Problem Solving

The EII theory, in part, attempts to integrate and thus unify (to some extent) existing theories of creative problem solving in two senses. First, most theories of creative problem solving focus on either a high-level stage decomposition (e.g., Wallas, 1926) or on a process explanation of only one of the stages (see the previous subsections). None of the above-mentioned theories provides a stage decomposition along with a process explanation of more than one stage (Lubart, 2001). Second, the process theories of incubation (e.g., S. M. Smith & Dodds, 1999) and insight (e.g., Mayer, 1995; Ohlsson, 1992, in press; Pols, 2002) are usually incomplete and often mutually incompatible. EII attempts to integrate the existing theories to make them more complete so as to provide a detailed description of the processes involved in key stages of creative problem solving. EII starts from Wallas's (1926) stage decomposition of creative problem solving and provides a detailed process-based explanation sufficient for a coherent computational implementation. (This last point is important because most of the aforementioned process theories are not detailed enough to be implemented as computational models.)

The basic principles underlying the EII theory are summarized in Table 1. As can be seen, EII is not just an integration/implementation of previously existing vague theories; it is a new theory, which focuses on the importance of implicit processing and knowledge integration in problem solving (see Sun et al., 2005). In the following subsection, the principles summarized in Table 1 are presented in more detail. This description is followed by theoretical and empirical justifications of the principles. This section ends with a discussion of EII's implications for psychological research on creative problem solving (i.e., explanation and integration of existing theories).

Basic Principles of the EII Theory

Principle 1: The coexistence of and the difference between explicit and implicit knowledge. The EII theory assumes the existence of two different types of knowledge, namely, explicit and implicit (Dienes & Berry, 1997; Dienes & Perner, 1999), residing in two separate modules (Sun, 2002). Explicit knowledge is easier to access and verbalize, and said to be often symbolic, crisper, and more flexible (Sun et al., 2001, 2005). However, using explicit knowledge requires more extensive attentional resources (Curran & Keele, 1993; Sun et al., 2005). In contrast, implicit knowledge is relatively inaccessible, harder to verbalize, often subsymbolic, and often more specific, more vague, and noisier (Sun, 1994, 2002). However, using implicit knowledge does not tap much attentional resource. As such, explicit knowledge and implicit knowledge are processed differently. According to the EII theory, explicit processes perform some form of rule-based reasoning (in a very generalized sense; E. E. Smith, Langston, & Nisbett, 1992; Sun, 1994) and represent relatively crisp and exact processing (often involving hard constraints; Sun et al., 2001),

Table 1
Principles of the Explicit–Implicit Interaction Theory

Basic principles
1. The coexistence of and the difference between explicit and implicit knowledge.
2. The simultaneous involvement of implicit and explicit processes in most tasks.
3. The redundant representation of explicit and implicit knowledge.
4. The integration of the results of explicit and implicit processing.
5. The iterative (and possibly bidirectional) processing.

Auxiliary principles
1. The existence of a (rudimentary) metacognitive monitoring process.
2. The existence of subjective thresholds.
3. The existence of a negative relation between confidence and response time.

while implicit processing is associative and often represents soft-constraint satisfaction (Evans, 2008; Sloman, 1996; Sun, 1994).

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 (E. R. Smith & DeCoster, 2000; Sun, 2002). This can be justified by the different representations and processing used to describe the two types of knowledge. As such, each type of process can end up with similar or conflictual conclusions that contribute to the overall output (Evans, 2007; see also Principle 4 below).

Principle 3: The redundant representation of explicit and implicit knowledge. According to the EII theory, explicit knowledge and implicit knowledge are often redundant, that is, they frequently amount to redescriptions of one another in different representational forms. For example, knowledge that is initially implicit is often later recoded to form explicit knowledge (through bottom-up learning; Sun et al., 2001, 2005). Likewise, knowledge that is initially learned explicitly (e.g., through verbal instructions) is often later assimilated and recoded into an implicit form, usually after extensive practice (top-down assimilation: Sun & Zhang, 2004).[3] There may also be other ways redundancy is created, for example, through simultaneous learning of implicit and explicit knowledge. Redundancy often leads to interaction (as described in Principle 4).

Principle 4: The integration of the results of explicit and implicit processing. Although explicit knowledge and implicit knowledge are often redescriptions of one another, they involve different forms of representation and processing, which may produce similar or different conclusions (Sun & Peterson, 1998); the integration of these conclusions may be necessary, which may lead to synergy, that is, overall better performance.

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 processes 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 may often use the integrated outcome as a new input. Reversing the direction of reasoning may sometimes carry out this process (e.g., abductive reasoning; Johnson & Krems, 2001; Pearl, 2000). Alternating between forward and backward processing has been argued to happen also in everyday human reasoning (Rips, 1994).

Auxiliary principles. In addition to the five principles presented so far, three auxiliary principles should be mentioned. These principles are less important because they are needed to account for the data, but alternative principles may be equally viable. Therefore they are not central to the fundamental theoretical framework of the EII theory. First, Principle 5 implies that a definitive result needs to be achieved to terminate the iterative process. This stopping criterion assumes a primitive form of metacognitive monitoring that can more or less accurately measure the probability of finding a solution (Bowers, Regehr, Balthazard, & Parker, 1990). In EII, this metacognitive measure is termed the *internal confidence level* (ICL). Second, there must be a threshold that defines what is meant by *definitive result.* This threshold can vary as a function of task demands, and there might even be several thresholds for different levels of confidence (Bowers et al., 1990; Ohlsson, in press). Lastly, a negative relationship between the ICL and the response time is assumed (as in, e.g., J. R. Anderson, 1991; Costermans, Lories, & Ansay, 1992).

Justification of the Principles

Principle 1: The coexistence of and the difference between explicit and implicit knowledge. There have been disagreements concerning what experimentally constitutes conscious accessibility (Dienes & Berry, 1997). It is also difficult to distinguish between explicit knowledge that is used when a task is being performed and explicit knowledge that is retroactively attributed to task performance (i.e., when verbal reports are given). Despite such difficulties, it is generally agreed that at least some part of performance is not consciously accessible under normal circumstances. Reber (1989) pointed out that "although it is misleading to argue that implicitly acquired knowledge is completely unconscious, it is not misleading to argue that implicitly acquired epistemic contents of mind are always richer and more sophisticated than what can be explicated" (p. 229). Voluminous experimental data testifying to this distinction can be found in Berry and Broadbent (1988); Cleeremans, Destrebecqz, and Boyer (1998); Dienes and Berry (1997); Karmiloff-Smith (1992); Mathews et al. (1989); Reber (1989); Seger (1994); Stanley, Mathews, Buss, and Kotler-Cope (1989); and Sun et al. (2001, 2005).

In general, explicit processing can be qualified as rule based in some way, whereas implicit processing is mostly associative (as argued by, e.g., Sloman, 1996; Sun, 1994). Explicit processing can involve the manipulation of symbols through the application of various explicit reasoning processes, for example, logical reasoning (Rips, 1994) and explicit hypothesis testing (Evans, 2002, 2006). The abstract nature of symbol manipulation allows for the application of knowledge in different but categorically similar situations (i.e., systematicity; Fodor & Pylyshyn, 1988). In contrast, implicit processing involves mostly instantiated knowledge that is holistically associated (Sun, 1994; Sun et al., 2001, 2005). Hence, implicit processing is often more situation specific and provides approximate matches in new situations (Sun, 1994), which limits the validity of its results. (Empirical evidence in support of these points can be found in the reviews cited above and thus is not detailed here.)

The above differences between explicit and implicit processing have important implications for the types of constraints that can be handled with each type of processing. Because explicit knowledge is often thought of as rule based, explicit processing can be viewed as an algorithm that satisfies hard constraints (as argued by, e.g., Sloman, 1996; Sun, 1994). For example, the proof of a theorem is done by using the rules and the axioms of mathematics (i.e., hard constraints), which must be completely satisfied. Such a task necessarily requires the use of explicit processes (along with implicit processes possibly). In contrast, inferring that robins and blue jays are similar can be done by using soft constraints (e.g., a similarity metric, as argued by, e.g., Sun, 1994). This is also in line with Dijksterhuis, Bos, Nordgren, and van Baaren (2006), who showed that an important reason for why unconscious thinkers

[3] This phenomenon is also referred to as automaticity (Hélie & Ashby, 2009; Hélie, Waldschmidt, & Ashby, 2010; Logan, 1988, 1992) or proceduralization (J. R. Anderson & Lebiere, 1998; Sun et al., 2001).

often made superior decisions had to do with the way decision makers weighed the relative importance of various soft constraints.

The last distinction between explicit and implicit processing, attentional difference, can be demonstrated, in particular, through a serial reaction time task involving a dual-task phase that cancels the beneficial effect of explicit knowledge while leaving implicit processing (mostly) untouched (Curran & Keele, 1993). Similar effects have been found in artificial grammar learning tasks, dynamic control tasks (for reviews, see Cleeremans et al., 1998; Sun et al., 2005), and perceptual categorization (Waldron & Ashby, 2001).

Principle 2: The simultaneous involvement of implicit and explicit processes in most tasks. One of the ways to show the simultaneous nature of explicit and implicit processing is to create a conflict situation (Evans, 2007). Processing hard (explicit) and soft (implicit) constraints simultaneously can result in different inferences, which can lead to a conflict (Evans, 2007; E. R. Smith & DeCoster, 2000). For instance, the similarity between the stimuli (implicit processing) has been shown to have a strong effect on rule-based categorization (explicit processing), which can lead to a conflict that suggests simultaneous implicit and explicit processing (Allen & Brooks, 1991; but see Lacroix, Giguère, & Larochelle, 2005). Similar results have been found in a syllogistic reasoning task (Evans, 2007).

Yet another line of evidence comes from research on skill acquisition. For instance, it was argued in Sun et al. (2005) that it is not necessary to select upfront explicit or implicit processes to tackle a particular problem. Most tasks are processed simultaneously explicitly and implicitly, and the task demands (e.g., complexity, structure, prior instructions, etc.) can make explicit or implicit processes more efficient. Hence, although performance might seem to be the result of (mostly) explicit or implicit processing by an external observer, both types of process are likely involved, and the observable (i.e., measurable) behavior results mostly from the more efficient process in a particular task setting. (For a detailed argument and review, see Sun et al., 2005.)

Principle 3: The redundant representation of explicit and implicit knowledge. Redundancy is very important in providing fault tolerance in cognitive systems (Russell & Norvig, 1995; von Newmann, 1956). For example, the brain is composed of millions of neurons that are known to be individually very noisy (Ma, Beck, Latham, & Pouget, 2006). Yet psychological processes are often robust (e.g., see Sun, 1994). Robustness can be achieved through redundancy.

A natural way of creating redundancy is to redescribe one kind of knowledge into the other. Early memory experiments in the context of the depth-of-processing hypothesis showed this strategy to be efficient for later recall (Craik & Tulving, 1975). Redescription of implicit knowledge into explicit knowledge is termed bottom-up learning (see, e.g., Sun et al., 2001), while the redescription of explicit knowledge into implicit knowledge is termed top-down assimilation (Sun & Zhang, 2004). Many psychological data have suggested the presence of bottom-up learning and top-down assimilation (as argued in Sun et al., 2005). Some of them are reviewed below.

Sun et al. (2001) proposed the idea of bottom-up learning and gathered much empirical evidence for it. In many experiments, the participants' ability to verbalize was independent of their performance (Berry & Broadbent, 1988). Furthermore, performance typically improved earlier than explicit knowledge that could be verbalized by participants (Stanley et al., 1989). For instance, in dynamic control tasks, although the performance of participants quickly rose to a high level, their verbal knowledge improved far slower: Participants could not provide usable verbal knowledge until near the end of their training (e.g., as shown by Stanley et al., 1989; Sun et al., 2005). This phenomenon has also been demonstrated by Reber and Lewis (1977) in artificial grammar learning. A more recent study of this phenomenon (Sun et al., 2001) used a more complex minefield navigation task. In all of these tasks, it appeared easier to acquire implicit skills than explicit knowledge (hence the delay in the development of explicit knowledge). In addition, the delay indicates that implicit learning may trigger explicit learning, and the process may be described as delayed explication of implicit knowledge (Karmiloff-Smith, 1992). Explicit knowledge is in a way extracted from implicit skills. Together, these data suggest the existence of bottom-up learning.

Top-down assimilation may be demonstrated through data on automaticity (Hélie et al., 2010; Logan, 1988, 1992). For instance, explicit processing (letter counting) is abandoned in favor of an implicit strategy (memory retrieval) in alphabetic arithmetic tasks (Logan, 1988). This form of automaticity is possible only once the explicit knowledge has been assimilated into implicit knowledge. Similar results can also be found in a dot-counting task (Logan, 1992), several lexical decision tasks (Logan, 1988), categorization (Hélie et al., 2010), and proceduralization experiments (e.g., J. R. Anderson & Lebiere, 1998; Sun et al., 2001).

Redundancy may also be created when simultaneous implicit learning and explicit learning are taking place. There is evidence that implicit and explicit knowledge may develop independently under some circumstances. Willingham, Nissen, and Bullemer (1989) reported data that were consistent with the parallel development of implicit and explicit knowledge. By using two different measures for assessing the two types of knowledge, they compared the time course of implicit and explicit learning. It was shown that implicit knowledge might be acquired in the absence of explicit knowledge and vice versa. The data ruled out the possibility that one type of knowledge was always preceded by the other. Rabinowitz and Goldberg (1995) similarly demonstrated parallel development of procedural and declarative knowledge in some conditions of an alphabetic arithmetic task.

Principle 4: The integration of the results of explicit and implicit processing. Simultaneous processing of explicit and implicit knowledge often leads to an output that is a combination of the results of explicit and implicit processing (Sun et al., 2001, 2005; Sun & Peterson, 1998). Such knowledge integration sometimes produces synergy (Sun & Peterson, 1998), which can lead to speeding up learning, improving performance, and facilitating transfer (Sun et al., 2005).

Knowledge integration is supported by recent neurological findings in insight problem solving (Bowden et al., 2005). According to Bowden and his colleagues (2005), problem solving is performed differently in the left and right brain hemispheres. The former is more closely related to language processing and strongly activates a limited set of concepts, while the latter is more related to imagery and provides diffused activation to a wider range of concepts. Hence, each hemisphere holds a different problem representation. Shortly before insight problems are solved, a neuronal burst sending a signal from the right hemisphere to the left hemisphere can be observed, thus yielding an integrated problem representation leading to the solution (Bowden et al., 2005). Similar phenomena have been found (behaviorally) in many other psychological tasks, such as the serial reaction time task (Curran & Keele, 1993), the finite-state grammar task (Mathews et al., 1989), the dynamic control task (Stanley et al., 1989), and the minefield navigation task (Sun et al., 2001). Many other similar indications exist to support the integration of explicit and implicit knowledge (see Sun et al., 2005, for a detailed review).

Principle 5: The iterative (and possibly bidirectional) processing. The often iterative/bidirectional processing assumed by the EII theory corresponds well with data, especially human reasoning data. For instance, forward and backward schemas were used to describe human performance in reasoning tasks (Rips, 1994). One important form of backward processing in human reasoning is abductive reasoning (Pearl, 2000). In abductive reasoning, one tries to infer the possible cause(s) following a set of observations. For instance, the floor can be observed to be wet, and one can infer that it might have rained earlier (i.e., the current hypothesis). Testing other effects that should be observed if this possible cause is correct can be used to refine the current hypothesis (e.g., if it rained earlier, other objects should also be wet). Johnson and Krems (2001) showed that human participants use this strategy and often use the current hypothesis to interpret new data. According to these authors, the main purpose of abductive reasoning is to control the growth of the complexity of the hypothesis space, which can quickly become unmanageable (e.g., by restraining the size of the search space by refining the current hypothesis). For this reason, abductive reasoning can be very useful in creative problem solving because insight problems are often overly complex and ill defined (Bowden et al., 2005). Hence, this form of backward processing, which is also consistent with the Bayesian interpretation of the rational analysis of cognition (J. R. Anderson & Lebiere, 1998), is used to initiate subsequent rounds of processing in EII.

In addition to the iterative nature of hypothesis refinement in abductive reasoning, iterative processing in EII is also supported by the heuristic-analytic theory (Evans, 2006). According to Evans (2006), participants repeatedly use heuristics (implicit processing) to generate models that are verified by analytic (explicit) processes in logical inference tasks. This generate-and-test algorithm is repeated until a satisfactory solution is found, which is consistent with EII's assumption about iterative processing. The creative problem-solving literature also provides countless examples of participants iteratively formulating and revising their hypotheses (through forward and backward processes; e.g., Bowden et al., 2005; Durso et al., 1994; Schooler et al., 1993).

Auxiliary principles. First, iterative processing ends when the ICL reaches a certain level according to EII. Empirically, the ICL can correspond to the FOK (or feeling of warmth) that has been measured in the metacognition literature (Bowers et al., 1990; Yaniv & Meyer, 1987) and is assumed by many theories of problem solving (e.g., Dorfman et al., 1996; Sun, Zhang, & Mathews, 2006) and many theories of memory search (e.g., Metcalfe, 1986; Metcalfe & Wiebe, 1987). These existing data and theories support the use of an ICL in EII. Second, Bowers et al. (1990) directly tested the presence of a hunch threshold and a solution threshold on FOKs. In addition, Dienes and Berry (1997) argued in favor of the presence of absolute and subjective thresholds to differentiate knowledge that is included in verbal reports from knowledge that is not included in verbal reports. This is consistent with the assumption of multiple thresholds on ICLs in EII (see also Ohlsson, in press). Also, when a response is output by a problem solver, the ICL may be used to estimate the confidence level that one reports (Costermans et al., 1992; Miner & Reder, 1994). Third, Costermans et al. (1992) showed that confidence levels are negatively related to response times, and their results suggest that this relation might be linear. Hence, EII assumes that response times are a negative (and possibly linear) function of the ICL when a response is output (see also J. R. Anderson, 1991). In addition to being consistent with empirical data, the linear relation is the simplest possible relation between the ICLs and the response times.

Accounting for Creative Problem Solving Using EII

The preceding assumptions allow for a conceptual model that captures the four stages of Wallas's (1926) analysis of creative problem solving (see Figure 1). First, Wallas described the preparation stage as involving "the whole traditional art of logic" (Wallas, 1926, p. 84). Hence, the preparation stage is mainly

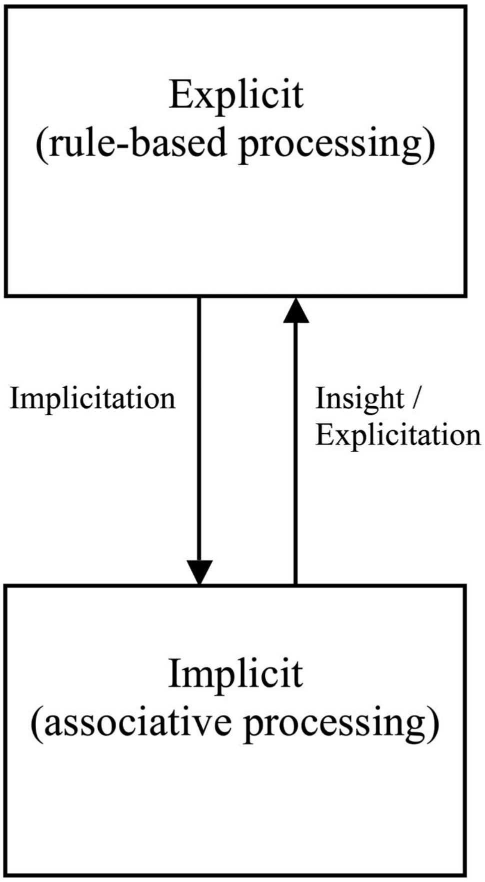


Figure 1. Information flow in the explicit–implicit interaction theory. The top level (explicit processing) is more heavily relied upon in the preparation and verification stages. The bottom level (implicit processing) is more heavily relied upon during incubation. Insight corresponds to the transfer of a solution from the bottom level to the top level.

captured by explicit processing in the EII theory. This is justified because explicit knowledge is usually rule based in some sense (Principle 1 of EII), which includes logic-based reasoning as a special case. Also, the preparation stage has to be explicit in EII because people are responding to (explicit) verbal instructions, forming representations of the problem, and setting goals (although implicit processes may also be involved to some lesser extent).[4]

In contrast, incubation relies more heavily on implicit processes in EII. According to Wallas (1926), incubation is the stage during which "we do not voluntarily or consciously think on a particular problem" (p. 86). This is consistent with our hypothesis regarding the difference of conscious accessibility between explicit and implicit knowledge (Principle 1 of EII). Moreover, incubation can persist implicitly for an extended period of time in Wallas's theory (see also Hadamard, 1954). This characteristic of incubation corresponds well with the above-mentioned hypothesis concerning the relative lack of attentional resource requirement in implicit processing. However, sometimes, explicit processing can occur in relation to the problem to be solved during incubation (see the conscious work theory of incubation, as reviewed in S. M. Smith & Dodds, 1999), and the EII theory is also consistent with this possibility (Principle 2 of EII).

The third stage, insight, is "the appearance of the 'happy idea' together with the psychological events which immediately preceded and accompanied that appearance" (Wallas, 1926, p. 80). In EII, insight is obtained by the crossing of a confidence threshold by the ICL, which makes the output available for verbal report (an auxiliary principle of EII). It is worth noting that the intensity of insight is often continuous (Bowden et al, 2005; Bowers et al., 1990; Hadamard, 1954; see also the associationistic theory of insight as reviewed in, e.g., Pols, 2002). Correspondingly, in the EII theory, the ICL is continuous. In particular, when the ICL of an output barely crosses the confidence 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 confidence threshold, a very intense experience can result (which, for example, can lead to running naked in the street, as Archimedes did). 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 (Principle 4 of EII).

Finally, the verification phase "closely resembles the first stage of preparation" (Wallas, 1926, 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 verification suggests that the insight solution might be incorrect, the whole process may be repeated by going back to the preparation stage (Finke et al., 1992; Hadamard, 1954; Wallas, 1926). 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).

A few example predictions (phenomena accounted for) by the EII theory are summarized in Table 2. Most of these predictions are explained in detail in the Simulations section of this article (while others may be found in, e.g., Hélie et al., 2008).

Reinterpretation of Existing Theories of Incubation Using EII

Unconscious work. Pure incubation is often essentially implicit (or unconscious; see Figure 1). In EII, information is simultaneously spread in explicit and implicit memories, with the latter being mostly responsible for incubation. As a result, the proposed theory provides a natural embodiment of the unconscious work theory of incubation (Dorfman et al., 1996; S. M. Smith & Dodds, 1999).

Conscious work. The assumed implicitness of incubation does not prevent EII from providing an interpretation of the conscious work theory of incubation (S. M. Smith & Dodds, 1999). As stated by Principle 2, most tasks are processed both explicitly and implicitly. Hence, incubation is in some sense partly a result of explicit processing. In addition, the result of the explicit processing can be below threshold (an auxiliary principle of EII), which would result in the problem solver being somewhat blind to this form of explicit processing (Karmiloff-Smith, 1992; S. M. Smith & Dodds, 1999).

Recovery from fatigue. EII assumes that explicit processes require more extensive attentional resources, whereas implicit processes are often effortless (Principle 1 of EII). Hence, according to EII, the preparation stage, which extensively involves explicit processing, can cause mental fatigue. In contrast, implicit processes do not wear out the problem solver as easily; incubation may proceed even when one is relaxing. Following a respite period (while incubating), the problem solver can again explicitly search for a solution to the problem. However, the state of its knowledge has been altered by the implicit processing that took place during incubation (Principle 5 of EII). The integration of implicit and explicit processing from this point on may lead to insight. This process reinterprets the recovery-from-fatigue hypothesis (S. M. Smith & Dodds, 1999).

Forgetting of inappropriate mental sets. According to the forgetting-of-inappropriate-mental-sets hypothesis, false assumptions are made during the preparation period; these assumptions must be forgotten to solve the problem. In the EII theory, the assumptions made during the preparation stage are used to contextualize subsequent processing (e.g., in the incubation phase). This contextualization can be altered by involvement in other problems, which can provide interference that speeds up the forgetting of past assumptions (Ashcraft, 1989). Thus, EII can account for the forgetting-of-inappropriate-mental-sets theory by recontextualizing its knowledge using explicit and/or implicit processing (depending on the interference task demands; see Hélie et al., 2008).

Remote association. In EII, problem solving is performed simultaneously by explicit and implicit processes (Principle 2 of EII), and implicit associations are continuously being retrieved during the incubation phase (mostly without awareness, according to Principle 1 of EII). Hence, if the problem solver is doing an intelligent search (i.e., she or he does not blindly retrieve the same association over and over again), a longer incubation period will cover more fully the space of associations and make the retrieval of remote associations more likely. This provides a natural corre-

[4] We would like to thank an anonymous reviewer for pointing this out.

Table 2
Predictions of the Explicit–Implicit Interaction Theory

Phenomena	Gist of mechanistic predictions/explanations
1. Incubation primes lexical decision (Yaniv & Meyer, 1987).	Incubation leads to implicit processing (contextualized by preparation), which activates certain words used in subsequent lexical decision tasks.
2. Incubation increases reminiscence in free recall (S. M. Smith & Vela, 1991).	Incubation increases the number of words recalled in the second free-recall test due to implicit retrieval during incubation.
3. Not all participants can solve insight problems (Durso, Rea, & Dayton, 1994); solvers' and nonsolvers' knowledge structures differ.	Some participants generate more diverse hypotheses, which increase their probability of solving insight problems.
4. Implicit processing is overshadowed by explicit processing (Schooler, Ohlsson, & Brooks, 1993).	An overly explicit mode of processing can reduce the amount of implicit processing (e.g., by reducing the weight of implicit processing in knowledge integration).
5. Incubation is differently affected by distracting activities (Hélie, Sun, & Xiong, 2008).	The distracting activity and the main task may involve a variety of cognitive processes and may or may not use the same cognitive processes.

Note. The conceptual and mechanistic explanations (predictions) of the first four phenomena are detailed in the Simulations section of this article. The explanation (prediction) of the last phenomenon is presented in Hélie et al. (2008).

spondence to the remote association theory of incubation (S. M. Smith & Dodds, 1999).

Opportunistic assimilation. When an iteration of processing does not generate a solution for the problem, the provisional result is used to initiate another iteration of processing (Principle 5 of EII; Evans, 2006). New environmental information can be considered (if available) and contribute to the next iteration of processing. The provisional result can be interpreted as a memory trace encoding the unsolved problem (from the previous iteration of processing) and can affect the processing of newly available environmental information. Together, they may lead to the solution of the problem. This form of priming provides an explanation for the opportunistic assimilation theory of incubation (S. M. Smith & Dodds, 1999).

Reinterpretation of Existing Theories of Insight Using EII

Constraint theory. In some sense, EII views explicit and implicit processing as hard- and soft-constraint satisfaction algorithms (Principle 1 of EII). Explicit processing is responsible mainly for the satisfaction of hard constraints (following hard rules), whereas implicit processing is responsible mainly for the satisfaction of soft constraints (Sun, 1994). When a large number of constraints are simultaneously satisfied, a higher ICL is produced (because it represents a better interpretation of the problem; Bowers et al., 1990). As argued earlier, insight is produced by the sudden crossing of a confidence threshold by the ICL (an auxiliary principle of EII), and furthermore, this is more likely to happen after (and as a result of) implicit–explicit knowledge integration (Principle 4 of EII). The afore-described process captures the constraint theory of insight (e.g., Mayer, 1995; Schooler & Melcher, 1995).

Fixation theory. According to EII, when the associations that have been initially retrieved do not lead to congruency of explicit and implicit processing, an impasse is likely to be reached (because the ICL is likely to be below the threshold). As postulated by the fixation theory (S. M. Smith, 1995), this impasse must be broken by knowledge restructuring (Ohlsson, 1992; Schilling, 2005). In EII, knowledge restructuring amounts to recontextualization of subsequent processing. Restructuring can be achieved by either or both explicit and implicit processing (in an iterative way). When the conclusions from the previous iteration of processing are used to initiate a subsequent round of processing, the knowledge used in the subsequent iteration of processing is in some sense recontextualized and restructured, which may remove unwarranted constraints.

Associationistic theory. In EII, implicit and explicit memories are searched in parallel, which can lead to the retrieval of many (including unlikely) associations. This process is akin to what is described by the associationistic theory (Pols, 2002; see also the spreading activation theory of insight: Yaniv & Meyer, 1987).

Evolutionary theory. The evolutionary theory of insight involves the formation of previously nonexistent associations (Campbell, 1960; Simonton, 1995). In EII, this can be accomplished by randomly probing implicit knowledge structures (without considering the existing associations, e.g., by randomly selecting representations and assuming that they are related). This procedure is in line with the principles underlying the EII theory and captures the blind variation process essential to evolution (Darwin's first principle). Once formed, these assumed implicit associations are evaluated, and one of them is selected by the explicit processes (according to their ICL; an auxiliary principle of EII and Darwin's second principle). The selected association is either output to effector modules (e.g., such as motor module, if a threshold is crossed) or used as the input for the subsequent iteration of processing (if the threshold is not crossed; an auxiliary principle of EII). This use of the selected association captures the retention process (which is Darwin's last principle).

Summary

Having reinterpreted some existing theories of incubation (six theories in total) and some existing theories of insight (four theories in total), we now proceed to a computational model that captures existing human data. Developing a computational model ensures the consistency of the theory, allows the possibility of testing alternate versions of the theory (e.g., by varying the parameters), and makes the proposed representations and processes clear and unambiguous.

A Connectionist Model Implementing the EII Theory

The computational model implementing the EII theory is based on the CLARION cognitive architecture (Sun, 2002; Sun et al., 2001, 2005; Sun & Peterson, 1998; Sun & Zhang, 2004, 2006). In previous work, the non-action-centered subsystem of CLARION has already been used extensively for capturing data of human reasoning (e.g., Sun, 1994; Sun & Zhang, 2004, 2006). Because it is assumed in EII that the preparation and verification stages of creative problem solving are mainly captured by explicit rule-based reasoning of some form, the non-action-centered subsystem in CLARION can readily provide the model and thus explanations for these two stages (based on the previous work). Hence, the following modeling and simulations focus instead on Wallas's (1926) two other stages of creative problem solving, namely, incubation and insight.

Modeling Incubation and Insight Using CLARION

In this subsection, details of the computational model for capturing incubation and insight are presented, based on the non-action-centered subsystem of CLARION (Sun et al., 2005; Sun & Zhang, 2006). As noted earlier, the preparation and verification stages are not included in the present model. Therefore, explicit processing in the model is presented only to the extent sufficient for simulating the incubation and insight stages in these tasks. More complex rule-based reasoning has been covered in previous work on CLARION (e.g., logical reasoning, variable binding, hypothesis testing; see Sun, 1992, 1994; Sun & Zhang, 2006), and the interested reader is referred to these earlier pieces. Also, because incubation and insight do not involve much procedural knowledge, the action-centered subsystem of CLARION is not included in the model, but details concerning this subsystem can be found in, for example, Sun et al. (2001, 2005).

This section presents an overview of the computational model along with some key equations. The reader may skip the technical details in this section on first reading without losing the thread of the discussion. A more complete mathematical specification is provided in the Appendix using matrix notation.

The top level. A sketch of the model is shown in Figure 2. In the top level, explicit knowledge is represented with localist representations, that is, each node represents a different concept or hypothesis, and a link between two nodes stands for an explicit rule between the two represented entities. Overall, the top level may be viewed as a linear connectionist network (with two layers, i.e., two sets of nodes, in this particular case). The implementation of rule-based processing is rudimentary here but sufficient for our purpose—to capture many data sets pertaining to creative problem solving (as shown in the Simulations section later). Unlike in most other connectionist networks, neither layer in the top level of the model is the input or the output; both can be used to fill either role. In the following discussion, we assume that information initially enters the model from the left in Figure 2 and exits from the right (for similar equations describing the flow of information in the opposite direction, see the Appendix).

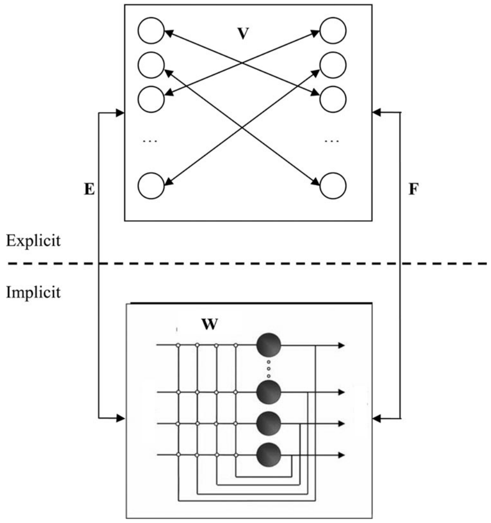


Figure 2. General architecture of the non-action-centered subsystem of CLARION. The letters refer to the connection matrices (see the Appendix for details).

Often, concepts are (redundantly) encoded in the bottom level (using distributed representations; Sun, 1994) and in the top level (using localist representations) of CLARION. If only the top-level representation is activated, a top-down signal may be sent to activate the corresponding representation in the bottom level (known as *implicitation*). Likewise, if the bottom level is activated, a bottom-up signal may be sent to activate the corresponding representation in the top level (known as *explicitation*). Hence, a stimulus is often processed in both the top and bottom levels in CLARION. The following equations describe the case where the top and bottom levels of CLARION are both activated to begin with. (Details of other cases are given in the Appendix.)

It should be noted that redundancy of representation (having equivalent forms of knowledge at the two levels) does not always implies coexistence or mutual activation across the two levels. Here, some alternative possibilities need to be pointed out. First, if the knowledge exists only at one level, there will be no mutual activation. Second, if equivalent knowledge does exist at the two levels but the link between them has not been established (i.e., the representational equivalence has not been established), there will be no mutual activation. Third, even when the representations across the two levels are linked, the links may not be used (e.g., due to distraction, lack of attention, low activation level, etc.). However, it is often the case that the equivalent forms of knowledge coexist at the two levels and that they will be able to access or activate each other (as generally hypothesized in CLARION; Sun et al., 2001, 2005; see also Sun, 1994, for detailed justifications). When mutual activation happens, complex insights are more likely to happen (as detailed later; see also Principle 4). These are the cases emphasized in this work.

When the left layer of the top level is activated, the activations are propagated in the top level using the following weighted sum:

$$y_i = \frac{1}{k1_i}\sum_{j=1}^{n} \mathrm{v}_{ij}\mathrm{x}_j, \tag{1}$$

where $\mathbf{y} = \{y_1, y_2, \ldots, y_m\}$ represents the activations of nodes in the right layer in the top level, $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ represents the activations of nodes in the left layer in the top level, $\mathbf{V} = (v_{ij})$ represents the (simplified) explicit rules (connecting x_j and y_i), and $k1_i$ is the number of nodes in the left layer that are connected to y_i ($k1_i \leq n$). Top-level node activations are binary (i.e., $x_j \in \{0, 1\}$ and $y_i \in \{0, 1\}$). Each node (x_j or y_i) represents an individual concept (using localist representations).[5] This transmission (node activation) rule ensures that the activation of y_i is equal to the proportion of its associates in the left layer (x_js) that are activated.[6]

The bottom level. In the bottom level of CLARION, implicit knowledge is represented by distributed activation patterns over a set of nodes and processed using a nonlinear attractor neural network (known as NDRAM; Chartier & Proulx, 2005). Computationally speaking, this network uses a nonlinear transmission function that allows the model to settle/converge into real-valued attractors (Hélie, 2008). The bottom-level representations are patterns (vectors) of bipolar node activations (i.e., $z_i = \pm 1$), and they are linked to the corresponding top-level representations by a set of linear equations (as detailed in the Appendix; see Equations A16 and A17).

The transmission in the bottom-level network is described by

$$z_{i[t+1]} = f(\sum_{j=1}^{r} w_{ij} z_{j[t]}), \tag{2}$$

where $\mathbf{z}_{[t]} = \{z_{1[t]}, z_{2[t]}, \ldots, z_{r[t]}\}$ is the state of the attractor neural network at time t (i.e., the activations of all the nodes in the network at time t), $\mathbf{W} = [w_{ij}]$ is a weight matrix that encodes the implicit associations among the nodes, and $f(\bullet)$ is a nonlinear function (as defined in the Appendix). Computationally speaking, this transmission rule is a regular synchronous update rule for attractor neural networks, and it ensures the convergence/settling of the model to a stable state (Hélie, 2008).

Transmission in the bottom level is iterative and remains in the bottom level until convergence/settling or a time limit is reached. Once one of these two criteria is met, the information is sent bottom-up (explicitation) for the integration of the two types of knowledge. Note that, following Sun and Zhang (2004), it has been estimated that each application of Equation 2 (called a *spin*) takes roughly 350 ms of psychological time.

Bottom-up explicitation. After the implicit processing is completed, the information is sent bottom-up in the following way:

$$y_{[\text{bottom-up}]i} = (k2_i)^{-1.1} \sum_{j=1}^{r} f_{ji} z_j, \tag{3}$$

where $\mathbf{y}_{[\text{bottom-up}]} = \{y_{[\text{bottom-up}]1}, y_{[\text{bottom-up}]2}, \ldots, y_{[\text{bottom-up}]m}\}$ represents the bottom-up activations of the nodes in the right layer of the top level (in Figure 2), z_j represents the activation of the jth node in the bottom level, $k2_i$ is the number of nodes in the bottom level (in $\mathbf{z}$) that are connected to $y_{[\text{bottom-up}]i}$ ($k2_i \leq r$),[7] and $\mathbf{F} = (f_{ij})$ is a weight matrix connecting the distributed bottom-level representations to their corresponding top-level representations (in the right layer of the top level: $\mathbf{y}$).[8] In words, Equation 3 translates the bottom-level activations into top-level activations (by reducing their dimensionality to m).

Integration. Once the bottom-up activation has reached the top level ($\mathbf{y}_{[\text{bottom-up}]}$), it is integrated with the activations already present in the nodes of the right layer ($\mathbf{y}$) of the top level using the Max function:

$$y_{[\text{integrated}]i} = \text{Max}[y_i, \lambda \times y_{[\text{bottom-up}]i}], \tag{4}$$

where $\mathbf{y}_{[\text{integrated}]} = \{y_{[\text{integrated}]1}, y_{[\text{integrated}]2}, \ldots, y_{[\text{integrated}]m}\}$ is the integrated activations of the nodes in the right layer of the top level and λ is a scaling parameter that determines how implicit the task processing is.[9]

The integrated activation pattern (vector) is further transformed into a Boltzmann distribution, which serves as the final activations of the corresponding nodes in this layer:

$$P(y_{[\text{integrated}]i}) = \frac{e^{y_{[\text{integrated}]i}/\alpha}}{\sum_{j} e^{y_{[\text{integrated}]j}/\alpha}}, \tag{5}$$

where α is a noise parameter (i.e., the temperature). The transformation above generates normalized activation patterns. In CLARION, each top-level node represents a hypothesis, and their normalized activation (the Boltzmann distribution) is the probability distribution of the hypotheses represented by these nodes. From this distribution (this set of final node activations), a hypothesis (a node) is stochastically chosen. Low noise levels in the equation above tend to exaggerate the probability differences, which lead to a narrow search of possible hypotheses and favor stereotypical responses. In contrast, high noise levels tend to reduce the probability differences, which lead to a more complete search of the hypothesis space (thus, the value assigned to α is constrained to be equal or larger during the incubation phase, compared with the preparation and verification phases; see, e.g., Martindale, 1995).

The statistical mode of the Boltzmann distribution (i.e., the probability of the most likely hypothesis or, equivalently, the

[5] Computationally speaking, this transmission rule is linear and represents the simplest case of neural networks. The normalizing factor ($k1_i$) prevents nodes with more incoming links from being more activated (on average).

[6] Note that, following Sun and Zhang (2004), it has been estimated that each application of Equation 1 takes roughly 1,500 ms of psychological time. However, top-level processing is done in parallel with bottom-level processing, so only the longest processing time is used (because the fastest process has to wait for the slowest for knowledge integration). In all the following simulations, the bottom-level processing time is slower (and used) because several iterations of bottom-level processing are performed before knowledge integration.

[7] The nonlinearity of the normalizing factor ($k2_i$) is used for capturing similarity-based processes not relevant to the present work. For detailed mathematical derivations, see Sun (1994).

[8] Note that in Equation 3, the transposition of the **F** weight matrix is used. This is because top-down processing (implicitation) uses the actual **F** weight matrix, whereas bottom-up processing (explicitation) uses the transposition of the **F** weight matrix (the same applies to the **E** weight matrix). Likewise, top-level processing from the left layer to the right layer uses the **V** weight matrix, whereas top-level processing from the right layer to the left layer uses the transposition of the **V** weight matrix. This use of transpositions to reverse the direction of processing substantially reduces model complexity (Kosko, 1988).

[9] The Max operator is often used to represent disjunction in fuzzy logic (Zadeh, 1988). Therefore, in a sense, the results of bottom-level and top-level processing are individually considered in the integrated activation vector.

activation of the most highly activated node) is used as the ICL as stipulated in EII.

Assessment. If the ICL (the statistical mode of the Boltzmann distribution) is higher than a predetermined threshold (i.e., ψ), the chosen hypothesis (represented by the chosen node) is output to effector modules (e.g., a motor module), and the response time (RT) of the model is computed:

$$\mathrm{RT} = a - b \times \mathrm{ICL}, \tag{6}$$

where a and b are the maximum response time and the slope of the response time curve, respectively. (Computationally, this equation was adopted because it is the simplest possible negative relation between the ICL and the response times, as found in Costermans et al., 1992, and Miner & Reder, 1994, and as previously used in, e.g., J. R. Anderson, 1991.)

If no response is output by the model (i.e., if the ICL is less than the predetermined threshold ψ), a new iteration starts with the chosen hypothesis (the chosen node) as the top-level stimulus. The result of previous implicit processing is treated as residual activations that are added to the bottom-up activations during the next iteration.

Specifically, the information in the right layer of the top level is sent back to the left layer of the top level:

$$x_i = \sum_{j=1}^{m} v_{ji} y_{[\mathrm{selected}]j}, \tag{7}$$

where x_i is the activation of the ith node in the left layer of the top level (see Figure 2), $\mathbf{y}_{[selected]} = \{y_{[selected]1}, y_{[selected]2}, \ldots, y_{[selected]m}\}$ is the bifurcated Boltzmann distribution in the right layer of the top level after a hypothesis has been chosen (i.e., the chosen hypothesis [node] has activation one and the remaining hypotheses [nodes] have activation zero), and v_{ji} is the explicit rule connecting y_j to x_i. Next, the result of Equation 7 is sent top-down to activate the bottom level (through implicitation; see the Appendix for mathematical details), and the processing starts anew (with a new iteration in the same way as described above). Intuitively, propagating the activation backward (from right to left in the top level) corresponds to abductive reasoning (i.e., if the chosen hypothesis in $\mathbf{y}$ is correct, what are the possible causes in $\mathbf{x}$?). Implicitation amounts to incorporating such possible causes into intuition (implicit processing). Starting a new iteration of processing after implicitation in both the top and bottom levels allows for inferences taking into consideration such possible causes (along with other information in the form of residual activations in the bottom level). This cumulating of inference processing is not random (even in high-noise conditions) because each new iteration of processing relies on the result from the previous iterations. The basic algorithm of CLARION is summarized in Table 3.

An example. We examine a prototypical insight problem-solving experiment using CLARION. In such an experiment, the participant provides an explanation for an ill-defined problem, such as follows (Schooler et al., 1993, p. 183):

> A giant inverted steel pyramid is perfectly balanced on its point. Any movement of the pyramid will cause it to topple over. Underneath the pyramid is a $100 bill. How would you remove the bill without disturbing the pyramid?

Table 3
Algorithm of the CLARION Model

1. Observe the current input information.
2. Simultaneously transmit the observed information in both levels (Equations 1 and 2).
3. Compute the integrated activation vector (Equations 3 and 4) and the hypothesis distribution (Equation 5).
4. Stochastically choose a response, and estimate the internal confidence level using the mode of the hypothesis distribution:
   a. If the internal confidence level is higher than a predefined threshold, output the chosen response to effector modules;
   b. Else, if there is time, go back to Step 1, and use the chosen response as the input (Equation 7).
5. Compute the response time of the model (Equation 6).

To capture and computationally explain this experiment with CLARION, the concepts (e.g., steel, pyramid, $100 bill, etc.) included in the problem description are represented in the left layer of the top level (see Figure 2), while the possible explanations are represented in the right layer of the top level. Because this is an open-ended problem, a large number of possible explanations (prior knowledge) are included in the model. Each concept and each explanation are represented by a different top-level node (in the left and the right layers, respectively), and these nodes are linked to form rules (representing culturally shared prior knowledge). Together, these nodes and links represent the explicit knowledge in the model.

In the bottom level of CLARION, each concept represented by a top-level node is also represented by a set of bottom-level nodes. Exemplars of the culturally shared explanatory rules coded in the top level are redundantly encoded in the bottom level (i.e., the attractor neural network is trained with these exemplars to create its corresponding attractors), which represents implicit knowledge in the model.

To simulate this task, the nodes representing the initial problem (in the left layer of the top level and the corresponding bottom-level representation) are first activated. This information is then transmitted to the right layer in the top level; in the meantime, the bottom-level activations are allowed to settle (converge). The stable state reached by the bottom level is sent bottom-up (explicitation) to be integrated with the top-level activations (in the right layer). The integrated activations are then transformed into a Boltzmann distribution (the final activations), and an explanation node is stochastically chosen on that basis. The statistical mode of the Boltzmann distribution is used to estimate the ICL, which is compared to a threshold. If the ICL is higher than the threshold, the chosen explanation is output to effector modules, and the process is over. Otherwise, the chosen explanation is sent backward to activate the left layer in the top level to infer possible causes for the chosen explanation (abductive reasoning). The activation in the left layer of the top level is used as the new stimulus to initiate another iteration of processing (to allow new inferences based on the possible causes). This iterative process ends when an explanation is output or the model runs out of time (if a time limit is given in the experiment).

Discussion

The CLARION model captures well the basic principles of the EII theory. First, the explicit knowledge in CLARION is represented in a localist fashion (i.e., one node = one concept), while implicit knowledge is represented using distributed representations. This is consistent with Principle 1 of EII (see Table 1) because this representational difference captures, to some extent, the difference of accessibility of explicit and implicit knowledge (see, e.g., Sun, 2002; Sun et al., 2001, 2005). Also, the top level of CLARION may carry out rule-based processing (of a rudimentary form in this particular work) because the links among the nodes (i.e., the connection matrix) may encode explicit rules of various forms (see, e.g., Sun, 1994; Sun et al., 2001). In contrast, the bottom level of the CLARION model implements a soft-constraint satisfaction algorithm because implicit knowledge in the bottom level is processed by an attractor neural network (Hertz, Krogh, & Palmer, 1991; Hopfield & Tank, 1985). These characteristics are exactly what is prescribed by EII (Principle 1 of EII).

Second, in line with Principle 2 of EII, most tasks are processed simultaneously in the top and bottom levels in CLARION. Simultaneous processing in CLARION is facilitated by the interlevel connections linking top- and bottom-level representations, which ensure that the corresponding representations in both levels are usually activated simultaneously (see the Appendix for further technical details). Hence, notwithstanding the location of initial activation, top- and bottom-level representations are usually simultaneously activated and processed in CLARION.

Third, each top-level node in CLARION corresponds to many bottom-level nodes; it is possible to encode (in different ways) the same associations in the bottom level and the top level. This characteristic of CLARION is in line with Principle 3 of EII. This redundant coding of information in CLARION is facilitated by the presence of top-down and bottom-up learning processes. For details of top-down learning and bottom-up learning in CLARION, which have been used to simulate a wide range of learning data, see Sun (2002) and Sun et al. (2001, 2005).

Fourth, because the top and bottom levels of CLARION process information differently, the results of top-level and bottom-level processing are integrated in CLARION. This integration process is in line with Principle 4 of EII and has been useful in modeling a wide range of human data in past simulations (e.g., Sun, 2002; Sun et al., 2001, 2005). Moreover, the integration function may lead to synergy, as amply demonstrated before (e.g., Sun & Peterson, 1998; Sun et al., 2005), which is in line with much human data in the implicit learning literature (e.g., Mathews et al., 1989; Stanley et al., 1989; Sun et al., 2001, 2005).

Fifth, processing in CLARION is iterative and bidirectional. The equations described in the preceding subsection go from left to right to left and so on (see Figure 2), although information flow can be reversed (see the corresponding equations in the Appendix). This is consistent with Principle 5 of EII. Moreover, the reasoning cycle in CLARION constitutes a generate-and-test process: Hypotheses are generated by the bottom level and made available by the explicitation process. Knowledge integration yields a Boltzmann distribution of hypotheses, and its mode is compared with predefined thresholds. This is the evaluation prescribed by EII.

In addition, the auxiliary principles included in EII are all incorporated into the CLARION model (see Table 1). Specifically, the mode of the Boltzmann hypothesis distribution is used to measure the ICL, as defined in the auxiliary principles of EII. Thresholds on the ICL (as discussed before) are used to choose between outputting a response and restarting the process. Response times in the CLARION model are a negative linear function of the ICLs, consistent with the auxiliary principles of EII and as suggested by empirical research (e.g., Costermans et al., 1992; Miner & Reder, 1994).

Simulations

Some experiments (i.e., Durso et al., 1994; Schooler et al., 1993; S. M. Smith & Vela, 1991; Yaniv & Meyer, 1987) that were mentioned at the beginning of this article to justify the notions of incubation and insight were simulated using the CLARION model. These experiments draw on well-established psychological paradigms (e.g., free recall, lexical decision, and problem solving) and are thus highly reliable. Given the broad scope of the approach in this article, the emphasis cannot be on extremely fine-grained modeling of the tasks involved. Hence, the simulations are coarser by necessity, which is inevitable given the nature of this approach.

All the simulation parameters are in Tables 4 and 5. Table 4 contains the task-related parameters, which were directly determined by the task input/output. Table 5 contains the free parameters, which, although not optimized, were tuned by hand using

Table 4
Task-Related Parameters Used in the Simulations

Parameter	Yaniv & Meyer (1987)	S. M. Smith & Vela (1991)	Durso, Rea, & Dayton (1994)	Schooler, Ohlsson, & Brooks (1993)
n	52	0	14	8
m	52	50	14	8
r	200	500	140	280
s	100	0	[a]	230
p	3	10	3	1
δ	0.10	0.49	0.40	0.40
Epochs	15	15	100	150

Note. n is the number of nodes in the left layer of the top level ($\mathbf{x}$), m is the number of nodes in the right layer of the top level ($\mathbf{y}$), r is the number of nodes in the bottom-level network ($\mathbf{z}$), s is the number of nodes in the bottom-level network that are connected to the left layer in the top level, p is the number of spins used to pretrain the bottom-level network, and δ is the slope of the transmission function in the bottom-level network. Other NDRAM parameters (i.e., the learning rate and memory efficiency) used to pretrain the bottom-level network were set to their default values throughout ($\eta = 0.001$, and $\zeta = 0.9999$; Chartier & Proulx, 2005).

[a] In the simulation of Durso et al. (1994), the same concepts were represented twice in the top level (once in each layer). Hence, a unique pool of bottom-level nodes was used.

Table 5
Free Parameters Used in the Simulations

Parameter	Yaniv & Meyer (1987)	S. M. Smith & Vela (1991)	Durso, Rea, & Dayton (1994)	Schooler, Ohlsson, & Brooks (1993)
λ	1.5	[a]	[a]	1.1
α	0.2	{0.06, 0.085}	$\{10^{-2}$–$10^{5}\}$	{0.12, 0.16}
Ψ	{0.715, 0.71, 0.69}	0.896	0.90	0.70

Note. λ scales the importance of implicit processing in the integrated activation, α is the temperature (randomness) in the Boltzmann distribution, and Ψ is the threshold on the internal confidence level.
[a] When no stimulus is presented to the model, processing has to be initiated from random activation in the bottom level. In these cases, the result of top-level processing is initially ignored by setting λ to a large value (e.g., $\lambda = 50$ or $\lambda = 500$).

reasonable values. (It should be noted that λ is constrained to be higher than 1 in the simulations below because the tasks were chosen to show the effect of incubation, which is mostly implicit.) Finally, because the verbal instructions (provided to participants before each experiment) and the preparation stage (as mentioned before) lead to contextualization of later processing of the incubation and insight stages (Wallas, 1926), only knowledge relevant to the simulated task was included in each simulation.

In each subsection below, an experiment is first reviewed in detail along with the resulting human data. Following each set of empirical data, the simulation setup is presented along with complete conceptual and mechanistic explanations of the task in the EII/CLARION framework. This is followed by the simulation results and a discussion of the implications of the simulation for creative problem solving in general and the EII theory in particular. A statistical threshold of α = .01 has been adopted throughout the article.

Incubation in a Lexical Decision Task

Yaniv and Meyer (1987) used a rare-word association task and a lexical decision task to test the unconscious work theory of incubation (Dorfman et al., 1996; S. M. Smith & Dodds, 1999). These tasks are detailed below.

Experimental setting. In Yaniv and Meyer's (1987) Experiment 1, each trial was initiated by the presentation of a definition to the participant, who had 15 s to find the associated word (the rare-word association task). If the participant was able to produce the associated word, the lexical decision task started. If the participant did not find the associated word, she or he was asked to estimate the FOK prior to starting the lexical decision task.

In the lexical decision task, the participant's task was to identify strings of letters as words or nonwords. Each rare-word association trial was followed by a block of six lexical decision trials. The block was composed of three types of strings: unrelated words (distractors), nonwords, and the response to the rare-word association trial (the target word). The prediction was that the participants who were unable to provide an answer in the rare-word association task but had a high FOK would be primed for the target word in the lexical decision task (leading to faster response times), but not those who had a low FOK. Likewise, correct responses in the rare-word association task would prime the target, but incorrect responses would not. If these were the cases, one might interpret these results as an indication that incubation is a form of unconscious processing and that incomplete (unconscious) processing might be sufficient to prime a target word (despite the failure to produce the target word).

Experimental results. Forty-four participants were tested in 52 rare-word association trials and associated $52 \times 6 = 312$ lexical decision trials. The results of interest were those obtained in the lexical decision task, factorized by the performance in the rare-word association task. As predicted, correct responses in the rare-word association task primed the target word in the lexical decision task, $t(2100) = 8.5$, $p < .001$. In contrast, incorrect responses in the rare-word association trial did not affect the performance of the subsequent lexical decision task (i.e., no priming for the target word), $t(2100) = 0.7$, *ns.*[10] In trials in which no response was given in the preceding rare-word association task, analyses of the response times showed a significant interaction between the FOK and the type of stimuli (targets vs. distractors), $t(1648) = 2.28$, $p < .05$. Gamma correlation coefficients (provided by Yaniv and Meyer, 1987) suggested that targets were faster than distractors when the FOK was high; this relation was reversed when the FOK was low (see Figure 3).

Simulation setup. In the top level of the CLARION model, the left layer was used to represent the words, while the right layer represented the definitions (see Figure 2). Each word and each definition were represented by a different node (i.e., using localist representations), and each word was associated to its definition by a link within the top level. In the bottom level of the CLARION model, half of the nodes were used to represent the words, while the remaining nodes were used to represent the definitions (both with distributed representations). Each word/definition was represented by randomly generated activation patterns in the bottom level.[11] The bottom-level network was pretrained to encode the associations between the words and their corresponding definitions. The values given to the task-related parameters were as shown in Table 4.[12]

[10] It should be noted that mean response time data in the lexical decision task following correct and incorrect responses in the rare-word association task were not available in Yaniv and Meyer (1987); only the test statistics were reported.

[11] Note that random representations were generated each time. One seed representation was generated once (randomly generated). However, a Gaussian noise vector ($\mu = 0$, $\sigma = 0.01$) was added each time to the definitions to represent individual differences. This corresponds to people (say, from the same linguistic community) using words relatively consistently in communication but possibly with some relatively minor individual variations.

[12] Note that the number of epochs used to pretrain the bottom level was kept to a minimum to represent the rareness of the associations. This increased the number of spins necessary for convergence in the bottom level during performance. If the associations used had not been rare, more epochs would have been used to pretrain the bottom level, convergence would have been faster during performance, and all the definitions would have been found within the allotted time.

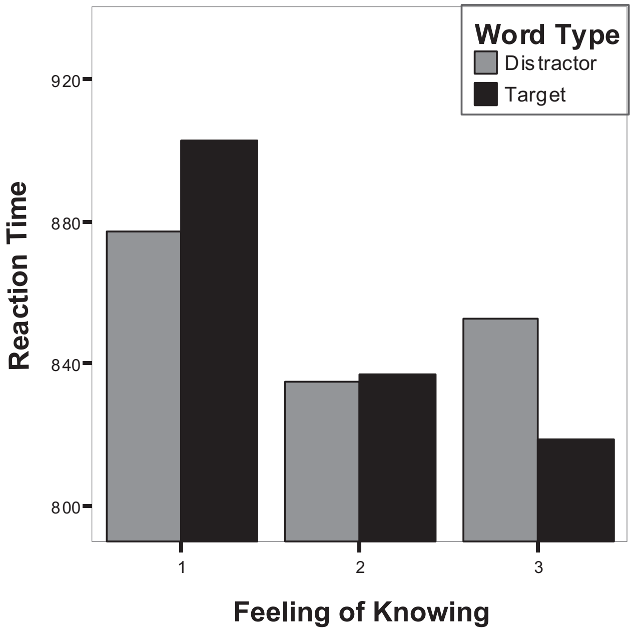


Figure 3. Lexical decision task response times from Yaniv and Meyer's (1987) Experiment 1 when no response was produced in the rare-word association task. The *x*-axis represents the feeling of knowing as measured after the rare-word association task, and the *y*-axis represent the response times (in ms) in the lexical decision task. Adapted from "Activation and Metacognition of Inaccessible Stored Information: Potential Bases for Incubation Effects in Problem Solving," by I. Yaniv and D. E. Meyer, 1987, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* p. 194. Copyright 1987 by the American Psychological Association.

To simulate a rare-word association trial, a stimulus activated the right (definition) layer in the top level and the corresponding representation in the bottom level. Explicit rules were applied in the top level (in this case, amounting to retrieving definition-word associations), and the information in the bottom level was processed for 42 spins[13] (with roughly 350 ms per spin, as hypothesized earlier; see also Libet, 1985), approximating the fact that human participants had 15 s. Following this processing, the outputs from both levels were integrated using the parameters shown in Table 5 and transformed into a Boltzmann distribution (which served as the activations of the nodes in the left layer of the top level). The statistical mode of the distribution (the maximum activation of the left-layer nodes in the top level) was used to estimate the ICL. Because there was no time for further iteration, a response was output if the ICL was higher than the first threshold in Table 5. Otherwise, no answer was provided, and the FOK was estimated using the ICL (as in the human experiment). If the ICL was higher than the second threshold in Table 5, the FOK was estimated as high, and if the ICL was lower than the last threshold in Table 5, the FOK was estimated as low. The remaining range was rated as medium.

For simulating the lexical decision task, three types of stimuli had to be represented. The target was the same word used in the corresponding rare-word association trial, and the distractors were the words used in other trials of the rare-word association task. Nonwords used randomly generated representations (real values within [0, 1]). Note that words (either distractors or targets) were represented explicitly in the top level (in the left layer), whereas nonwords were not.

Following each rare-word association trial, six lexical decision trials were conducted. Because the stimuli were presented rapidly, a normally distributed noise pattern (a noise vector) was added to each stimulus ($\mu = 0, \sigma = 0.05$). The information was transmitted within the CLARION model as follows. First, a stimulus activated a node in the left layer of the top level of the CLARION model and the corresponding implicit (bottom-level) representation. Activations were transmitted simultaneously in the top level and the bottom level. The bottom level underwent six spins, as human participants had a maximum of 2 s ($6 \times 350 = 2{,}100$ ms). Residual activations from the end of the rare-word association trial were present in the bottom level, which added to the result of current bottom-level processing (for technical details, see the Appendix). The output from the bottom level was integrated with the activations in the right layer of the top level using the Max function and transformed into a Boltzmann distribution (which served as activations for the nodes in the right layer of the top level). A response was stochastically selected, and the ICL was computed (as explained before) and used to estimate the response time of the model (with $a = 1{,}530$ and $b = 1{,}380$). Note that no threshold was used on the ICL because a response had to be output.

Rationale and explanations.

Conceptual explanation based on EII. According to the EII theory, a rare-word association trial produces a simultaneous search at the explicit and the implicit levels (Principle 2 of EII). 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 does not necessarily lead to solutions). In contrast, according to EII, implicit memory search is more likely to retrieve the desired association if given enough time because soft-constraint satisfaction can allow 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 (for the no-response participants). 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 is in line with Yaniv and Meyer's (1987) results.

Mechanistic explanation based on CLARION. In the CLARION-based computational simulation, the FOK from the rare-word association trial should have a strong effect on the response time to the target during the subsequent lexical decision trials because the FOK is represented by the ICL (which estimates

[13] A spin is a round of synchronous updating of all the nodes in the bottom-level neural network (i.e., an application of Equation 2; see The Bottom Level subsection).

the efficiency of processing toward the target word during the rare-word association trial). Hence, in a sense, the residual activations determine the amount of priming in the lexical decision trials. If the priming is relevant to a lexical decision trial (i.e., if the stimulus in the lexical decision trial is the target), the residual activations reduce the response time because the new bottom-level processing is consistent with the result of previous bottom-level processing and additively combining the activations magnifies the effect. In contrast, when the stimulus in the lexical decision trial is not the target (i.e., it is a distractor or a nonword), the priming is irrelevant, and the residual activations can increase the response time because previous processing (toward the target) can result in more noise in the Boltzmann response distribution. This mechanistic explanation leads directly to Yaniv and Meyer's (1987) results.

Simulation results. Three thousand simulations were run (each corresponding to a human participant in the experiment), each containing 52 rare-word association trials (each stimulus was seen once during the rare-word association trials), each followed by six lexical decision trials, exactly as in the human experiment. Figure 4a shows the response times in the lexical decision trials split by performance in the rare-word association trials (correct vs. incorrect) and stimulus type. As can be seen, targets were recognized faster than distractors in the lexical decision task when the correct response was provided in the rare-word association task (as predicted). Targets and distractors also slightly differed when an incorrect response was given in the rare-word association task, but the response time difference was much smaller in this case.

The same statistical analyses as in Yaniv and Meyer (1987) were performed on the simulated participants. Target recognition was significantly faster than distractor recognition when a correct response was given in the rare-word association task, $t(2976) = 6.87$, $p < .0001$, as in Yaniv and Meyer's results. As in Yaniv and Meyer's results, this difference between targets and distractors was not statistically significant when an incorrect response was given in the rare-word association task, $t(2221) = 1.91$, *ns*. This suggests that the small difference between target and distractors when an incorrect response was produced in the rare-word association task may be attributed to random variation (especially considering the high statistical power from several thousand simulated participants). These simulation results above are all in line with the results from Yaniv and Meyer's Experiment 1.[14]

Of more interest are the trials in which no response was given in the rare-word association task. Figure 4b shows the response times in the lexical decision trials split by FOK and stimulus type. As can be seen, the FOK (from the corresponding rare-word association trial) had a strong effect on the difference between response times to targets and distractors. As predicted, targets were faster than distractors when the FOK was high, but this relation was reversed for low FOK. The interaction in a Stimulus Type × FOK analysis of variance (ANOVA) reached statistical significance, $F(2, 5998) = 42.87$, $p < .0001$. Further decomposition of the analysis showed that targets were faster than distractors when the participants rated their FOK as high, $F(1, 2999) = 12.51$, $p < .0001$. The opposite effect was found for low FOK: Distractors were faster than target words, $F(1, 2999) = 67.85$, $p > .0001$. These statistically significant differences were not present for medium FOK, $F(1, 2999) = 5.05$, *ns*. All these results are in line with Yaniv and Meyer's (1987; see Figure 3).

Discussion. The simulation results obtained with the CLARION model matched well the human data of Yaniv and Meyer (1987). The reproduction of these qualitative and quantitative results supports the psychological plausibility of the proposed model and the adequacy of the EII theory. Several effects were simultaneously reproduced without varying the free parameters across tasks. The model was not designed specifically to simulate this task but was well supported by fundamental theoretical considerations (Sun, 2002). Overall, CLARION captured and mechanistically explained the human data demonstrating the effect of incubation in a lexical decision task.

Incubation in a Free-Recall Task

S. M. Smith and Vela (1991) studied the effect of incubation on the number of new words recalled during the second free-recall phase in a two-phased free-recall experiment. This measure is referred to as *reminiscence*.

Experimental setting. The participants had 5 min 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 min. Once the first free-recall test was completed, the participants had a 0-, 1-, 5-, or 10-min 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 restudying them). Two hundred twenty-one participants were tested in this 3 × 4 design, and the dependant variable was reminiscence.

Experimental results. A Test Duration × Incubation Interval ANOVA was performed on reminiscence (see Figure 5a). There was no effect of test duration, $F(2, 209) = 0.27$, *ns,* but incubation interval had a significant effect on reminiscence, $F(3, 209) = 9.40$, $p < .01$. 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 min) yielded similar reminiscence scores and that these scores were smaller than those obtained for longer incubation intervals (5 and 10 min, respectively, which did not differ statistically). Subsequent experiments showed that the effect of the incubation interval on reminiscence was significant only during the first minute of the second free-recall test (S. M. Smith & Vela, 1991).

Simulation setup. To simulate this task, only the right layer was used in the top level of CLARION (see Figure 2), and each node represented a different line drawing (word).[15] In the bottom level of CLARION, all the concepts (each represented by a top-level node) were encoded with a common pool of nodes using distributed representations, and a different bottom-level distributed representation was randomly generated for each line drawing (word).

To simulate the first recall test, a random pattern activated the bottom level, and the activations were propagated within the

[14] While we could not directly compare the simulated response times with human data, the estimates were reasonable, and all the statistical effects were reproduced.

[15] This can be accomplished in the CLARION model by setting the number of nodes in the left layer to zero and the knowledge integration parameter to a large value (causing top-level rules to be ignored).

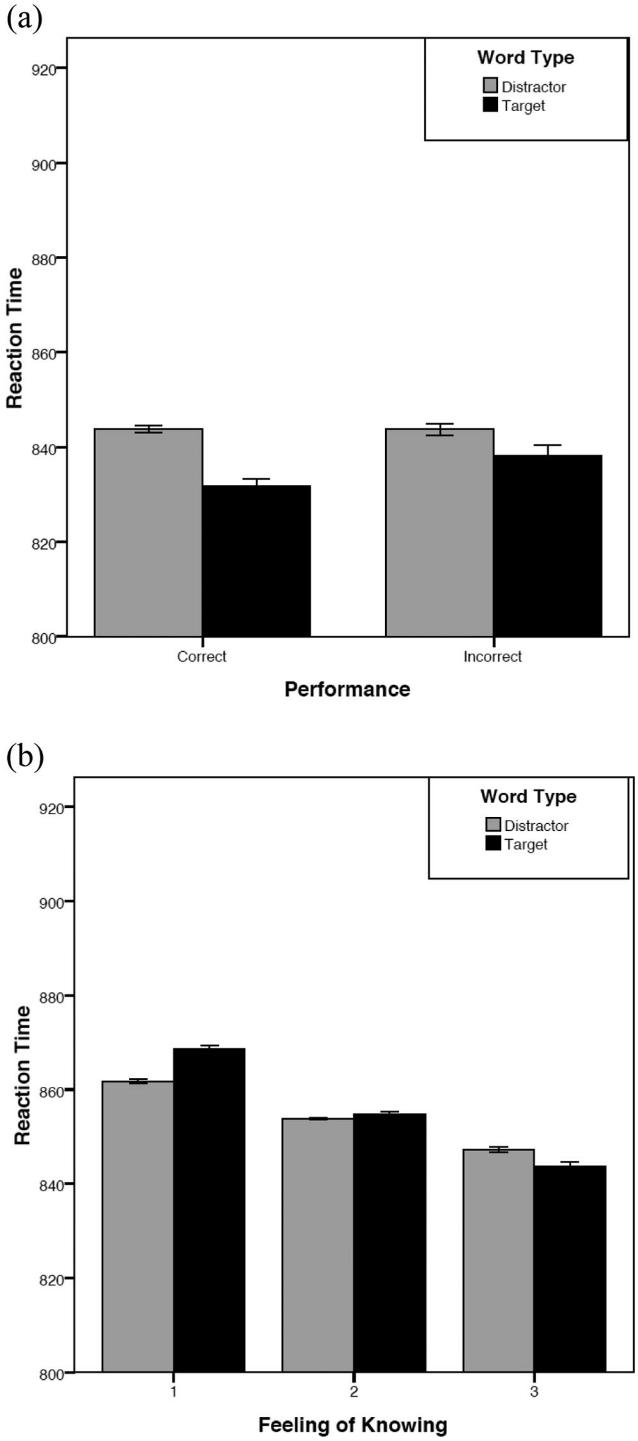


Figure 4. a: Simulated response times in the lexical decision task when an answer was given in the rare-word association task. The *x*-axis represents the performance in the rare-word association task (correct vs. incorrect), and the *y*-axis represent the response times (in ms) in the lexical decision task. b: Simulated response times in the lexical decision task when no answer was given in the rare-word association task. The axes are the same as in Figure 3. In both panels, error bars indicate standard errors.

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 Boltzmann distribution (using the first noise-level value from Table 5). The mode of the distribution (the maximum activation of the right layer of the top level) was used to estimate the ICL, which was compared with the threshold. 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 to start the process again. As in all other simulations, each spin in the bottom level took 350 ms of psychological time. Hence, the durations of recall were 171, 343, and 686 spins (for 1, 2, and 4 min, respectively).

Simulation of the incubation period was basically the same as that of the first recall test except for the following: (a) The noise level in the Boltzmann response distribution was increased to the second value in Table 5. (b) If an item was recalled, it was stored in a buffer memory (J. R. Anderson & Milson, 1989). The incubation intervals were 0, 171, 857, and 1,714 spins (for 0, 1, 5, and 10 min, 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 the human experiment of S. M. Smith and Vela (1991), most words were recalled during the first minute of the second test (as mentioned earlier). The CLARION parameter values were as shown in Tables 4 and 5.

Rationale and explanations.

Conceptual explanation based on EII. According to the EII theory, parallel memory searches are conducted in explicit and implicit memories during the free-recall tests (Principle 2 of EII). However, the incubation period is different: Principle 1 of the EII theory stipulates that explicit memory search requires more attentional resources, whereas implicit memory search is mostly automatic. Thus, mostly implicit processes are deployed during the incubation phase, and words are being retrieved from implicit memory during that period (but not much from the explicit memory). These additional words are output at the beginning of the second test, increasing the number of words recalled in this second test (but not the first test). According to the EII theory, reminiscence increases as the number of words recalled in the second test becomes larger compared with the number of words recalled in the first test (on the average, i.e., by statistical facilitation). This conceptual explanation is in line with S. M. Smith and Vela's (1991) results.

Mechanistic explanation based on CLARION. In CLARION, words are being generated (recalled) from the bottom level during the recall tests. Because the eventual effect of incubation is to increase the number of words recalled during the second recall test (but not the first; see the previous conceptual explanation), the likelihood of recalling new items in the second test should be increased due to incubation. In contrast, test length should affect the total numbers of recalled items during both the first and second tests (i.e., the effects of test length should be roughly the same for the two recall tests). This is different from the effect of incubation because it does not change much the likelihood of recalling new (additional) words during the second recall test (because the key to increasing the likelihood of recalling new words is the difference between the numbers of words recalled during the two recall tests).

As a result, only incubation should increase reminiscence in the CLARION simulation, which is consistent with S. M. Smith and Vela's (1991) human data.

Simulation results. Twelve thousand simulations were run (1,000 in each of the 12 conditions). The results are shown in Figure 5b. As predicted, the mean reminiscence scores were positively affected by the incubation length. The mean reminiscence scores were 1.73, 2.03, 3.04, and 3.81 for the 0-, 1-, 5-, and 10-min incubation intervals, respectively, which is similar to the human data. However, the duration of the recall tests did not show such a clear pattern. In particular, test duration seemed to have a positive effect on reminiscence when there was no incubation interval, no effect for moderate incubation intervals (1 and 5 min), and a negative effect for a long incubation interval (10 min). However, unlike the effect of incubation interval, all these effects of test duration are small: The biggest group difference within each incubation level is smaller than one word (0.739). A Test Duration × Incubation Interval ANOVA was performed on the reminiscence scores to confirm these observations. First, the incubation length had a significant effect on reminiscence, $F(3, 11998) = 1{,}661.34$, $p < .0001$, as in the human data. Post hoc Tukey analyses showed that all incubation levels were statistically different ($p < .0001$). Second, the main effect of test duration did not reach statistical significance, $F(2, 11988) = 0.78$, *ns,* as in the human data.[16]

Discussion. CLARION was able to reproduce the effect of incubation on reminiscence found by S. M. Smith and Vela (1991). The main difference between the simulation and the human experiment was that the simulation made the simplifying assumptions that (a) words are recalled independently during each recall test and (b) the two recall tests are independent. This does not seem to be the case with human participants as many effects of words and test dependencies (e.g., priming) have been observed (e.g., Cohen, 1963). These differences could probably be resolved by modeling the dependencies between the words (e.g., by using top-level rules or the correlation between the bottom-level representations) and by adding a recency-based base-level activation (J. R. Anderson & Milson, 1989). However, the focus of this simulation was not to capture the minute details of free recall.[17] Overall, CLARION was successful in capturing the data concerning the effect of incubation on reminiscence in a free-recall experiment.

Insight in Problem Solving

Experimental setting. Many theories of insight assume that insight is the consequence of knowledge restructuring (e.g., Mayer, 1995; Pols, 2002; Schilling, 2005; Schooler & Melcher, 1995; S. M. Smith, 1995). Because declarative knowledge has often been modeled using graphs (Schilling, 2005), Durso et al. (1994) hypothesized that insight could be observed by constructing and comparing participants' knowledge graphs before and after insight had occurred. To test this hypothesis, the participants were asked to explain the following story:

> A man walks into a bar and asks for a glass of water. The bartender points a shotgun at the man. The man says, "Thank you," and walks out. (Durso et al., 1994, p. 95)

The participants' task was to explain why the sight of the shotgun replaced the man's need for a glass of water (i.e., because he had the hiccups). To explain this story, the participants had 2 hr to ask the experimenter yes–no questions. After this questioning period, the participants were split into two groups (solvers and nonsolvers) and asked to rate the relatedness of pairs of concepts using a Likert-type scale (Likert, 1932). These ratings were used to construct the solvers' and nonsolvers' knowledge graphs (via the Pathfinder scaling algorithm; Schvaneveldt, Durso, & Dearholt, 1989).

Experimental results. Twelve participants tried to explain the story, and only half of the participants successfully accomplished the task. The resulting aggregated knowledge graphs were as shown in Figure 6. As can be seen, the solvers' knowledge graph (see Figure 6a) differed from the nonsolvers' (see Figure 6b) by 12 edges. These differences reflected a shift of the focal points of the graph (i.e., the center and median of the graph) from "Bartender" to "Relieved." Furthermore, the correlation between the two graphs was essentially zero (Durso et al., 1994). (Note that no further statistical analysis was provided by Durso et al. 1994.)

Simulation setup. To simulate this task, each concept was represented by a separate node in each layer in the top level of CLARION. The nonsolvers' graph (see Figure 6b) was assumed to represent common prior knowledge (i.e., common prior semantic associations) and was thus precoded as rules in the top level (linking corresponding nodes across the two layers of the top level) before the simulation started. In the bottom level, each concept was represented using 10 nodes (with distributed representations, which were randomly generated). The explicit rules (in the top level) were also coded as implicit associations in the bottom level (i.e., the idea of redundant encoding) through pretraining the bottom-level network using example stimuli consistent with the top-level rules. The values given to the parameters were as shown in Tables 4 and 5.

As in the simulation of the free-recall task (S. M. Smith & Vela, 1991; see the subsection Incubation in a free-recall task), hypothesis generation was initiated by a random activation pattern in the bottom level of CLARION. This pattern was further processed by the neural network (by repeatedly applying the nonlinear transmis-

[16] The interaction between the factors also reached statistical significance, $F(6, 11988) = 59.70$, $p < .0001$. 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). This difference could probably be resolved by directly modeling the upper limit of short-term memory capacity using an extra parameter (to avoid a ceiling effect caused by the upper limit of short-term memory being set by the complex interaction of several free parameters) and adding more short-term memory details (see also footnote 17).

[17] It should be noted that the emphasis of this simulation was on simulating the effect of incubation on reminiscence, not free recall per se. A more complete simulation of this experiment would require more complex memory processes (such as short-term memory) and more parameters. Although available in the CLARION cognitive architecture (Sun, 2002), these detailed memory processes and parameters are not included here for the sake of focus and clarity. As a result, a detailed comparison of the simulated numbers of recalled words with corresponding human data is not attempted here.

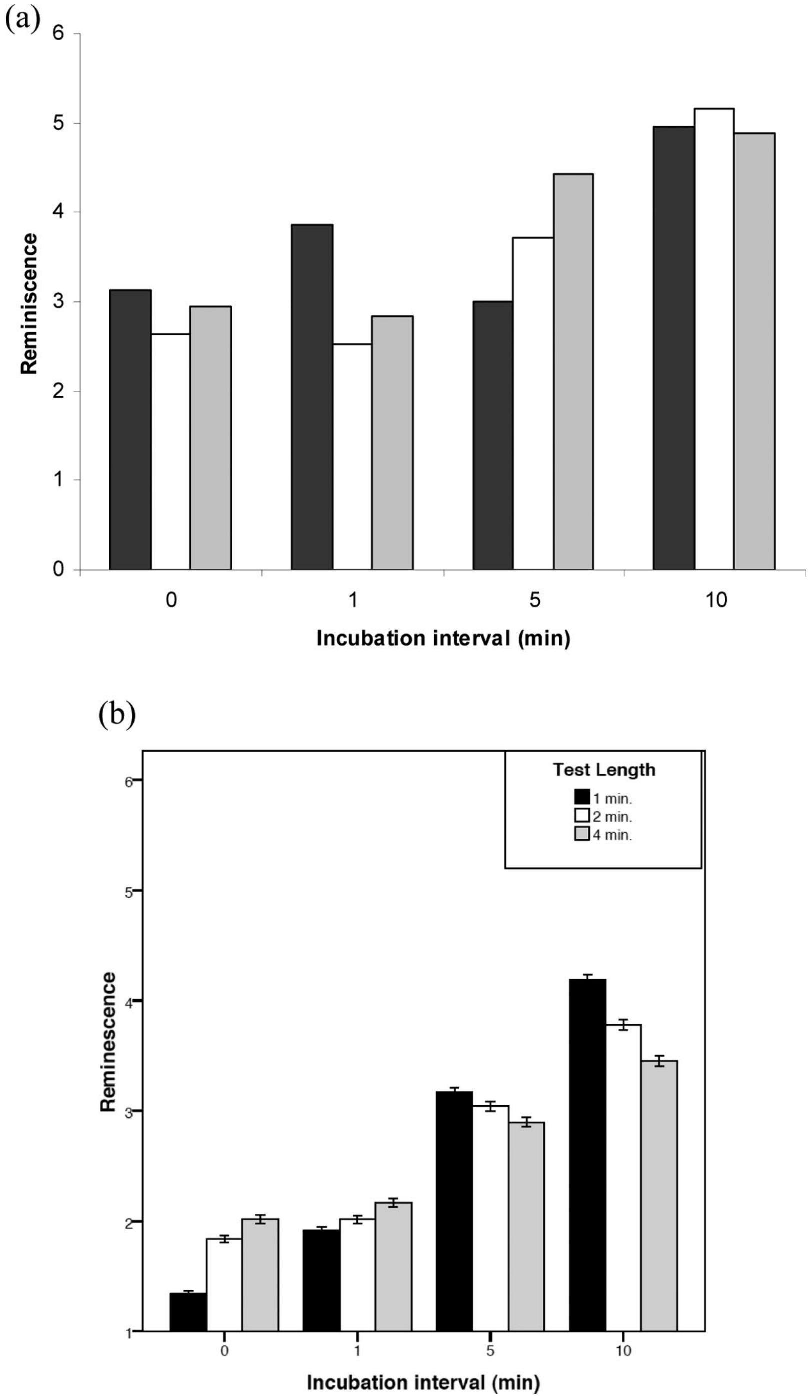
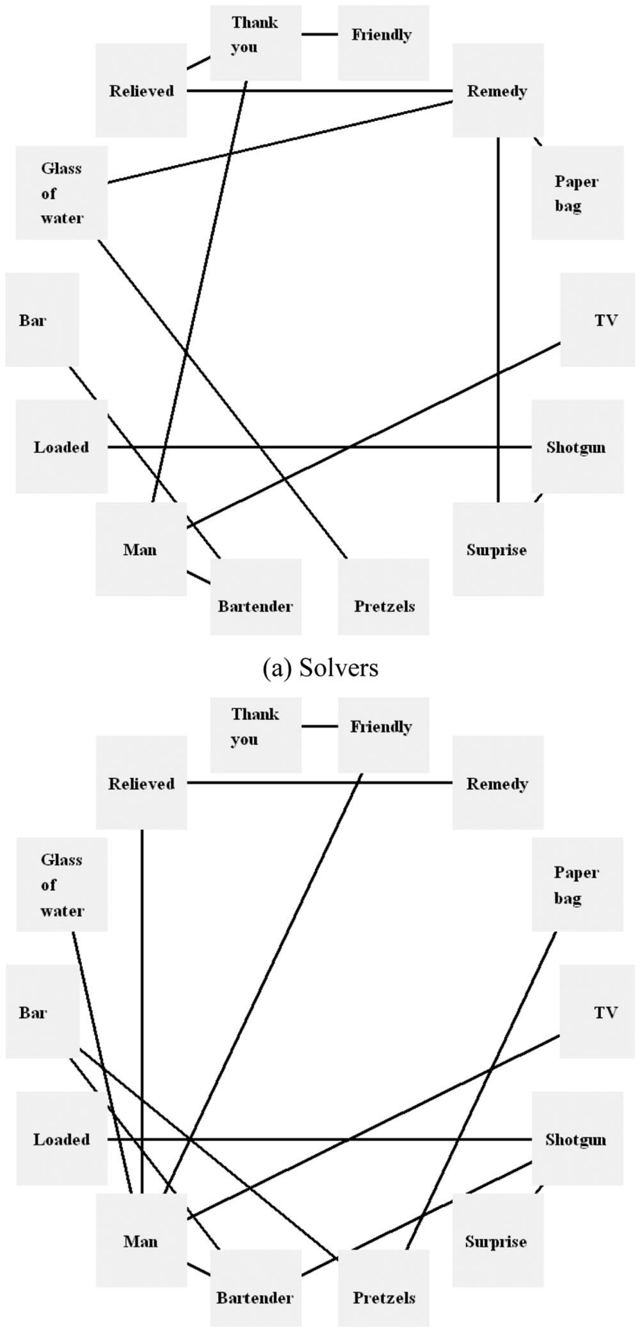


Figure 5. a: Reminiscence effect found in S. M. Smith and Vela's (1991) Experiment 1. b: Simulated reminiscence effect. Error bars indicate standard errors. In both panels, the black bars represent 1-min tests, the white bars represent 2-min tests, and the grey bars represent 4-min tests.

sion function) until convergence (settling) of the bottom-level neural network. The resulting stable state was sent bottom-up to activate the right layer of the top level (through explicitation) and integrated with the result of explicit processing (although the scaling parameter was set to ignore the effect of rule-based processing due to the absence of rules relevant to finding the solution to this problem). The integrated result was transformed into a Boltzmann distribution, which served as activations for the right-layer nodes in the top level. The mode of the distribution (the maximum activation in the layer) was used to estimate the ICL to determine if a question was to be asked of the simulated experimenter.



(a) Solvers

(b) Non-solvers

Figure 6. Knowledge graphs inferred by the participants in Durso, Rea, and Dayton's (1994) Experiment 1.

If the ICL was higher than a chosen threshold, a question was to be asked of the simulated experimenter. Questions concerned a direct link between a node in the right layer and a node in the left layer (both in the top level). A node in the right layer was first stochastically chosen based on the Boltzmann distribution in place; then, abductive reasoning was performed to activate the left layer in the top level; the activations in the left layer were also transformed into a Boltzmann distribution (which served as activations of the left-layer nodes), and a node from the left layer was stochastically chosen. A question was then asked concerning the existence of a link between the chosen node in the right layer and the chosen node in the left layer (both in the top level). If the answer was yes (i.e., when the link was present in the solvers' graph as shown in Figure 6a), a new rule (link) was added to the top level (i.e., to the explicit knowledge of the simulated participant); otherwise, the corresponding rule (if it existed in the top level) was removed from the top level. If the ICL was too low to come up with a question to the experimenter, the same procedure was used, except that the top-level weight matrix (explicit knowledge) was not modified.

In all cases, the activations present in the left layer of the top level were sent top-down (i.e., implicitation) for another iteration of processing (see the Appendix for mathematical details). This iterative process ended if a solution was found (i.e., the top-level explicit knowledge of the simulated participant was identical to the solvers' graph; see Figure 6a) or 20,571 iterations had occurred in the bottom level ($20{,}571 \times 350$ ms $= 7{,}199{,}850$ ms ≈ 2 hr, as in the human experiment).

Most associationistic theories of insight argue that a more diffused search in memory is more likely to yield a creative solution (e.g., Campbell, 1960; Martindale, 1995; Mednick, 1962). In CLARION, this phenomenon may be captured using the noise (temperature) parameter in the Boltzmann distribution (see the Modeling Incubation and Insight Using CLARION subsection). The present simulation served to test the adequacy of this hypothesis. The noise (temperature) level, used for selecting a node based on the Boltzmann distribution, was varied between 10^{-2} and 10^{5} (with an increment of 1 for the exponent). One thousand simulations were run with each of these noise levels.

Rationale and explanations.

Conceptual explanation based on EII. After the participant is read the story, she or he engages in explicit memory retrieval and implicit memory search (incubation). However, explicit processing is mostly rule based (Principle 1 of EII), which only brings up stereotypical semantic associations from the words included in the story. In contrast, the gradient of associations is flatter in implicit memory (Martindale, 1995; Mednick, 1962): The search is more diffused, and thus, more remote (creative) associations can be retrieved using soft-constraint satisfaction (Hadamard, 1954). Hence, according to the EII theory, implicit processing allows the retrieval of approximate hypothetical associations that differ from those retrieved explicitly. These implicit associations are then integrated with the result of explicit processing (Principle 4 of EII). If the chosen integrated association is deemed plausible (i.e., if the ICL is high enough), a question concerning the validity of this association is put to the experimenter. If the experimenter confirms the association, it is added into explicit knowledge; otherwise, it is removed. This process is iterated, and explicit and implicit processing are reinitiated with the new state of the knowledge. This iterative process ends when the participant finds the correct solution or the allowed time elapses.

Mechanistic explanation based on CLARION. During the questioning period, a random activation pattern is used to initiate processing and randomly sample the implicit associations (i.e., the preexisting implicit knowledge; see Figure 6b). However, each time an implicit association is sent bottom-up (through explicitation), noise is added in constructing the Boltzmann distribution, and a hypothesis is stochastically chosen. Low noise should result in a higher probability of choosing the most likely hypothesis according to the existing knowledge structure, which tends to be uncreative (and often counterproductive in this particular context). However, when more noise is added during the construction of the Boltzmann distribution, hypotheses that are somewhat inconsistent with the currently existing knowledge structure are more likely to be sampled. This can lead to altering the connection patterns of the top-level explicit knowledge structure (through questions and answers as described earlier), which may eventually lead to something resembling the correct solution (the solvers' knowledge graph; see Figure 6a). This process constitutes a typical generate-and-test algorithm (Russell & Norvig, 1995). This mechanistic explanation is in line with the human results obtained by Durso et al. (1994).

Simulation results. The mean performance by noise level was as shown in Figure 7. As can be seen, the CLARION model was generally able to modify its explicit representations based on yes–no questions to the experimenter. As predicted, this ability to produce a new explicit representation of the problem was positively related to the noise level (α) in low noise conditions but leveled off with a reasonable amount of noise. This was confirmed by a between-subject ANOVA. The effect of the noise level on the mean number of edges differing between the solution proposed by CLARION and the solvers' graph (see Figure 6a) was highly significant, $F(7, 7992) = 12{,}193.40$, $p < .0001$. More precisely, Tukey post hoc analyses showed that low noise levels resulted in poor performance and that each increment between 10^{-2} and 10^{0} significantly improved the performance of the model ($p < .01$). From that point on, increasing the noise level did not significantly improve the performance of the model. Overall, the noise (temperature) parameter in CLARION changed the probability of correctly solving an insight problem, which is in line with associationistic theories of insight. (Note that the absence of statistical analysis in Durso et al., 1994, limited our ability to compare the simulated data with the human data. However, the simulated results clearly showed the shift of the problem representation in the model, as in the human data.)

Discussion. The preceding analysis showed that the performance of CLARION improved as the noise level was increased. Martindale (1995) equated creativity to a more thorough exploration of the solution space, which increased the probability of finding creative solutions to ill-defined problems. This can be modeled by altering the noise level (i.e., temperature, stochasticity) in the search process. A lower noise level suggests a more timid exploration of the solution space and thus the generation of common, uncreative solutions (which may not solve ill-defined problems). Adding more noise (increasing stochasticity) initiates a more complete investigation of the possible solutions, thus allowing less frequent solutions to be sampled. These infrequent solutions might be responsible for insight (see also the associationistic theory of insight; e.g., Pols, 2002).

In particular, this interpretation of the simulation results is in line with the evolutionary theory of insight (e.g., Campbell, 1960; Simonton, 1995). According to this theory, noisy hypotheses are implicitly generated and explicitly evaluated (somewhat similar to the heuristic-analytic theory; Evans, 2006). According to this theory, more creative individuals would implicitly generate a greater number of hypotheses. In modeling, this amounts to creative and uncreative people having different noise (temperature/stochasticity) levels (with the former being modeled with a higher noise level). To summarize, CLARION was successful in simulating the search process leading to insight in problem solving, and the simulation results were in line with previous theories of insight and creativity. This constitutes converging evidence for the EII theory.

Overshadowing in Problem Solving

The implicit learning literature has repeatedly shown that explicitly looking for rules and regularities can impair performance when none exist or when they are difficult to extract (Berry & Broadbent, 1988; Reber, 1989; Sun et al., 2005). This overshadowing of implicit processing by explicit processing is robust and also present in insight problem solving (Schooler et al., 1993; Schooler & Melcher, 1995). A typical insight problem is addressed below.

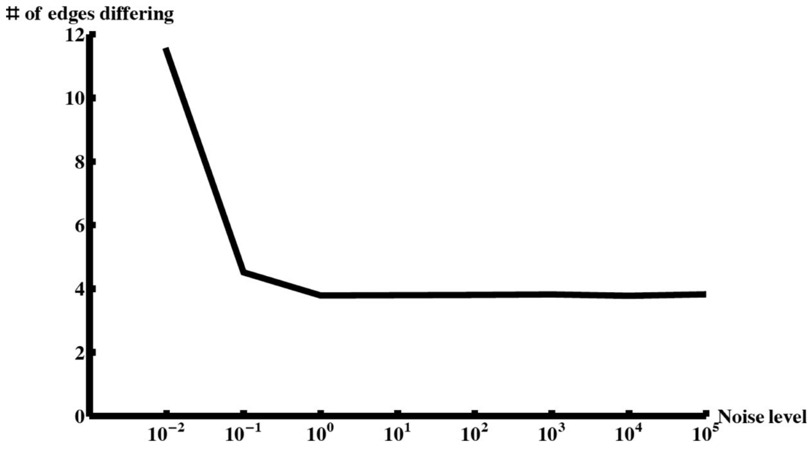


Figure 7. Number of edges differing between the solutions found by the CLARION model and the solvers' knowledge graph (see Figure 6a). The *x*-axis represents the noise level (temperature) in the Boltzmann distribution.

Experimental setting. Schooler et al. (1993) asked participants to solve the following problem:

> A dealer in antique coins got an offer to buy a beautiful bronze coin. The coin had an emperor's head on one side and the date 544 B.C. stamped on the other. The dealer examined the coin, but instead of buying it, he called the police. Why? (Schooler et al., 1993, p. 182)

Each of the 82 participants had 2 min to solve this problem. Following this initial problem-solving period, half of the participants were assigned to an unrelated task, while the remaining half were asked to verbalize their problem-solving strategies. In both cases, the interruption period lasted 90 s and was followed by another 4-min attempt to solve the initial problem. The dependant variable was the proportion of insight problems solved by the participants.

Experimental results. The results were as shown in Figure 8 (gray bars). After the participants spent the interruption period working on an unrelated task, 45.8% of the insight problems were solved. In contrast, only 35.6% of the problems were solved after the participants verbalized their problem-solving strategies. This constitutes a decrease of nearly 30% in the proportion of solved problems. According to Schooler et al. (1993), this statistically significant difference, $t(74) = 2.13$, $p < .05$, was the consequence of an overly verbal mode of problem solving, which prevented the participants from reaching an insight.

Simulation setup. As argued earlier, the participants only relied on the information present in the problem following the preparation period: the coin material, the carved pattern, the coin date, and the dealer's decision. The first three features were independent, while the last feature was a function of the first three (i.e., the antique dealer only buys high-quality items). In addition, the relative sizes of bottom-level (implicit) representations were made to reflect the task constraints. For instance, the dealer's decision was to be explained and thus should not be changed by the neural network settling (convergence) process in the bottom level. Hence, the dealer's decision was represented by more nodes in the bottom level. Accordingly, other features were represented by fewer nodes in the bottom level (in particular, the emphasis on the antique nature of the coin in the original problem suggested that the date might be problematic and thus the date was represented by even fewer nodes).

More precisely, each of the problem features (i.e., coin material, carved pattern, coin date, dealer's decision) was represented by two nodes in the left layer of the top level (see Figure 2): the date (good, bad), the material (good, bad), the carved pattern (good, bad), and the dealer's decision (buy, do not buy). In the right layer of the top level, eight abstract explanations were locally represented (using eight nodes in total).[18] In addition, the dependency between the dealer's decision and the other features was coded into the top level: When all the coin features were good, the dealer bought the coin; otherwise, the dealer did not buy the coin.

In the bottom level of the CLARION model, each concept (represented by each top-level node) was represented by a randomly generated distributed pattern. As previously indicated, the problem constraints suggested that some features were more relevant or more reliable than others, which led to the use of distributed representations of variable sizes: The coin date was represented by 30 bottom-level nodes, the dealer's decision was represented by 100 bottom-level nodes, and the remaining features (and the explanations) were represented using 50 bottom-level nodes each (all with distributed representations). Eight training stimuli (i.e., one for each explanation, thus representing all the possible cases) were generated by concatenating the random representations (the training served to reencode the top-level rules in the bottom level).

To simulate this task, a stimulus was first presented to the left layer in the top level, as well as to the bottom level, representing a good date, good material, a good carved pattern, but a refusal of the dealer to buy the coin. This stimulus did not correspond to any exemplar used to pretrain the bottom level and was inconsistent with the precoded rules in the top level (in the cases used for pretraining, coins with all the good features were bought). The stimulus was transmitted through the top-level connections and the bottom-level neural network settling (convergence) process. As in all the other simulations, the resulting bottom-level stable state was sent bottom-up (through explicitation), integrated with the result of top-level processing in the right layer of the top level, and transformed into a Boltzmann distribution (using the first noise value from Table 5). A node was stochastically chosen based on the distribution (as the model response), and the statistical mode of the distribution was used to estimate the ICL. If the ICL was higher than a predefined threshold, the response was output to effector modules. Otherwise, the response was used to initiate another round of processing (by transmitting the activation backward from the right layer to the left layer in the top level and then implicitation, as discussed before). As explained earlier, this constituted a form of abductive reasoning to infer possible causes of the selected explanation. These possible causes were used for the next iteration of processing. As in all the other simulations, each spin in the bottom level took approximately 350 ms of psychological time, so the first period of problem solving lasted a maximum of 343 spins (because human participants had 2 min).

The interruption period lasted 257 spins in the simulation (because the interruption period lasted 90 s for human participants). During this time, the participants who were assigned to an unrelated task continued to generate implicit hypotheses to explain the initial problem because implicit processing did not require much attentional resource and thus might go on during the performance of the unrelated task. The simulation runs representing these participants continuously worked on the problem (with the second noise value from Table 5). In contrast, the verbalization group did not have this incubation period because verbalization prevented the participants from working on the task and forced them into an explicit mode (Schooler et al., 1993).

Finally, both conditions had another 4-min period of implicit and explicit processing to solve the initial problem (i.e., a maximum of 686 spins). During this final problem-solving period, the noise parameter of the simulated unrelated interruption participants was reset to its initial value (to refocus the search). The dependent variable was the proportion of simulations that selected

[18] This number of explanations was determined by the Cartesian product of the first three features (recall that the dealer's decision was assumed dependent on the three other features). Each explanation represented one configuration of activation in the left layer. The correct explanation represented a good material, a good carved pattern, a bad date, and a refusal to buy the coin.

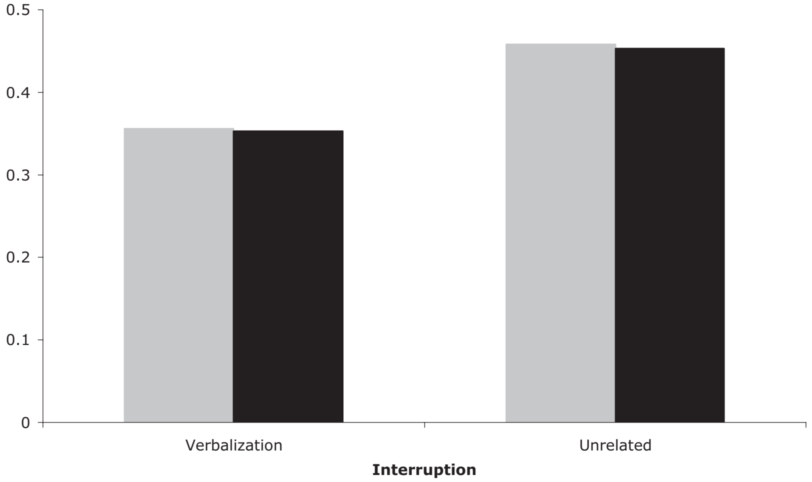


Figure 8. Proportion of correct explanations selected by the participants in Schooler, Ohlsson, and Brooks's (1993) Experiment 1 (gray bars) and by the CLARION model (black bars). The *x*-axis represents the distracting activity during the interruption period.

the correct explanation for the insight problem. The values of all the parameters were as shown in Tables 4 and 5.

Rationale and explanations.

Conceptual explanation based on EII. According to the EII theory, both explicit processing and implicit processing are initiated by the problem (Principle 2 of EII). However, insight problems are more likely to be solved by the implicit processes because rule-based processes are ineffective in solving such problems (Bowden et al., 2005). As in the earlier explanation of Durso et al.'s (1994) experiment, implicit hypotheses are generated using implicit knowledge and then verified using explicit knowledge. When the participants were interrupted to take part in an unrelated activity, hypotheses were still being generated implicitly (as in the explanation of S. M. Smith and Vela's, 1991, reminiscence data). In contrast, participants who had to verbalize their problem-solving strategies could not generate implicit hypotheses easily (because they were likely stuck in an explicit processing mode). When the participants went back to working on the problem, the verbalization group had fallen behind, so the overall probability of the verbalization group's solving the problem was lower than that of the unrelated interruption group.

Mechanistic explanation based on CLARION. In this simulation, only the bottom level of the CLARION model can generate the correct explanation because the top level can only produce stereotypical responses that are the direct consequences of its precoded explicit rules. In contrast, the bottom level involves a neural network settling (convergence) process that can be viewed as performing soft-constraint satisfaction (as discussed before; see Hertz et al., 1991; Hopfield & Tank, 1985). Because more nodes were used to represent the dealer's decision than the other features in the bottom level, the dealer's decision was considered a stronger constraint. Hence, the activation pattern (vector) was more likely to be pulled toward existing attractors that satisfied this constraint (which were likely to be the correct explanations for this task). The verbalization group did not benefit as much from this soft-constraint satisfaction process because the implicit processes were disengaged during the interruption period. This explanation clarifies the mechanistic processes underlying overshadowing in human insight problem solving (Schooler et al., 1993).

Simulation results. Five thousand simulations were run in each condition. The simulation results were as shown in Figure 8 (black bars). As predicted, simulations of the verbalization condition were less likely to select the appropriate solution to the coin problem than were the simulations of the unrelated interruption condition. Only 35.3% of the simulation runs in the verbalization condition selected the correct explanation for the problem, whereas 45.3% of the simulation runs in the unrelated problem condition selected the correct explanation (compared with 35.6% and 45.8% in the human data, respectively). This difference between the simulated verbalization and unrelated interruption conditions was reliable according to a binomial test, $B(5000, 0.353) = 2{,}265$, $p < .0001$.[19] The fit to the human data was also excellent: The difference between the human and the simulation data was smaller than 0.5%. CLARION thus successfully simulated the data related to overshadowing in insight problem solving. (More detailed statistical analysis comparing the simulated data with human data was impossible due to the unavailability of the complete human data.)

Discussion. CLARION did a good job of simulating the overshadowing effect of explicit processes on implicit processes (Schooler et al., 1993). In CLARION, overshadowing was captured by disengaging implicit processing (under proper circumstances). The data were captured because the bottom level carried out soft-constraint satisfaction (Hertz et al., 1991; Hopfield & Tank, 1985) that can weigh some constraints (e.g., the dealer's decision) more heavily than others (e.g., the coin date). The activation pattern (the activation vector) was pulled toward existing attractors that satisfied the stronger constraint, which is a natural explanation of the phenomenon.

[19] A binomial statistic was used here because each simulation produced a single output classified as correct or incorrect.

General Discussion

In the literature, creative problem solving has been used as an umbrella term that encompasses research on incubation and insight in problem solving (Bowden et al., 2005). As a result of the diversity of the field, existing theories of creative problem solving are notably fragmentary and often contradictory to each other (Lubart, 2001). The main goal of this work was to provide a new theory of creative problem solving that allows a somewhat coherent integration and unification (to some extent) of abstract theories of creative problem solving (e.g., stage decompositions; Wallas, 1926) with process theories that focus on the detailed explanations of particular stages (e.g., Dorfman et al., 1996; Mayer, 1995; Ohlsson, 1992, in press; Pols, 2002; Schilling, 2005; Schooler & Melcher, 1995; Simon, 1966; Simonton, 1995; S. M. Smith, 1995; S. M. Smith & Dodds, 1999). Furthermore, the proposed theory aimed at integrating and unifying the existing process theories by performing a reinterpretation of their assumptions, operations, and predictions.

The EII theory is an initial attempt at such a theory. EII relies mainly on five basic principles, and each of the principles is motivated by psychological as well as theoretical considerations (see the section *EII: An Integrative Theory of Creative Problem Solving*). In this article, we have shown how the EII theory may be used to capture Wallas's (1926) stage decomposition of creative problem solving and have reinterpreted six process theories of incubation and four process theories of insight. In addition, the EII theory has been able to capture and provide an explanation for many data sets that support the notions of incubation and insight in human cognition, including the four detailed in this article. The conceptual explanations provided by the EII theory are intuitively appealing.

In addition to providing high-level conceptual explanations, the formulation of EII was sufficiently precise to be implemented as a computational model based on the CLARION cognitive architecture (Sun, 2002; Sun et al., 2001, 2005). This implementation led to quantitative simulation results that closely matched human data, including the data showing the effect of incubation in a lexical decision task (Yaniv & Meyer, 1987), the data showing the effect of incubation in a free-recall task (S. M. Smith & Vela, 1991), the data showing the effect of knowledge restructuring in achieving insight (Durso et al., 1994), and the data showing the effect of overly explicit processing on insight problem solving (overshadowing; Schooler et al., 1993).

It should be mentioned that the emphasis of the above-mentioned simulations was not on fine-grained modeling of each task involved but on a broad-stroke coverage of a variety of tasks, thus making the model coarser by necessity. As a result, the simulations may have overlooked a few phenomena in these tasks that we consider to be of secondary importance (e.g., fitting the exact number of recalled words in S. M. Smith & Vela, 1991) or may have focused on only a few data sets of a task instead of all (e.g., ignoring the data related to hypermnesia in S. M. Smith & Vela, 1991). This oversight may actually be beneficial. As in any function approximation or data-fitting situations, a balance has to be stricken between fitting data faithfully and avoiding fitting noise in the data (i.e., overfitting). Coarser grained modeling may be beneficial in this regard. Finally, very importantly, a broad-scoped but coarse-grained synthesis of a range of data is essential to the goal of understanding the general principles of cognition (Newell, 1990).

This being said, the formulation of the EII theory and the development of the CLARION model to capture human data in creative problem solving have important implications beyond the scope of creative problem solving. In particular, implications for other research on creativity are examined below.

Implications for Psychological Theories of Creativity

Empirical phenomena. The notion of creativity has been defined almost as many times as there have been studies on creativity published (for reviews, see, e.g., Mayer, 1999). However, there are two common themes in most definitions: novelty and usefulness. To be deemed creative, an idea must be novel to the individual who generated it and useful according to some set of criteria.

Many different approaches have been used to study creativity (Mayer, 1999). For instance, research in the psychometric tradition, which focuses on the comparison of people who score high and low in creativity tests, has found that creative individuals work hard, prefer to make their own agenda, strive for originality, and are more flexible than uncreative individuals (Hayes, 1989). While the psychometric tradition has been around for nearly a century, most current research efforts belong to the psychological approach. This approach is mainly experimental (or quasi-experimental; Runco & Sakamoto, 1999) and focuses on the processes involved in creative thinking. Mostly, this line of research involves studying creative problem solving and the factors that affect such problem solving (Mayer, 1999). For instance, research has suggested that the information provided to participants could either increase or decrease the creativity of their solutions (Runco & Sakamoto, 1999). Also, there seems to be an optimal level of expertise for creative work: Novices rarely contribute significant creative work, but high levels of expertise often lead to the automatic production of overlearned solutions (Runco, 2004). In addition, focused attention, anxiety, and rewards tend to decrease creativity (Runco, 2004; Runco & Sakamoto, 1999). Consistent with the personality traits found to be associated with creativity by psychometric research, intrinsic motivation seems to be the most important factor leading to creativity (for full reviews and for other factors, see Runco, 2004; Runco & Sakamoto, 1999).

The EII theory can provide intuitively appealing explanations for the cognitive factors found to affect creativity. First, most creative solutions in EII are explained by implicit processing or by its integration with explicit processing (as hypothesized by Runco & Sakamoto, 1999; for a neuropsychological argument, see Dietrich, 2004). The effect of task instructions on creativity can sometimes be explained with inducing more or less explicit modes of processing. For example, we have shown that relying mostly on explicit processes in insight problem solving can lead to the overshadowing effect found in human participants (see the explanation of the simulation of Schooler et al., 1993). Second, experimental instructions can also affect the content of explicit knowledge (see the explanation of the simulation of Durso et al., 1994), and tasks that rely mainly on newly acquired explicit knowledge, which has not yet been recoded into implicit knowledge (Principle 3 of EII), can also produce an overly explicit mode of processing (and, consequently, uncreative solutions). Third, however, too little explicit knowledge would also produce low-quality responses because the context for implicit processing, which follows from mostly explicit processing in the preparation stage, could be erroneous or insufficient. Fourth, the effects of focused attention, anxiety, and rewards can all be accounted for by EII. According to Runco and Sakamoto (1999), all these factors are one and the same: Anxiety focuses the participants' attention on the stress-generating stimulus, while rewards focus their attention on whatever leads to rewards. Hence, they all amount to focused attention, which is represented by an overly explicit mode of processing in EII (Principle 1 of EII). To summarize, according to the EII theory, implicit processing plays a major role in the explanation of most creative solutions, and the cognitive factors known to decrease creativity often lead to an overly explicit mode of processing (and hence the decrease of creativity).

Theoretical considerations. The EII theory of creative problem solving is not the earliest theory that proposes an explanation for the cognitive processes involved in creativity (see, e.g., Campbell, 1960; Finke et al., 1992; Johnson-Laird, 1988; Mednick, 1962). One of the most successful previous theories of creativity is Geneplore (Finke et al., 1992). According to this theory, the creative process is composed of two distinct processing components: a generative process and an exploratory process. First, the generative processes (such as memory retrieval, association, and analogical transfer) are used to construct mental representations known as preinventive structures. When several preinventive structures have been generated, the exploration processes (such as attribute finding, conceptual interpretation, and hypothesis testing) are employed to interpret and evaluate these structures. If one or a combination of preinventive structures is sufficient to solve the problem at hand (i.e., it meets all the constraints), the creative product is output, and work on the problem is over. Otherwise, the cycle starts anew.

While this theory provides a useful high-level description of the creative process, it does not include a more fine-grained analysis of the realization of the processing components. The evolutionary theory of creativity (Campbell, 1960; Johnson-Laird, 1988), which is very similar to the evolutionary theory of insight, can be used to explain the formation of preinventive structures (the generative process) and their interpretation/evaluation (the exploratory process). Like the evolutionary theory of insight, the evolutionary theory of creativity assumes Darwin's three principles (i.e., blind variation, evaluation/selection, and retention), which roughly map onto the generation, exploration, and output of a creative solution in Geneplore. Also, as in the evolutionary theory of insight, solution generation and selection are assumed to be unconscious in the evolutionary theory of creativity (and only the selected solution reaches consciousness, as reviewed earlier). Hence, using the evolutionary theory of creativity to enhance Geneplore would lead to the prediction that creative individuals are generally unaware of the underlying process. This can account for the apparent suddenness of creative ideas.

The EII theory can capture the Geneplore theory of creativity, similar to EII's reinterpretation of the evolutionary theory of insight (Campbell, 1960). According to the EII theory, the generative process can be captured by implicit processing because it captures both memory retrieval and association (via soft-constraint satisfaction). Furthermore, explicitation and knowledge integration in EII can be used to capture the exploratory process. These processes have been used in EII to provide an intuitive explanation for attribute finding (see the simulation of Schooler et al., 1993), conceptual interpretation (see the simulation of S. M. Smith & Vela, 1991), and hypothesis testing (see the simulation of Durso et al., 1994). Finally, in EII, the integrated output is used to estimate the ICL. High ICLs indicate that the proposed solution meets most constraints, whereas low ICLs indicate important violations. In the former case, a creative solution is output, while in the latter case, information is sent top-down for another iteration, which captures the reinitialization of the generative process in Geneplore.

Implications for Computational Theories of Creativity

Early on, we argued that an important contribution of EII is its precision, which allows for a computational implementation of the theory. EII and its implementation (in CLARION) thus have important implications for computational research on creativity. Here, we present a few examples of artificial intelligence models along with a rough description of how CLARION could capture the related phenomena.

One of the acknowledged sources of creativity in artificial intelligence is analogies in problem solving (Langley & Jones, 1988). According to most theories, analogies are found by matching the structure of a new problem with the structure of previously solved problems (Gentner & Markman, 1997). Langley and Jones (1988) assumed that the search for an adequate analogy would be performed implicitly and that the identification of a useful analogy would produce an insight. While Schank and Cleary (1995) did not acknowledge the existence of implicit processes, they argued that this theory could be implemented by using explanation patterns (e.g., plans, scripts). Using these knowledge structures in unusual contexts violates top-down expectations, which constitutes creativity (Schank & Cleary, 1995). Accordingly, Schank and Cleary argued that research on creativity should focus on finding the explanation patterns that are shared in a culture, how they are normally accessed, and how they are (creatively) misapplied. For a computational model to be creative, a set of heuristics to access explanation patterns, a set of heuristics to adapt old patterns to new situations, and a set of heuristics to keep seemingly useless hypotheses alive are needed (Schank & Cleary, 1995). These ideas have been applied in commonsense adaptive planners such as PLEXUS (Alterman, 1988).

According to EII (and CLARION), it appears that finding analogies by applying previously encoded scripts and plans is a constraint satisfaction problem. Implicit knowledge in CLARION is captured by using an attractor neural network in the bottom level, which has been shown to constitute a parallel soft-constraint satisfaction algorithm (Hertz et al., 1991; Hopfield & Tank, 1985). Hence, if the explanation patterns were encoded as attractors in the bottom level of CLARION, it would naturally apply existing explanation patterns to new situations by the neural network settling (convergence) process. The resulting stable state would then be sent bottom-up (via explicitation) for integration, and if the bottom-up activation were somehow in line with the top-level activation, integration would likely lead to crossing the insight threshold. This insight would represent the application of a known explanation pattern in a novel situation (described by the input information that entered the system). This CLARION-based reinterpretation has the advantage of avoiding the use of psychologically dubious memory representations and processes (such as symbolic indexing and symbolic local memory retrieval; see Schank & Cleary, 1995). Also, soft-constraint satisfaction appears to be a natural substrate for the flexible use of explanation patterns (although the issue of representing complex structural relationships in connectionist networks is still a difficult one; see, e.g., Sun, 1992, 1994).

A second example of creativity research in artificial intelligence involves the use of connectionist networks (e.g., Boden, 2004; Duch, 2006; Martindale, 1995). According to Martindale (1995), a noisy neural network, where a random signal is added to the connection weights or inserted in the activation function, can be used to model an evolutionary theory of creativity. Also, the noise level can be used to represent the flatness of the associative hierarchy in creative individuals by making the activation more homogeneous (Mednick, 1962). Hence, more creative individuals could be modeled by using more noise, whereas less creative individuals would be modeled by using less noise. This addition of noise in neural networks is essentially similar to Duch's (2006) chaotic activation and Boden's (2004) R-unpredictability (i.e., pragmatic unpredictability).

This line of computational models of creativity is compatible with EII and its implementation in CLARION. In CLARION, both explicit knowledge and implicit knowledge are modeled using connectionist networks, and responses are stochastically chosen through a Boltzmann distribution. This distribution includes a noise parameter that has been shown to affect the probability of solving insight problems (see the simulation of Durso et al., 1994). Hence, the EII theory and its implementation in CLARION are fully compatible with the aforementioned connectionist theories of creativity.

To summarize, CLARION is able to roughly capture some computational models of creativity (at a conceptual level). In addition, theoretically, CLARION can also provide similar high-level conceptual reinterpretations for computational models of scientific discovery (e.g., Newell, Shaw, & Simon, 1962; for reviews, see Boden, 2004; Rowe & Partridge, 1993) and creative analogy (e.g., Hofstadter & Mitchell, 1994; Rowe & Partridge, 1993). However, implementation of these computational theories constitutes a major undertaking by itself in terms of time and effort. Because this is tangential to the focus of the present article, we do not delve into it here.

Future Work

While proposing a unified framework to study creative problem solving is an important step, the EII theory needs to address a more fundamental problem—the role of implicit processes in problem solving. The role of implicit processes has been suggested by several studies in both deductive reasoning (e.g., Evans, 2006; Sun, 1994) and inductive reasoning (Heit, 1998; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Sun, 1994). In both cases, the similarity between the entities involved affects rule application and rule generation through softening the hard constraints involved in rule-based processing. The simulations in this article mostly captured the effect of similarity through bottom-up knowledge integration. We have shown in this article that this knowledge integration process can be used to capture empirical human data related to creative problem solving. Future work should be devoted to capturing the effect of similarity (through bottom-level processing) on inductive and deductive reasoning, as well as the complementary effect—the effect of explicit knowledge (e.g., explicit rules in deductive or inductive reasoning) on similarity (Tenenbaum & Griffiths, 2001).

References

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19.

Alterman, R. (1988). Adaptive planning. *Cognitive Science, 12,* 393–421.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Applications of a neural model. *Psychological Review, 84,* 413–451.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96,* 703–719.

Ashcraft, M. H. (1989). *Human memory and cognition.* Glenview, IL: Scott Foresman.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit–explicit distinction. *British Journal of Psychology, 79,* 251–272.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms* (2nd ed.). London, England: Routledge.

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

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology, 22,* 72–110.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67,* 380–400.

Chartier, S., & Proulx, R. (2005). NDRAM: A nonlinear dynamic recurrent associative memory for learning bipolar and nonbipolar correlated patterns. *IEEE Transactions on Neural Networks, 16,* 1393–1400.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2,* 406–416.

Cohen, B. H. (1963). Recall of categorized word lists. *Journal of Experimental Psychology, 66,* 227–234.

Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 142–150.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 189–202.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4,* 3–23.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22,* 735–808.

Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review, 11,* 1011–1026.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006, February 17). On making the right choice: The deliberation-without-attention effect. *Science, 311,* 1005–1007.

Dodds, R. A., Smith, S. M., & Ward, T. B. (2002). The use of environmental clues during incubation. *Creativity Research Journal, 14,* 287–304.

Dodds, R. A., Ward, T. B., & Smith, S. M. (in press). A review of experimental literature on incubation in problem solving and creativity. In M. A. Runco (Ed.), *Creativity research handbook* (Vol. 3). Cresskill, NJ: Hampton Press.

Dorfman, J., Shames, V. A., & Kihlstrom, J. F. (1996). Intuition, incubation, and insight: Implicit cognition in problem solving. In G. Underwood (Ed.), *Implicit cognition* (pp. 257–296). New York, NY: Oxford University Press.

Duch, W. (2006). Computational creativity. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 435–442). Vancouver, British Columbia, Canada: IEEE Press.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58,* 1–113.

Durso, F. T., Rea, C. B., & Dayton, T. (1994). Graph-theoretic confirmation of restructuring during insight. *Psychological Science, 5,* 94–98.

Evans, J. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128,* 978–996.

Evans, J. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13,* 378–395.

Evans, J. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13,* 321–339.

Evans, J. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278.

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications.* Cambridge, MA: MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3–71.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52,* 45–56.

Hadamard, J. (1954). *The psychology of invention in the mathematical field.* New York, NY: Dover.

Hayes, J. R. (1989). Cognitive processes in creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 135–145). New York, NY: Plenum Press.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248–274). Oxford, England: Oxford University Press.

Hélie, S. (2008). Energy minimization in the nonlinear dynamic recurrent associative memory. *Neural Networks, 21,* 1041–1044.

Hélie, S., & Ashby, G. F. (2009). A neurocomputational model of automaticity and maintenance of abstract rules. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 1192–1198). Atlanta, GA: IEEE Press.

Hélie, S., Sun, R., & Xiong, L. (2008). Mixed effects of distractor tasks on incubation. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 1251–1256). Austin, TX: Cognitive Science Society.

Hélie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics, 72,* 1013–1044.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation.* Redwood City, CA: Addison-Wesley.

Hofstadter, D. R., & Mitchell, M. (1994). The copycat project: A model of mental fluidity and analogy making. In K. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory: Vol. 2. Analogical connections* (pp. 31–112). Norwood, NJ: Ablex Publishing.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA, 79,* 2554–2558.

Hopfield, J. J., & Tank, D. W. (1985). "Neural" computation of decisions in optimization problems. *Biological Cybernetics, 52,* 141–152.

Isaak, M. I., & Just, M. A. (1996). Constraints on thinking in insight and invention. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 281–325). Cambridge, MA: MIT Press.

Jausovec, N., & Bakracevic, K. (1995). What can heart rate tell us about the creative process? *Creativity Research Journal, 8,* 11–24.

Johnson, T. R., & Krems, J. F. (2001). Use of current explanations in multicausal abductive reasoning. *Cognitive Science, 25,* 903–939.

Johnson-Laird, P. N. (1988). Freedom and constraint in creativity. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 202–219). New York, NY: Cambridge University Press.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50,* 109–135.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* Cambridge, MA: MIT Press.

Kohler, W. (1925). *The mentality of apes.* New York, NY: Liveright.

Kohonen, T. (1972). Correlation matrix memories. *IEEE Transactions on Computers, 21*(C), 353–359.

Kosko, B. (1988). Bidirectional associative memories. *IEEE Transactions on Systems, Man, and Cybernetics, 18,* 49–60.

Lacroix, G. L., Giguère, G., & Larochelle, S. (2005). The origin of exemplar effects in rule-driven categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 272–288.

Langley, P., & Jones, R. (1988). A computational model of scientific insight. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 177–201). New York, NY: Cambridge University Press.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8,* 529–566.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140,* 1–55.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 883–914.

Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal, 13,* 295–308.

Ma, W., Beck, J., Latham, P., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9,* 1432–1438.

Maier, N. R. F. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology, 12,* 181–194.

Martindale, C. (1995). Creativity and connectionism. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 249–268). Cambridge, MA: MIT Press.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1083–1100.

Mayer, R. E. (1995). The search for insight: Grappling with Gestalt psychology's unanswered questions. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 3–32). Cambridge, MA: MIT Press.

Mayer, R. E. (1999). Fifty years of creativity research. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 449–460). New York, NY: Cambridge University Press.

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

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 288–294.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition, 15,* 238–246.

Miner, A. C., & Reder, L. M. (1994). A new look at feeling of knowing: Its metacognitive role in regulating question answering. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47–70). Cambridge, MA: MIT Press.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1962). The processes of creative thinking. In H. E. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary approaches to creative thinking* (pp. 63–119). New York, NY: Atherton Press.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (Vol. 1, pp. 1–44). London, England: Harvester Wheatsheaf.

Ohlsson, S. (in press). *Deep learning: How the mind overrides experience*. Cambridge, England: Cambridge University Press.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, England: Cambridge University Press.

Pols, A. J. K. (2002). *Insight problem solving.* (Unpublished doctoral dissertation). University of Utrecht, Utrecht, the Netherlands.

Rabinowitz, M., & Goldberg, N. (1995). Evaluating the structure-process hypothesis. In F. E. Weinert & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 225–242). Mahwah, NJ: Erlbaum.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219–235.

Reber, A. S., & Lewis, S. (1977). Toward a theory of implicit learning: The analysis of the form and structure of a body of tacit knowledge. *Cognition, 5,* 333–361.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press.

Rowe, J., & Partridge, D. (1993). Creativity: A survey of AI approaches. *Artificial Intelligence Review, 7,* 43–70.

Runco, M. A. (2004). Creativity. *Annual Review of Psychology, 55,* 657–687.

Runco, M. A., & Sakamoto, S. O. (1999). Experimental studies of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 62–91). New York, NY: Cambridge University Press.

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach.* Upper Saddle River, NJ: Prentice Hall.

Schank, R. C., & Cleary, C. (1995). Making machines creative. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 229–247). Cambridge, MA: MIT Press.

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

Schooler, J. W., & Melcher, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 97–133). Cambridge, MA: MIT Press.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122,* 166–183.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. H. Bowers (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 249–284). New York, NY: Academic Press.

Seger, C. (1994). Implicit learning. *Psychological Bulletin, 115,* 163–196.

Simon, H. A. (1966). Scientific discovery and the psychology of problem solving. In R. G. Colodny (Ed.), *Mind and cosmos: Essays in contemporary science and philosophy* (pp. 22–40). Pittsburgh, PA: University of Pittsburgh Press.

Simon, H. A. (1972). Theories of bounded rationality. In C. B. McGuire & R. Radner (Eds.), *Decision and organization: A volume in honor of Jacob Marschak* (pp. 161–176). Amsterdam, the Netherlands: North-Holland.

Simonton, D. K. (1995). Foresight in insight? A Darwinian answer. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 465–494). Cambridge, MA: MIT Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science, 16,* 1–40,

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108–131.

Smith, S. M. (1995). Fixation, incubation, and insight in memory and creative thinking. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 135–156). Cambridge, MA: MIT Press.

Smith, S. M., & Dodds, R. A. (1999). Incubation. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (pp. 39–43). San Diego, CA: Academic Press.

Smith, S. M., & Vela, E. (1991). Incubated reminiscence effects. *Memory & Cognition, 19,* 168–176.

Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 41*(A), 553–577.

Sun, R. (1992). On variable binding in connectionist networks. *Connection Science, 4,* 93–124.

Sun, R. (1994). *Integrating rules and connectionism for robust commonsense reasoning.* New York, NY: Wiley.

Sun, R. (2002). *Duality of the mind: A bottom-up approach toward cognition.* Mahwah, NJ: Erlbaum.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science, 25,* 203–244.

Sun, R., & Peterson, T. (1998). Autonomous learning of sequential tasks: Experiments and analyses. *IEEE Transactions on Neural Networks, 9,* 1217–1234.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review, 112,* 159–192.

Sun, R., & Zhang, X. (2004). Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research, 5,* 63–89.

Sun, R., & Zhang, X. (2006). Accounting for a variety of reasoning data within a cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence, 18,* 169–191.

Sun, R., Zhang, X., & Mathews, R. (2006). Modeling meta-cognition in a cognitive architecture. *Cognitive Systems Research, 7,* 327–338.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640.

von Newmann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies* (pp. 43–98). Princeton, NJ: Princeton University Press.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review, 8,* 168–176.

Wallas, G. (1926). *The art of thought.* New York, NY: Harcourt, Brace.

Willingham, D. B., Nissen, M., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1047–1060.

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 187–205.

Zadeh, L. (1988). Fuzzy logic. *Computer, 21,* 83–93.

(*Appendix follows*)

Appendix

Mathematical Specification of the CLARION Model

This Appendix is a technical description of the non-action-centered subsystem of CLARION in a matrix algebra form (using the implementation known as CLARION-H). An overview of the algorithm is presented in Table 3 in the main text.

The Top Level

In the top level, explicit knowledge is locally represented using binary vectors $\mathbf{x}_i = \{0, 1\}^n$, $\mathbf{y}_j = \{0, 1\}^m$, and $\|\mathbf{x}_i\| = \|\mathbf{y}_j\| = 1$ (in Figure 2 in the main text, $\mathbf{x}_i$ and $\mathbf{y}_j$ are vectors describing the activation in the left and right layers of the top level, respectively), and $\|\bullet\|$ is the Euclidean norm. These layers are connected using the matrix $\mathbf{V}$, and the associations can be exactly retrieved using the following linear transmission rule:

$$\mathbf{y}_i = (\mathbf{V}\mathbf{N}_1)\mathbf{x}_i,$$

$$\mathbf{x}_i = (\mathbf{V}^{\mathbf{T}}\mathbf{N}_2)\mathbf{y}_i, \tag{A1}$$

where $\mathbf{N_1}$ and $\mathbf{N_2}$ are defined as

$$\mathbf{N_1} = \begin{pmatrix} \|\mathbf{v}_1\|^{-2} & 0 & \ldots & 0 \\ 0 & \|\mathbf{v}_2\|^{-2} & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \|\mathbf{v}_{\mathrm{n}}\|^{-2} \end{pmatrix}$$

$$\mathbf{N}_2 = \begin{pmatrix} \|\mathbf{v}_1^{\mathbf{T}}\|^{-2} & 0 & \ldots & 0 \\ 0 & \|\mathbf{v}_2^{\mathbf{T}}\|^{-2} & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \|\mathbf{v}_{\mathrm{m}}^{\mathbf{T}}\|^{-2} \end{pmatrix}, \tag{A2}$$

where $\mathbf{x}_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$ is the activation in the left layer, $\mathbf{y}_i = \{y_{i1}, y_{i2}, \ldots, y_{im}\}$ is the activation in the right layer, and $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is the matrix containing the rules (see Equation A14).

The $\mathbf{N_1}$ and $\mathbf{N_2}$ matrices are used to normalize the activation so that the activation of a node cannot be higher than one (it is equivalent to the $k1_i$s in the main text). To do that, the number of associates of each node must be determined and used to divide the summed activation (the dot product). This number can be obtained by counting the number of nonzero elements in each row of the $\mathbf{V}$ matrix (to find the number of associates for each node in $\mathbf{x}_i$) or in each column (to find the number of associates for each node in $\mathbf{y}_i$). Because the $\mathbf{V}$ matrix is binary, this can be calculated using the squared norm. As a result, if proportion p of the associates of y_{ij} are activated in $\mathbf{x}_i$, the activation of y_{ij} is p. Note that, following Sun and Zhang (2004), it has been estimated that each application of Equation A1 takes roughly 1,500 ms of psychological time.

The Bottom Level

In CLARION, the bottom level has been implemented by an attractor neural network (J. A. Anderson, Silverstein, Ritz, & Jones, 1977; Hopfield, 1982) using NDRAM (Chartier & Proulx, 2005). Each top-level association is redundantly encoded using a random vector $\mathbf{z}_i = \mathbf{t_{1\mathit{i}}} + \mathbf{t_{2\mathit{i}}}$, where $\mathbf{t_{1\mathit{i}}} = \{-1, 1\}^s \cup \{0\}^{r-s}$ is a vector representing the first s nodes in the bottom level, which are connected to the left layer in the top level using matrix $\mathbf{E}$, while $\mathbf{t_{2\mathit{i}}} = \{0\}^s \cup \{-1, 1\}^{r-s}$ is a vector representing the remaining $r - s$ nodes in the bottom level, which are connected to the right layer in the top level using matrix $\mathbf{F}$ (see Figure 2 in the main text).

The distributed representation of the stimulus ($\mathbf{z}_i$) is transmitted in the bottom level using this nonlinear transmission rule (Chartier & Proulx, 2005):

$$\mathbf{z}_{i[t+1]} = f(\mathbf{W}\mathbf{z}_{i[t]}), \tag{A3}$$

$$\forall_{\,j},\ 1 \leq j \leq r{:}f(z_{ij[t]})$$

$$= \begin{cases} +1, \ z_{ij[t]} > 1 \\ (\delta + 1)z_{ij[t]} - \delta z_{ij[t]}^3, \ -1 \leq z_{ij[t]} \leq 1, \\ -1, \ z_{ij[t]} < -1 \end{cases} \tag{A4}$$

where $\mathbf{z}_{i[t]} = \{z_{i1[t]}, z_{i2[t]}, \ldots, z_{ir[t]}\}$ is the distributed representation after t spins in the network (there is a total of p spins per trial) and $0 < \delta < 0.5$ is the slope of the transmission function. This network is guaranteed to settle in a stable state (Hélie, 2008). Following Sun and Zhang (2004), we assumed that each spin in the bottom level takes roughly 350 ms of psychological time.

Bottom-up Transmission (Explicitation)

Once the bottom-level processing is completed, the information is sent bottom-up using the following equations. If the initial stimulus first activated the left layer in the top level, the bottom-up activation is transmitted to the right layer of the top level:

$$\mathbf{y}_{[\text{bottom-up}]} = (\mathbf{F}^{\mathbf{T}}\mathbf{N}_3)\mathbf{z}_{i[p]}, \tag{A5}$$

where $\mathbf{y}_{[\text{bottom-up}]}$ is the bottom-up signal sent to the right layer in the top level and $\mathbf{z}_{i[p]}$ is the bottom-level activation after p spins.

Otherwise, if the initial stimulus first activated the right layer in the top level, the bottom-up activation is transmitted to the left layer of the top level:

$$\mathbf{x}_{[\text{bottom-up}]} = (\mathbf{E}^{\mathbf{T}}\mathbf{N}_4)\mathbf{z}_{i[p]}, \tag{A6}$$

where $\mathbf{x}_{[\text{bottom-up}]}$ is the bottom-up signal sent to the left layer in the top level. $\mathbf{N_3}$ and $\mathbf{N_4}$ are the following square diagonal matrices:

$$\mathbf{N}_3 = \begin{pmatrix} \|\mathbf{f}_1^{\mathbf{T}}\|^{-2.2} & 0 & \ldots & 0 \\ 0 & \|\mathbf{f}_2^{\mathbf{T}}\|^{-2.2} & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \|\mathbf{f}_{\mathrm{r}}^{\mathbf{T}}\|^{-2.2} \end{pmatrix}$$

$$\mathbf{N}_4 = \begin{pmatrix} \|\mathbf{e}_1^{\mathbf{T}}\|^{-2.2} & 0 & \ldots & 0 \\ 0 & \|\mathbf{e}_2^{\mathbf{T}}\|^{-2.2} & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \|\mathbf{e}_{\mathrm{r}}^{\mathbf{T}}\|^{-2.2} \end{pmatrix}, \tag{A7}$$

where $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_r\}$ and $\mathbf{E} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_r\}$ are the matrices linking the top- and bottom-level representations (see Equations A16 and A17). Like $\mathbf{N_1}$ and $\mathbf{N_2}$, the $\mathbf{N_3}$ and $\mathbf{N_4}$ matrices are counting the number of nonzero elements in each column of matrices $\mathbf{F}$ and $\mathbf{E}$ so that if a node in the top level is associated to d nodes in the bottom level, its total activation is divided by $d^{1.1}$. The exponent, which is not present in top-level activation, was added to bottom-up activation to capture similarity-based processing (Sun, 1994).

Top-Down Transmission (Implicitation)

If the left layer in the top level is activated, the corresponding distributed representation is activated in the bottom level:

$$\mathbf{z}_i = \mathbf{E}\mathbf{x}_i = \mathbf{t}_{1i}. \tag{A8}$$

Likewise, if the right layer in the top level is activated, the corresponding distributed representation can be activated in the bottom level:

$$\mathbf{z}_i = \mathbf{F}\mathbf{y}_i = \mathbf{t}_{2i}. \tag{A9}$$

Integration

Once the bottom-up activation has reached the top level, it is integrated with the activation already present in this level using the Max function. Hence, if the initial stimulus activated the left layer of the top level, the integrated activation vector is located in the right layer of the top level:

$$\forall_j,\ 1 \le j \le m\text{: } y_{[\text{integrated}]j} = \text{Max}[y_{ij}, \lambda \times y_{[\text{bottom-up}]j} + \lambda_r \times \text{residual}_j], \tag{A10}$$

where $\mathbf{y}_{[\text{integrated}]} = \{y_{[\text{integrated}]1}, y_{[\text{integrated}]2}, \ldots, y_{[\text{integrated}]m}\}$ is the integrated activation vector, $\mathbf{y}_{[\text{bottom-up}]} = \{y_{[\text{bottom-up}]1}, y_{[\text{bottom-up}]2}, \ldots, y_{[\text{bottom-up}]m}\}$ is the bottom-up activation (Equation A5), λ is a scaling parameter that determines how implicit the processing is, residual$_j$ is the bottom-up activation resulting from the final state of the bottom level at the end of the previous processing iteration at node *j,* and λ_r is a scaling parameter. λ_r is set to 1 in the Yaniv and Meyer (1987) simulation (because of the two-task sequence) and to 0 in all the other simulations reported here (because there is no task sequence).

Likewise, if the initial stimulus activated the right layer in the top level, the integrated activation vector is located in its left layer:

$$\forall_j,\ 1 \le j \le n\text{: } x_{[\text{integrated}]j} = \text{Max}[x_{ij}, \lambda \times x_{[\text{bottom-up}]j} + \lambda_r \times \text{residual}_j], \tag{A11}$$

where $\mathbf{x}_{[\text{integrated}]} = \{x_{[\text{integrated}]1}, x_{[\text{integrated}]2}, \ldots, x_{[\text{integrated}]n}\}$ is the integrated activation vector, and $\mathbf{x}_{[\text{bottom-up}]} = \{x_{[\text{bottom-up}]1}, x_{[\text{bottom-up}]2}, \ldots, x_{[\text{bottom-up}]n}\}$ is the bottom-up activation (Equation A6).

In all cases, the integrated activation vector is further transformed into a Boltzmann probability distribution:

$$P(y_{[\text{integrated}]i}) = \frac{e^{y_{[\text{integrated}]i}/\alpha}}{\sum_j e^{y_{[\text{integrated}]j}/\alpha}}, \quad \text{or} \tag{A12}$$

$$P(x_{[\text{integrated}]i}) = \frac{e^{x_{[\text{integrated}]i}/\alpha}}{\sum_j e^{x_{[\text{integrated}]j}/\alpha}},$$

where α is the temperature (randomness parameter). Equation A12 is also called the hypothesis distribution in the main text and replaces the activation in the integrated activation vector. In other words, each $y_{[\text{integrated}]i}$ or each $x_{[\text{integrated}]i}$ represents a hypothesis, and Equation A12 represents the probability distribution of the hypotheses. The statistical mode of Equation A12 is used to estimate the internal confidence level (ICL).

Assessment

The statistical mode of the hypothesis distribution is used to estimate the ICL of the model, and a hypothesis is stochastically chosen. If the ICL is higher than a predetermined threshold (ψ), the chosen hypothesis is output to effector modules; else, the iterative process continues with the chosen hypothesis as the top-level stimulus (Equation A1) and the scaled previous bottom-up activation as the new residual activation in the bottom level (Equation A5 or Equation A6, scaled by λ).

If a response is output, the response time of the model is computed as follows:

$$\text{RT} = a - b \times \text{ICL}, \tag{A13}$$

where *a* and *b* are the maximum response time and the slope, respectively.

Learning

In the top level, explicit knowledge is represented using weight matrix $\mathbf{V} = (v_{ij})$, which was trained to encode the explicit rules using standard Hebbian learning (Kohonen, 1972):

$$\mathbf{V} = \sum_i \mathbf{y}_i \mathbf{x}_i^{\mathbf{T}}, \tag{A14}$$

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 of the restriction on the stimuli; see the Top Level section above).

In the bottom level, implicit knowledge is represented by the **W** weight matrix, which is pretrained to encode the implicit associations using a contrastive Hebbian learning rule (Chartier & Proulx, 2005):

$$\mathbf{W}_{[t]} = \zeta \mathbf{W}_{[t-1]} + \eta\left(\mathbf{z}_{i[\mathbf{0}]}\mathbf{z}_{i[\mathbf{0}]}^{\mathbf{T}} - \mathbf{z}_{i[p]}\mathbf{z}_{i[p]}^{\mathbf{T}}\right), \tag{A15}$$

(*Appendix continues*)

where $\mathbf{W}_{[t]}$ is the weight matrix at trial t, $0 < < \zeta \leq 1$ is a memory efficiency parameter, and $0 < \eta < \frac{1}{2(1 - 2\delta)r}$ is a general learning parameter (for a demonstration, see Chartier & Proulx, 2005). Note that Equation A15 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**:

$$\mathbf{E} = \sum_i \mathbf{t}_{1i}\mathbf{x}_i^{\mathbf{T}}, \tag{A16}$$

$$\mathbf{F} = \sum_j \mathbf{t}_{2j}\mathbf{y}_j^{\mathbf{T}}, \tag{A17}$$

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 (defined in the Bottom Level section above).

Received July 5, 2007
Revision received February 17, 2010
Accepted February 27, 2010 ■

Correction to Trope and Liberman (2010)

The article "Construal-Level Theory of Psychological Distance," by Yaacov Trope and Nira Liberman (*Psychological Review*, 2010, Vol. 117, No. 2, pp. 440-463), contained a misspelling in the last name of the second author in the below reference. The complete correct reference is below. The online version has been corrected.

Pronin, E., Olivola, C. Y., & Kennedy, K. A. (2008). Doing unto future selves as you would do unto others: Psychological distance and decision making. *Personality and Social Psychology Bulletin, 34,* 224–236.

DOI: 10.1037/a0020319