Using Meta-Cognition for Regulating Explanatory Quality Through a Cognitive Architecture

John Licato • Ron Sun • Selmer Bringsjord

Rensselaer Polytechnic Institute
Troy, NY, USA
{licatj,rsun,selmer}@rpi.edu

Abstract. Recent years have seen a renewed interest in cognitive systems with the ability to explain either external phenomena or their own internal reasoning processes while solving problems. Some successful models of explanation-generation have made use of structured representations, reasoned over using analogical or deductive mechanisms. But before such models can be adapted for use in real-world situations, they need to incorporate additional features associated with explanation-generation. For example, generated explanations may differ qualitatively based on the explanandum's domain; e.g., explanations rooted in physical causality to explain physical phenomena vs. folk-psychology explanations that rely on propositional attitudes (believes, knows, intends, ...). This may affect the generated explanations in both explicit and implicit ways. We tackle both the explicit and implicit effects of this cognitive feature and incorporate them into a comprehensive cognitive architecture: CLARION (especially its meta-cognitive and non-action-centered subsystems).

Keywords: Explanation, Cognitive Architecture, CLARION, Analogy, Deductive Reasoning, Meta-Cognition

1 Introduction: Features of Explanations

The importance of a cognitive system's ability to explain its results, or the actions of others, and to produce *useful* explanations, is being increasingly realized by AI researchers. But as has been known for quite some time now, there are a variety of explanations that might be considered useful. For example, if one wishes to tell some cognitive system **W** that a chicken crossed the road (which happened to require movement in an eastward direction), there are at least two different ways of presenting more or less the same thing:

E_1 Chicken C wanted to cross the road.
E_2 Muscle contractions in chicken C propelled it eastward.

These two explananda refer to the same event at different levels of abstraction by invoking different concepts. The type of explanation (or alternately, explanans) that might be deemed an appropriate response to each of these explananda differ as well. An explanation whose language features many propositional attitudes of the chicken

(e.g. *"C believes," "C knows," "C wants," etc.*) may be appropriate for explaining E_1, but may not constitute a satisfactory explanation in response to E_2. An explanation rooted in physical causality (referring to the normal properties of muscle contractions, for example) may be the other way around: it would be appropriate for E_2 but less so in response to E_1. In short, the presentation of the explanandum affects the sort of explanation that is most appropriate.

The question hinted at in the above example, of whether to root an explanation in physical causality or propositional attitudes, reflects a parallel one faced by cognitive systems: What factors are used by agents to determine which qualitative features of an explanation are appropriate? In the present paper, we explore and model one answer to this question: that the concepts used in the *presentation* of the explanandum affect the explanans in both implicit and explicit ways. We model these ways using the Meta-Cognitive Subsystem (MCS) of the CLARION cognitive architecture.

We do not hope, nor do we attempt, to resolve any questions regarding whether one type of explanation is *better* than another. Although discussion in the philosophical literature of the so-called *intentional stance* [3], the normative views of Hempel [8], and so on, are fascinating and informative, we are here only concerned with modeling the cognitive processes that lead humans to choose one style of explanation over another.

The remainder of this paper proceeds as follows. After further motivating the features the modeling of which is our target, will first discuss related previous work in modeling explanation-generation, in order to set the stage for the communication of our own, and to introduce concepts we use in this paper such as metaknowledge, metacognition, and so on (§2). In section 3, we present the cognitive architecture CLARION, and briefly discuss recent developments in its representational capabilities which make it possible for us to do the work we present herein. We close with brief demonstrations in section 4, and section 5 concludes with final remarks.

1.1 Effects of the Explanandum's Presentation

The type of feature of explanation-generation we aim to model here, which we refer to as **F** effects for convenience, are the effects that the presentation of the explanandum has on the explanation generated. If the explanandum e is a simple fact about some world, let us define the *full* explanandum E as the explanandum plus all of the contextual facts required to understand the explanandum. For example, to return to the earlier example, the position of the chicken relative to the road, the position of the road relative to the four cardinal directions, and so on, are all examples of facts comprising E. The presentation of the full explanandum $P(E)$ is a particular form of the full explanandum E. This distinction is important. E_1 and E_2 might be considered two partial presentations of the same full explanandum, but they differ in their presentations.

F effects, then, are those which the presentation of the full explanandum exhibits on the explanations generated. We can further subdivide these into $\mathbf{F_e}$ effects, and $\mathbf{F_i}$ effects; these are explicit and implicit effects, respectively. Examples of both in the psychological literature are numerous, e.g. see [23, 13].

Determinations of similarity based on simple featural overlap might be considered an implicit process, or one that operates primarily using the representations on CLARION's lower level [23], if the features in question are predominantly micro features not

immediately verbalizable. Such similarity processes are known to be used in analogical reasoning, particularly in the initial stages, which use surface similarity to select source analogs from long-term memory [9, 17, 7].

But explicit processes may play a large role in explanation as well. One way to identify explicit processes, or those that operate primarily using the representations like those on CLARION's top level, is to perform experiments on human subjects that require them to verbalize their thoughts in some way. In explanation, one example relates to the so-called "self-explanation effect," in which children who verbalize their explanations seem to be able to improve the quality of their learning, and learn more [2]. This effect also applies to adults who actively create explanations for their own use [1]. Furthermore, explaining the reasoning of the beliefs and the reasoning of others also directly enhances learning [19]; this suggests that encouraging development of theory of mind may be helpful in teaching [31].

Our basic hypothesis for the modeling of **F** effects in the present paper is that the knowledge structures used to construct explanations are selected based on parameters in the metacognitive system, which themselves may be influenced, either explicitly or implicitly, by the concepts used in the explanandum's presentation.

2 Metacognition and Explanation Generation in Cognitive Systems

In this section we provide an overview of some recent modeling of metacognition in order to give the reader a feel for the state of the art in the field, and to clarify the present paper's contribution. Explanation, and in particular the modeling of explanation using analogy, has been tackled before. Thagard (2012) divides the computational models of explanation thus far into four types: probabilistic; those based on artificial neural networks; logical; and those based on schemas or analogy . The approach described in this paper falls in between the last two of these four types, since the template-matching system which we describe in the next section allows for both rule-based deductive reasoning and a form of analogical reasoning.

Hummel and Landy [11] propose that in explanation-generation, there are at least three types of flexibilities required by the representations and underlying processes: relational flexibility, the ability to see one concept as possibly playing multiple roles; semantic flexibility, the ability to exploit partial or imperfect matches between the objects and relations comprising an explanandum and the objects and relations encoded in potentially relevant domains in long-term memory; and an ability to map to, and transfer elements from, multiple domains in long-term memory simultaneously. However, the third type of flexibility can lead to a variant of the type-token problem (i.e. ambiguity about whether two elements have the same referent) against which Gentner's one-to-one constraint [6] is often used for defense. To fix this, they have their system decide whether two units correspond within the context of a certain source analog (which effectively implements a context-sensitive variant of the one-to-one constraint), and model the system using LISA [10–12].

Friedman and Forbus [4] and Friedman [5] propose a tiered framework in which explanations sit in a layer above that of justifications, which itself sits above a con-

cept level. They demonstrate qualitative shifts in explanation-generation by exploiting metaknowledge that provides information about the structures in each tier. They do not, however, model explanation-generation for external preferences, but instead focus on the self-explanation effect. Tailoring explanations based on the beliefs of others may involve many types of reasoning, including modeling theory of mind [16], or having the ability to represent nested beliefs (e.g. "I know that the person I'm talking to believes that I believe *X*.").

Let us make two broad observations from the preceding summaries of literature. First, we see a form of metacognition in the work by Friedman and Forbus [4], in that metaknowledge about the structures in each tier is produced, manipulated, and reasoned over by the system. It is this sense of metacognition which we propose to utilize in this paper, in order to (among other things) qualitatively constrain the types of explanations which are generated by our model. The idea of qualitatively different explanations connects to our second observation, which is that the current body of work modeling explanation generation does not adequately address the cognitive processes which vary the qualitatively different types of explanations and selects the ones which are most appropriate.

Therefore, the work we propose in this paper distinguishes itself from the above approaches, on whose shoulders our work stands, in four key ways. First, our approach distinguishes between the full explanandum and its presentation. Second, we assume that this presentation affects a metacognitive system which in turn constrains the type of explanation that is generated. Third, we propose the use of specialized knowledge structures, such as *templates* and *constraint chunks* (both of which are described shortly), to allow such constraints to take the form of highly expressive knowledge structures.

Finally, we acknowledge both explicit and implicit effects of the explanandum on the explanation generation, and model both using the cognitive architecture CLARION, in such a way as to take advantage of the features it provides. We next summarize the aspects of CLARION we have used.

3 Explanation Generation in CLARION

CLARION is an integrative cognitive architecture with a several key features that we take advantage of here. These features include dual representation, a division of cognitive subsystems in a way that has previously been demonstrated to be psychologically plausible, and a flexible knowledge framework which can capture sub-conceptual, unstructured-conceptual, and structured-conceptual knowledge simultaneously [23, 25, 14]. CLARION consists of two levels: an explicit top level and an implicit bottom level. The top level typically contains knowledge structures and localist representations (which may or may not be linguistic concepts) and the bottom level often contains micro features and distributed representations. (Micro features, for our purposes here, can be defined informally as low-level constructs that correspond to properties which are not necessarily explicit, often because they are features that are not paid attention to by the agent. For example, a micro feature chunk may correspond to a certain brightness of a certain hue of the color red, or a very specific sound that can be heard precisely at three minutes in to a specific performance of Beethoven's 9th Symphony.) The top/bottom

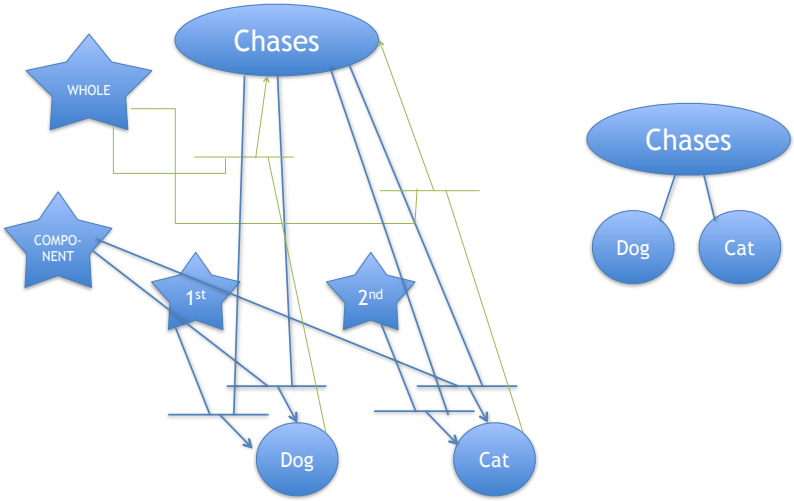

Fig. 1. A knowledge structure representing the proposition $CHASES(DOG, CAT)$. CDCs are pictured as star-shaped. On the right is the simplified version, which omits the CDCs and many of the ARs, though they are there (just not pictured).

level division is reflected in each of CLARION's subsystems: the Motivational, Action-Centered, Non-Action-Centered, and Meta-Cognitive Subsystems (MS, ACS, NACS, and MCS, respectively). A primary focus of CLARION has been psychological plausibility, and much work has been devoted to defining mechanisms within its subsystems that are tied to known psychological phenomena and processes [22, 26, 27].

The focus in the present paper is exclusively on an interaction between the NACS and MCS. In particular, recent work by Licato et al. has demonstrated how structured knowledge can be represented and reasoned over using no more than the psychologically plausible mechanisms already defined in the NACS [14, 15]; we use and expand on this method of representing structured knowledge to model explanation-generation and its metacognitive control below.

The NACS contains declarative knowledge, or general knowledge about the world that is not action-centered, which is often used for making inferences on the basis of its knowledge. The top level of the NACS contains localist chunks linked to units on the bottom level called DV pairs (Dimension-Value) pairs. The DV pairs can be linked to each other, and the chunks can also be linked to each other. However, the links between chunks are a special type of directed link called Associative Rules (ARs), which are represented pictorially using arrows. All of the links between top and bottom level units have weights that can be changed over time. This unique structure allows CLARION to define a *directed* similarity measure between two chunks [30, 21, 26]. This simple similarity measure can be used as part of larger algorithms used for analogical reason-

ing, deductive reasoning, and general behaviors defined over structured representations [14].

The MCS [24, 28] contains knowledge concerning the agent's cognitive processes and their outcomes, and also includes mechanisms that allow for active monitoring, regulation, and orchestration of the agent's cognitive processes (often toward some pragmatic goal that may be set by the MS). Like the other subsystems, the MCS is divided into a top and bottom level; however, not much work has been focused on fully exploiting both levels productively. In [24] and [28], the MCS was mostly used as the place where parameters which weighted processes in other subsystems were housed. In this paper, we propose to expand on the role of the MCS by having it hold structured knowledge analogous to that already defined in the NACS [14].

Structured knowledge in the NACS is achieved by first allowing top-level chunks to differentiate into types: object chunks, proposition chunks, template chunks, etc. These chunks are then linked using ARs and specialized chunks called Cognitively Distinguished Chunks (CDCs). For example, the proposition $Chases(Dog, Cat)$ can be represented as in Figure 1.

3.1 Templates

Analogical and deductive reasoning are carried out by defining special structures called Templates. These are essentially NACS structures that have been grouped under a single Template Chunk (TC) using a CDC defined for that purpose. In deductive reasoning, a template can specify the antecedent and consequent portions of a rule separately, so that when a structure sufficiently matches the antecedent portion, the consequent contains information on how to transfer the matched knowledge structure to create a new inference. Analogical reasoning can also be modeled by converting potential source analogs into templates and relaxing the match requirements. Matching structures to templates uses an Ant Colony Optimization algorithm inspired by [18], where the Template itself determines what types of matches are acceptable [15].

Explanation-construction proceeds as follows. We assume that we are given a knowledge base of templates. Each template is either a single structure, in which case it is to be used as a source analog for analogical matching and inference, or the template consists of an antecedent and consequent portion, in which case it is to be used as a deductive inference rule (e.g. Figure 2).

Given some knowledge structure K and template T, if a match is found from K to T (using the minimum conditions for an acceptable match specified by T itself), then a new structure K' is created from the elements of K and the instructions provided by T (these instructions are not explicitly stated by T in any way, rather they are implicit in the template's structure itself).[1]

Each template is grouped under a single Template Chunk (TC). The chunks in each template may each be linked to DV pairs in the NACS bottom level, and the template's TC is linked to a disjunction of all DV pairs linked to all non-CDC chunks in the template.

[1] For further detail, we direct interested readers to [14].

Algorithm 1 The Template Selection algorithm. This is used to filter out the template structures and select a subset of them based on how much they satisfy the constraints.

Require: Beliefs or knowledge the agent holds $B = \{B_i\}$
Require: A set of template chunks $T = \{T_i\}$
Require: A set of CCs $C = \{C_i\}$
  Define $\phi = 0.2$
  for all $T_i \in T$ **do**
    Set the activation level of T to ϕ
  end for
  for all $C_i \in C$ **do**
    if C_i is an excitatory chunk **then**
      Set C_i's activation level to $2 * \phi$ [
    else if C_i is an inhibitory chunk **then**
      Set C_i's activation level to $-2 * \phi$]
    end if
  end for
  Perform one iteration of Similarity-Based Reasoning to propagate activations
  return Active set T_A, consisting of the n $T_i \in T$ with the highest activation levels (typical value for n is between 5 and 10).

3.2 Constraint Chunks and the General Explanation-Construction Algorithm

We can now introduce a new type of chunk, which we will call a Constraint Chunk (CC). A CC is a chunk that resides on the top level of the MCS, and is used to either bias the parameters of cognitive processes based in the other (non-MCS) subsystems, or to point to the TC of a template which serves as a inviolable rule to constrain cognitive processes. The precise way in which it performs this biasing function is described shortly in the present section.

Just as the NACS chunks are linked to distributed units on the NACS bottom level, CCs are also linked to distributed units on the bottom level of the MCS. However, unless a similarity measure is defined between elements on the bottom levels of the NACS and MCS, no similarity measure will exist between chunks on their top levels. At least for this project, then, the design decision was made to allow the NACS and MCS to draw from a common pool of bottom-level distributed units, so that the same similarity measures used between two chunks of the NACS could be used from NACS to MCS chunks.[2]

Explanation generation is a simple backward-chaining process that starts with a set of knowledge structures $B = \{B_i\}$ corresponding to beliefs or knowledge that the agent holds, which are not part of the full explanandum, a set of templates $T = \{T_i\}$, a set of CCs $C = \{C_i\}$, and a full explanandum E.

The algorithm will start by selecting the relevant template structures. This requires that we have a set of CCs which are currently created manually in order to allow external users to set the qualitative features of the desired explanation, but the CCs are in such

[2] This design decision is partially justified by CLARION's view that meta-cognitive processes are intermeshed with other processes, and although the MCS is treated as a separate subsystem, it should really be viewed as closely integrated with the processes of the other subsystems [28].

a form that they can later be set autonomously by the motivational or action-centered subsystems. To carry out our demonstrations, we create two types of CCs: excitatory CCs, used to bias certain templates into being selected; and inhibitory CCs, which instead suppress and constrain the templates selected. Inhibitory and excitatory CCs can be single chunks, or they may also serve as TCs for templated structures in the NACS.

Next, the algorithm selects $T_i \in T$ subject to the constraints set by the CCs. It does this by activating all templates a fixed amount, and then activating excitatory CCs, allowing the activation to propagate using similarity-based reasoning [20, 26] (a single iteration was sufficient, though we could perform more later), and further activate certain templates. If any excitatory CCs serve as TCs for templated structures, then that structure is matched with the structures in T, and successful matches further activate those templates. Next, inhibitory CCs are activated, but rather than further activating similar templates, it lowers their activations.

As a result, we have a degree of activation for each $T_i \in T$ which reflects the constraints defined by the CCs. We collect the top n template chunks with the highest activations. This resulting set of templates is called the *active template set* (T_A). The pseudocode for the creation of T_A is shown in Algorithm 1.

The backward-chaining process can now begin. The algorithm starts by defining S as the set of facts in the full explanandum E. The templates are momentarily reversed: If some fact $s \in S$ matches the conclusion portion of a template in T_A, inference is performed on the antecedent portion of that template to produce a new set of facts, which replace s in S. If any of these newly added facts match beliefs in B, they are removed from S. This constitutes a single iteration of the backward-chaining process, which repeats until either S is empty, no more facts are found that can be added to S, or a preset time limit is reached. The remaining facts in S are then outputted as abductive assumptions.

We offer the pseudocode describing the general explanation-construction algorithm in Algorithm 2.

Algorithm 2 The General Explanation Generation algorithm.

Require: Beliefs or knowledge the agent holds $B = \{B_i\}$
Require: A set of active templates T_A obtained through Algorithm 1.
Require: Set of facts $S = \{s_i\}$ in full explanandum E.
  Let *currAssumptions* $\leftarrow S$
  while *currAssumptions* $\not\subseteq B$ or timeout not yet reached **do**
    for all $t \in T_A$ **do**
      if Consequent of t matches some $f \in$ *currAssumptions* and $f \notin B$ **then**
        Let A = The facts comprising the antecedent of t
        currAssumptions $\leftarrow$ (*currAssumptions* $- \{f\}) \bigcup A$
      end if
    end for
  end while
  return *currAssumptions* as the abductive explanation of E.

4 Demonstrations

Our very brief proof-of-concept demonstrations will serve as examples for testing the model we describe in this paper. These examples attempt to construct explanations when given a small knowledge-base, using the analogical comparison and transfer mechanisms defined in the NACS and the constraints in the MCS.

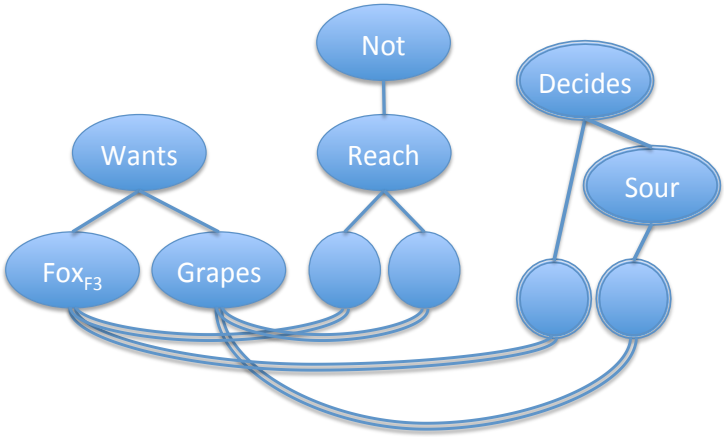

Fig. 2. A template representing the inference that a certain fox (the subscript $F3$ is meant to denote that it is a particular fox from a story with the label $F3$) wants grapes that he cannot reach, and therefore he decides that those grapes are sour. Following the notation defined in [14], the chunks with double lines are part of the consequent, and the horizontal double-lines connecting chunks are *identity links*. Assume that there is a template with chunks a, b connected by an identity link. Next, the template-matching algorithm may attempt to match two chunks a' and b' to a and b, respectively. But because of the identity link, a' and b' must have an extremely high similarity (using the measure defined in [26]).

4.1 Modeling F_i and F_e Effects

In order to clarify how we model the implicit and explicit effects of full explanandum presentation on explanation, we present a simple example demonstration that generates explanations for why a chicken crossed the road. The two full explananda, presented here in English for readability, are:

E_i The chicken decided to cross the road; the chicken was heading East.
E_p The chicken's body moved, crossing the road; the chicken was heading East.

Note that there is a very subtle difference in presentation: E_i invokes the concept of "deciding" whereas E_p does not. The algorithm will construct a new CC by simply creating a new chunk whose connected DV pairs are the disjunction of the DV pairs connected to the chunks in $P(E)$, the presentation of the full explanandum. This new CC bias the templates selected in the explanation-generation step, and thus will allow us to test $\mathbf{F_i}$ constraints. The templates provided to the system would include:

- If there is wind blowing east, and that wind is blowing on an object *o*, then *o* will move east.
- If *c* wants to achieve goal *g*, and *g* requires that action *a* happen, then *c* will decide to perform action *a*.
- If there is an object *o* that is East of *c*, and *c* likes *o*, then *c* will want to achieve the goal of moving East.
- If the chicken wants to achieve the goal of moving East, then the action of the chicken crossing the road must happen.

We now run the explanation-generation algorithm, and output the top explanation generated. When full explanandum E_i was used, the explanation generated the majority of the time (presented here again in English for readability) was:

> *Assume there is an object o that is East of the chicken. Assume the chicken likes o. The chicken will want to achieve the goal of moving East. The action of the chicken crossing the road must happen. The chicken will decide to cross the road.*

Whereas when E_p was used, the explanation was:

> *Assume that there is wind blowing east. Assume that wind is blowing on the chicken. The chicken will move east.*

Explicit effects are modeled by creating an inhibitory CC that is also the template chunk for a structure corresponding to the proposition $p =$ "The wind is blowing east." This will attempt to prevent any explanations that have p as one of its intermediate structures.

We ran the trial with E_p as the full explanandum, except this time the inhibitory CC corresponding to p is included. As expected, the explanations which require that the wind is blowing east are suppressed, and the explanation is generated as if E_i were provided instead.

5 Conclusion / Future Work

It is increasingly important that cognitive systems be able to explain and justify their conclusions and choices to the humans they will inevitably work with. For such systems, generating qualitatively different types of explanations may be essential. Using the work we have presented in this paper, such a thing can be accomplished with a few parameter changes in a meta-cognitive system. These parameters may be changed

autonomously according to contextual factors, or by normal processes rooted in CLARION's subsystems, such as the ACS, MCS, or MS. We have presented a model that can explain produce explanations at different levels of abstraction, like E_1 and E_2 in this paper's introduction.

The work here is certainly not complete; a much wider variety of explanations must eventually be addressed. For example, the ability to justify behaviors using a proof defined in a fully formalized logic is (for some domains) a glaring absence to be tackled soon, but the work in this paper can be used as a springboard for moving in that direction.

An obvious next step is to flesh out the proof-of-concept demonstration briefly described in this paper, and to examine how it performs when provided with a much larger knowledge-base. Furthermore, more sophisticated deductive reasoning is necessary to augment the part of our explanation-generation algorithm that uses inhibitory CCs corresponding to full structures. In the demonstration we presented herein, $p =$"The wind is blowing east." was used to find and suppress templates that may have led to intermediate propositions equivalent to p. But if a template leads to a logically equivalent proposition such as "The wind is not not blowing east," our algorithm would have failed.

Finally, our current system does not demonstrate learning. If the templates drawn on by the explanation generator are insufficient, then presumably a human would eventually learn a new set of templates, somehow; this is not modeled in the present work. Clearly, there is much to do.

This work was funded by grant N000141310342 from the Office of Naval Research.

References

1. Chi, M.T., Bassok, M., Lewis, M.W., Reimann, P., Glaser, R.: Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science 13(2), 145–182 (1989)
2. Chi, M.T., De Leeuw, N., Chiu, M.H., Lavancher, C.: Eliciting self-explanations improves understanding. Cognitive Science 18(3), 439–477 (1994)
3. Dennett, D.: The Intentional Stance. The MIT Press (1989)
4. Friedman, S.E., Forbus, K.: An integrated systems approach to explanation-based conceptual change. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence. Atlanta, GA (2010)
5. Friedman, S.E., Forbus, K.: Repairing incorrect knowledge with model formulation and metareasoning. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (2011)
6. Gentner, D.: Structure-Mapping: A Theoretical Framework for Analogy. Cognitive Science 7, 155–170 (1983)
7. Gentner, D., Rattermann, M.J., Forbus, K.: The Roles of Similarity in Transfer: Separating Retrievability from Inferential Soundness. Cognitive Psychology 25, 524–575 (1993)
8. Hempel, C.: Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. Free Press, New York (1965)
9. Holyoak, K.J., Koh, K.: Surface and structural similarity in analogical transfer. Memory and Cognition 15(4), 332–340 (1987)
10. Hummel, J.E., Holyoak, K.J.: A Symbolic-Connectionist Theory of Relational Inference and Generalization. Psychological Review 110, 220–264 (2003)

11. Hummel, J.E., Landy, D.H.: From analogy to explanation: Relaxing the 1:1 mapping constraint...very carefully. In: Kokinov, B., Holyoak, K.J., Gentner, D. (eds.) New Frontiers in Analogy Research: Proceedings of the Second International Conference on Analogy. Sofia, Bulgaria (2009)
12. Hummel, J.E., Licato, J., Bringsjord, S.: Analogy, explanation, and proof. Frontiers in Human Neuroscience (In Press)
13. Kahneman, D.: Thinking, Fast and Slow. Farrar, Straus and Girous (2011)
14. Licato, J., Sun, R., Bringsjord, S.: Structural Representation and Reasoning in a Hybrid Cognitive Architecture. In: Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN) (2014)
15. Licato, J., Sun, R., Bringsjord, S.: Using a Hybrid Cognitive Architecture to Model Children's Errors in an Analogy Task. In: Proceedings of CogSci 2014 (2014)
16. Pynadath, D.V., Rosenbloom, P., Marsella, S.C., Li, L.: Modeling two-player games in the sigma graphical cognitive architecture. In: Proceedings of the Sixth Conference on Artificial General Intelligence (AGI-13) (2013)
17. Ross, B.H.: Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. Journal of Experimental Psychology: Learning, Memory, and Cognition 15(3), 456–468 (1989)
18. Sammoud, O., Solnon, C., Ghédira, K.: An Ant Algorithm for the Graph Matching Problem. In: 5th European Conference on Evolutionary Computation in Combinatorial Optimization (EvoCOP 2005). Springer (2005)
19. Siegler, R.S.: How does change occur: A microgenetic study of number conservation. Cognitive Psychology 28(3), 225–273 (1995)
20. Sun, R.: Robust Reasoning: Integrating Rule-Based and Similarity-Based Reasoning. Artificial Intelligence 75(2) (1995)
21. Sun, R.: Schemas, logics, and neural assemblies. Applied Intelligence 5.2, 83–102 (1995)
22. Sun, R.: From Implicit Skills to Explicit Knowledge: A Bottom-Up Model of Skill Learning. Cognitive Science 25(2), 203–244 (2001)
23. Sun, R.: Duality of the Mind: A Bottom Up Approach Toward Cognition. Lawrence Erlbaum Associates, Mah- wah, NJ (2002)
24. Sun, R.: The motivational and metacognitive control in clarion. In: Gray, W. (ed.) Modeling Integrated Cognitive Systems. Oxford University Press, New York, New York, USA (2007)
25. Sun, R.: Autonomous generation of symbolic representations through subsymbolic activities. Philosophical Psychology (2012)
26. Sun, R., Zhang, X.: Accounting for Similarity-Based Reasoning within a Cognitive Architecture. In: Proceedings of the 26th Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates (2004)
27. Sun, R., Zhang, X.: Accounting for a Variety of Reasoning Data Within a Cognitive Architecture. Journal of Experimental and Theoretical Artificial Intelligence 18(2) (2006)
28. Sun, R., Zhang, X., Mathews, R.: Modeling meta-cognition in a cognitive architecture. Cognitive Systems Research 7, 327–338 (2006)
29. Thagard, P., Litt, A.: Models of scientific explanation. In: Thagard, P. (ed.) The Cognitive Science of Science, chap. 3. The MIT Press (2012)
30. Tversky, A.: Features of Similarity. Psychological Review 84(4), 327–352 (1977)
31. Wellman, H.M., Lagattuta, K.H.: Theory of mind for learning and teaching: The nature and role of explanation. Cognitive Development 19(4), 479–497 (2004)