First Annual Conference on Advances in Cognitive Systems Poster Collection (2012) 103-118

Interactive Analogical Retrieval

Swaroop S. Vattam SVATTAM@CC.GATECH.EDU
Ashok K. Goel GOEL@CC.GATECH.EDU
School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332 USA

Abstract

In many real-world contexts, human reasoners may not possess *a priori* knowledge of source analogues required to address target problems by analogy. Under such circumstances, a human might interact with her external information environment to obtain the needed source analogue(s). In this paper, we investigate *interactive analogical retrieval* wherein the retrieval process is mediated by the external information environment. We first report on empirical findings and then propose a theoretical model of interactive analogical retrieval.

1. Introduction

Reasoning by analogy is ubiquitous in human cognition and thus has received significant attention in cognitive systems research. Research on analogy has resulted in numerous theories and models with considerable variety of mechanisms for explaining different aspects of analogical reasoning, including retrieval, mapping, transfer, learning, etc (e.g., Carbonell, 1986; Clement, 2008; Falkenhainer, Forbus, & Gentner 1989; Forbus, Gentner, & Law 1995; Gentner, 1983, 1989; Goel & Bhatta, 2004; Hofstadter, 1995; Holyoak & Thagard, 1989; French & Hofstadter, 1991; Kolodner, 1994; Nersessian, 2009; Schank, 1983; Thagard *et. al.*, 1990). However, these theories and models typically focus on the situations in which knowledge of *source* analogues required to address the *target* problem is *a priori* encoded in the long-term memory of the cognitive agent. As a result, almost all current models of analogy emphasize an internal memory-based retrieval process. But one of the open research questions related to analogy is what happens when the analogist does not have prior knowledge of source analogues in the first place, resulting in a failure of the reminding process. One possibility is that the analogist interacts with her external information environment in order to obtain the relevant source analogues. The purpose of this article is to provide account of analogical retrieval process that explains how an external information environment mediates the process of analogical retrieval. We call this *interactive analogical retrieval (IAR).*

In developing an understanding of IAR, our focus has been on explaining how source analogues are obtained from the environment in the concrete context of *biologically inspired design* (Bar-Cohen, 2011; Benyus, 1997; French, 1994; Vincent & Mann, 2002; Yen & Weissburg, 2007). Biologically inspired design involves creative use of analogies to biological systems in order to develop solutions for complex design problems (e.g., designing a device for acquiring water in desert environments based on the analogous fog-harvesting capability of

Namibian beetle). Finding the right biological analogues is one of the critical first steps in biologically inspired design. But designers, typically situated in the domain of engineering, are usually unfamiliar with the domain of biological systems and have to significantly rely on their external socio-cultural-technological environment in order to find their sources of inspiration. Therefore, biologically inspired design provides an excellent real-world context in order to study the phenomenon of IAR.

2. The Domain

Biologically inspired design is one of the important emerging movements in engineering design. The paradigm espouses use of analogies to biological systems in generating conceptual designs for new technological innovations. This paradigm has inspired many designers in the history of design, such as Leonardo Da Vinci, the Wright brothers, etc. But it is only over the last generation or so that the paradigm has become a movement, fueled by a growing need for environmentally sustainable design on the one hand, and driven by the desire for design creativity and innovation on the other. Some examples of important innovations emerging from this paradigm include Velcro (inspired by the attachment mechanism of burr seeds), hearing aids with enhanced directional hearing (inspired by fly's auditory system), drag-reducing surfaces (inspired by shark skin), dry adhesives (inspired by attachment mechanism of gecko feet), self-cleaning surface coatings (inspired by lotus leaf), next generation wind turbine technology (inspired by the structure of flippers of humpback whales), etc.

The practice of biologically inspired design remains largely *ad hoc* with no well-established communities of practice. Accepted methodologies, best practices, or tools for systematic transfer of knowledge from biology to engineering are currently lacking. Consequently, the flow of ideas, concepts, principles, etc. from biology to engineering is mostly incidental or solution-driven. Incidental here means that the origin of the biological source of inspiration is either serendipitous or happens through ad hoc associations between people. Solution-driven implies that the problem-solving process goes from solutions to problems rather than other way around: it begins with a biological source and looks for human problems to apply this solution to.

At the same time, more and more engineers are taking an interest in biologically inspired design as this paradigm is gaining traction in the engineering community. One implication then is that engineers working on design challenges are likely to proactively look for biological sources of inspiration rather than start with a source or wait for accidental encounters with biology, which shifts the emphasis from solution-driven to problem-driven biologically inspired design. But although engineers may be experts in their respective domains, they are likely to be novices in the domain of biological systems. In order to promote biologically inspired design, the needs of designers coming from engineering have to be better understood and fulfilled.

From a cognitive standpoint, biologically inspired design is an instance of *design by analogy* (Goel, 1997). Recent cognitive research on creative design has explored the use of analogies in proposing solutions to design problems in the conceptual phase of the design process. But current conceptions of design by analogy appear to be a result of the following traditional characterization of analogical reasoning in the context of design: analogical design involves reminding and transfer of elements of a solution for one design problem to the solution for another design problem, where the selected design elements can be components, relations between components, or configurations of components and relations. That is, given a problem

P_{new} and a (partial, possibly null) solution S_{new} for P_{new}, analogical reasoning involves retrieval of a familiar problem P_{old} from memory with a solution S_{old}, and transfer of selected elements from S_{old} to S_{new}. Keeping in mind the context of biologically inspired design, we depart from this traditional characterization by assuming that P_{old} and S_{old} are *not* available in the memory and has to be obtained through interaction with the external environment.

3. The Task of Interest

Biologically inspired design is a complex activity that encompasses many tasks and sub-tasks. However, the focus of our research is limited to one of the key initial tasks of biologically inspired designing, namely *bio-inspiration seeking* - the task of finding biological systems relevant to the technology being designed. From a cognitive standpoint, the task of bio-inspiration seeking is an instance of interactive analogical retrieval, the phenomenon that we are interested in understanding. But, there are an estimated 5 to 15 million species of biological organisms. If one takes into account different levels of organization of biological systems like cellular-, organ-, and ecosystem-levels, then this estimated number of biological systems increases by an order of magnitude or more. Furthermore, novice bio-inspired designers coming from engineering are not familiar with the extent, scope, and richness of biology. They may be aware of only a small fraction of this vast space of biological systems that can be drawn upon in order to develop their design solutions. The near limitless availability of biological systems to draw upon coupled with designers' lack of knowledge of this vast domain of biological systems makes bio-inspiration seeking an intellectually challenging task.

Our studies of biologically inspired design (discussed in next section) show that it is a common practice among designers to search online in order to find their biological sources of inspiration. However, these studies also indicate that the online information environments upon which designers rely do not adequately support the task of online bio-inspiration seeking. Therefore, in spite of having online access to vast amounts of biological information, designers often struggle to find their biological sources of inspiration using the online approach. The reliance on online information environments coupled with the lack of adequate support in those environments makes an intellectually challenging task even more difficult to perform.

4. Empirical Observational Studies

We conducted two *in situ* studies of biologically inspired design practice in order to understand the phenomenon of interactive analogical retrieval as it occurs in its natural setting. These studies were conducted in Fall 2006 and Fall 2008 respectively. Details of these studies can be found in other published sources (Vattam, Helms, & Goel 2008; Helms, Vattam, & Goel 2009; Vattam, Helms, & Goel 2010).

4.1 Study Setting and Methodology

Both studies were conducted in the context of an introductory course on biologically inspired design at Georgia Institute of Technology. ME/ISyE/MSE/PTFe/BIOL 4803 is a senior-level project-based interdisciplinary course that is structured into lectures, found object exercises, and a semester-long design project. These design projects group an interdisciplinary team of 4-6 students together based on similar interests. Instructors ensure that each team has at least one designer with a biology background and a few from different engineering disciplines. Each team identifies a problem that can be addressed by a biologically inspired solution, and develops a design based on one or more biological design cases. Each team has one or more faculty as mentors who give expert advice as and when needed. Yen et al. (2010) describe the course in more detail.

As external observers (in the Fall 2006 study) and participant observers (in the Fall 2008 study), we attended almost all the classroom sessions, collected all the course materials, documented lecture content, and observed teacher-designer and designer-designer interactions in the classroom. But the focal point of our investigation was the design projects. A total of ten biologically inspired design projects were documented in these studies. We attended the design meetings of selected teams many times to observe firsthand how the design process unfolded. We took field notes, collected all the design related documentation produced by the teams, and also collected their idea journals. We analyzed the gathered data focusing on the processes and the products of the designers. In terms of the practices, we observed and documented frequently occurring problem-solving and representational activities of designers. In terms of the design products, we observed and documented the "design trajectory" – the evolution of the conceptual design over time.

4.2 Findings

We found that several key aspects characterized the online bio-inspiration seeking activity. First, it involved a search for one or more cross-domain analogies between the target technology that was the subject of design and source biological systems, mediated by several kinds of online information environments (predominantly those which gave access to scholarly biology articles like Web of Science, Google Scholar, etc.). Second, it was characterized by the application of unique strategies such as "biologizing the problem" and the use of abstractions such as functions, mechanisms, principles, constraints, etc., in order to bridge the engineering-biology divide during the search process. Third, it was characterized by a process that was not only collaborative, but consisted of three stages: pre-search stage consisted of team-level activities used to come to a shared understanding of the problem, establish information needs, negotiate division of labor, etc.; during-search stage consisted of individual information-seeking activity in order fulfill the identified information needs; and after-search stage consisted of representation-construction activities and information organization and sharing activities. Fourth, the individual information-seeking process was highly exploratory and open-ended, took up a lot of designers' time and yielded relatively small number of information resources that contained actually relevant (analogous) biological systems. Fifth, although going online to find sources of inspiration was common, perhaps inevitable, designers found this approach to be very challenging and prone to low success rate, causing frustration, suboptimal choice of sources, and source fixation problems. In particular, the following three problems surfaced prominently.

- *Low rate of encountering relevant information resources*: The duration of the online search process for seeking bio-inspiration was typically in the order of several weeks (with an average of two to three hours per week). During this search process, designers often go for long periods without finding information resources that contains an actually analogous biological system. Most of the time they encounter documents containing systems that are either superficially similar or literally similar to the target problem. In other words, the relative frequency of encountering actually useful information resources in this context is very low. This can be contrasted with our everyday online information seeking experiences where we frequently find useful information resources in response to our information needs and do so with relative ease.

- *High rate of recognition errors*: When designers encounter biology articles or documents in online environments, they have difficulty in recognizing if an article contains an analogous biological system or not. In other words, they are prone to making errors in judgment about the true utility of information resources that they encounter in their search process. They might dismiss a resource as having low utility even though it might actually be a high-utility resource (false negative), or they might select a low-utility resource and spend a lot of time and effort consuming it, only to realize later that it was not useful (false positive). Both false positives and negatives exact considerable costs on the overall information seeking process.
- *Difficulty in comprehending information resources*: Assuming that they recognize an article as having a potentially analogous biological system, designers still face the challenge of comprehending that article and developing a sufficiently rich mental model of the biological system(s) contained in it. Because 'designerly' ways of coming to know a biological system by consuming existing biological texts (especially scholarly articles) were difficult, designers often worked with incorrect or incomplete mental models which negatively affected their design solutions down the road.

5. Interactive Analogical Retrieval (IAR)

Our *in situ* observational studies pointed out the challenges of online bio-inspiration seeking. But we need to explain those challenges. In order to provide an explanation, we develop a model of *interactive analogical retrieval (IAR)*: a descriptive model of the cognitive process that underlies the task of bio-inspiration seeking. IAR combines two existing theoretical frameworks: *Analogical Retrieval by Constraint Satisfaction (ARCS)*, (Thagard *et. al.* 1990), and *Information Foraging Theory* (Pirolli 2007). ARCS is a model of analogical retrieval that explains how sources analogues are retrieved from the long-term memory, but it is silent regarding retrieval from the environment. On the other hand, Information Foraging Theory explains how people seek information in online information environments in general, but it does not take into account the peculiarities of analogical retrieval.

5.1 Theoretical foundations

ARCS is a cognitive model of analogical retrieval which posits that in order to access sources (schemas held in long-term memory) that are considered analogous to a target (a target problem/situation schema held in short-term memory), the access mechanism should simultaneously consider three constraints: *semantic similarity* (the overlap in terms of the number of similar concepts between the target and potential sources), *structural similarity* (the overlap in terms of the higher-order relationships between the target and potential sources), and *pragmatic similarity* (the overlap in terms of the pragmatic constraints or goals surrounding the target and potential sources). It is these three pressures acting simultaneously that distinguish analogical retrieval from other kinds of information access mechanisms. Although there are several other models of analogical retrieval in cognitive science and AI literature that one can draw upon, (such as MAC/FAC (Forbus, Gentner, & Law, 1995) and Case-based retrieval (Kolodner, 1994)), the choice of ARCS was driven by its ability to explain our observational data (Vattam 2012).

Information Foraging Theory (Pirolli, 2007) is a theory that was developed to explain the general information seeking behavior of people in online information environments, but makes no theoretical distinction between seeking analogies versus seeking information to satisfy more mundane information needs. According to this framework, information seeking in online environments is analogous to how animals forage for food in their natural environments. Similar to their animal counterparts, this theory posits that information seekers navigate from one *information region* to another in an information environment that is inherently patchy in nature, from one *information patch* to another within a region, and use *information scent* to guide this navigation process. Furthermore, this theory has also demonstrated that information seekers adapt their behavior to the structure of information environment in which they operate such that the system as a whole (comprising of the information seeker, the information environment, and the interactions between the two) tries to maximize the ratio of expected value of the knowledge gained to the total cost of interaction.

The IAR model developed here takes the general model of information foraging and specializes it to situations where the information seeker is seeking source analogues for target problems or situations. This is done by introducing the notion of the three pressures of analogical retrieval (from the ARCS model) into the general model of information foraging.

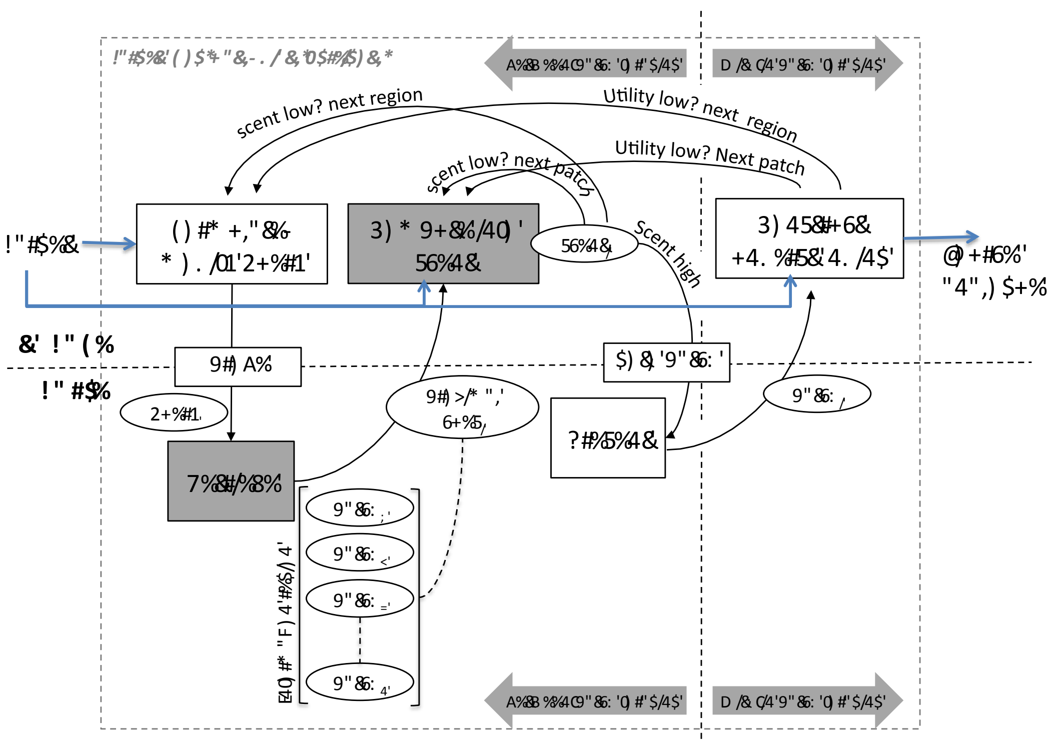


Figure 1: Interactive analogical retrieval model

5.2 A High-Level Information Processing Model of IAR

In interactive analogical retrieval, the high-level function of retrieving a source analogue is accomplished by two iterative processes that constitute the general information seeking behavior: *between-patch* and *within-patch* foraging processes (Figure 1).

5.2.1 Between-patch Foraging

Between-patch foraging (the process depicted on the left side of the vertical dotted line in Figure 1) explains the navigation process where the information seeker browses the information environment looking for high-utility information resources to consume. In the context of biologically inspired design, high-utility information patches correspond to documents containing analogous biological systems. In this process, upon probing an information environment with some information goal in mind, a forager encounters numerous information patches (e.g., Web pages, online articles, etc.) that compete for the forager's attention. These patches may or may not contain the information relevant to the forager's goals. Forager then expends some amount of time and effort navigating from one patch to another until one that can be exploited is found.

The structure of the online information environments has evolved to exhibit certain regularities in the distribution of information resources and the navigation mechanisms that lead to those resources. One such regularity is that when foragers encounter patches in the online information environment, they cannot perceive the contents of those patches all at once. Rather they perceive snippets of information representative of the distal information patches. These snippets of information are referred to as *proximal cues* or scent cues - cues that users can perceive in their local information environment to judge the utility of distal information patches and can choose to either navigate towards or away from those patches. Proximal cues are intended to represent tersely the content that a forager will encounter by choosing a particular patch. For example, proximal cues found in the Google environment consists of text on the blue hyperlinks plus the snippets of text following each link in the search results.

The perception of proximal cues associated with information patches is referred to as *information scent* of a patch. Information scent is also a measure of the perceived relevance of an information patch based on the cues. If proximal cues are perceived to have high information scent, a forager will assess that the patch associated with that scent is likely to lead to information relevant to forager's goals and vice versa. Between-patch foraging uses information scent in the context of interactive analogical retrieval as follows. Given a target problem or situation:

1. The analogist probes the environment by formulating and issuing a *query*. This query is context-dependent and draws upon the target problem.
2. In response, the environment retrieves and conveys an *information region* consisting of a set of *information patches* $\{(P_1,\{c_{11},c_{12},\ldots\}),(P_2,\{c_{21}, c_{22},\ldots\})\ldots\}$, where P_i is an information patch and c_{ij}'s are the proximal cues associated with the patch P_i.
3. Forager perceives *information scent* of the patches, an estimation of how relevant different patches are to the target, based only on the visible proximal cues: $\{(P_1, S_1), (P_2, S_2)\ldots\}$, where P_i is an information patch and S_i is the information scent that a forager associates with the patch P_i based on the match between the proximal cues and the target.
4. If the information scent of an information patch exceeds a certain threshold, it is considered relevant (high perceived utility). Therefore, the forager goes to that patch (by acting on the environment like clicking the associated hyperlink), at which point the environment

presents the information patch to the forager. This initiates the *within-patch foraging* process.

5. If the scent does not exceed the threshold, it is considered irrelevant (low perceived utility), one of two things can happen as depicted in Figure 1: (i) the agent can stay within the same information region but loop back to Step 4 for processing the next patch in the region, or (ii) the agent can abandon the current information region and loop back to Step 1 in order to look for more fruitful regions.
6. Finally, there is *uncertainty* relationship between perceived information scent and the actual relevance of distal information patch – in some cases the scent might be high but the patch might turn out to be irrelevant and vice versa.

5.2.2 Within-patch Foraging

Once the forager picks up scent of a potentially useful information patch, the forager goes to that patch and starts consuming information in that patch, in what is termed as the within-patch foraging process (the process depicted on the right side of the vertical dotted line in Figure 1). In the context of bio-inspiration seeking, this process involves comprehending the contents of an article and constructing a mental model of biological system(s) discussed in that article. In the within-patch foraging process, the agent is also simultaneously evaluating the actual utility of the patch by comparing/aligning/mapping the emerging mental model of the biological system against the target problem. In case of successful evaluation, the agent has obtained a source analogue. At any point, if this evaluation indicates a low utility of the current patch, the between-patch process is initiated. One of two things can happen when this transition occurs as depicted in Figure 1: (i) the agent can stay within the same information region but loop back to Step 4 (of between-patch foraging process above) for processing the next patch in the region, or (ii) the agent can abandon the current information region and loop back Step 1 (of between-patch foraging process above) in order to look for more fruitful regions.

5.2.3 Incorporating Pressures of Analogical Retrieval

According to the ARCS model, it is the pressures of analogical retrieval that differentiates a mundane information retrieval task from analogical retrieval task. There are two boxes (or sub-processes) in this analogical retrieval process where the pressures of analogical retrieval (semantic, structural, and pragmatic) might apply: “Retrieve” and “Compute information scent.” These are depicted as boxes shaded in gray color in the Figure 1. This is so because these two sub-processes rely on the notion of similarity. The “Retrieve” process uses some notion of similarity that is built into the search algorithm in order to access information patches. The “Compute information scent” process computes the perceived utility of an information patch by computing the similarity between the target and the proximal cues associated with the patch. Wherever the notion of similarity is applicable in this process, the pressures of analogical retrieval come into play. Out of these two sub-processes, “Compute information scent” requires an explanation; the “Retrieve” sub-process is implemented in the environment (e.g., Google search mechanism) and hence beyond the scope of this model.

While the information scent model provided in the original information foraging framework adequately explains the scent perception for non-analogy information seeking tasks, it has to be adapted in the present context such that it takes into account the three pressures of analogical retrieval. Thus, a new information scent perception model is presented here. This model explains how information scent is computed taking into consideration the pressures of analogical retrieval.

5.2.4 Pressurized Information Scent Model

Pressurized Information Scent Model (PRISM) is a model of information scent perception in the context of interactive analogical retrieval. The model assumes the presence of an organized store of associated concepts (associative semantic memory) to which representations of particular episodes are linked (episodic memory). An analogist initiates the interactive analogical retrieval process with a particular target problem or situation in mind. PRISM assumes that the cognitive system of the analogist has represented the target problem in cognitive structures called *target schema*. For our purposes, a schema is defined as an explicit, declaratively-represented mental construct representing either an encountered or expected aspect of the world (Turner, 1994).

With a target problem in mind, the analogist forages the information environment for source analogues. During the between-patch foraging process, the analogist encounters a set of information patches with associated proximal cues as shown in Figure 2. PRISM assumes that the goal of the analogist is to perceive (calculate) the information scent of each patch based on the proximal cues associated with that patch. The information scent of a patch will then allow the analogist to make judgment about the utility of going to that patch. This in turn allows the analogist to navigate the set of encountered information patches in the order of highest to lowest expected utility.

When the analogist encounters proximal cues in the environment, PRISM assumes that the cognitive system of the analogist will represent those cues in cognitive structures called *scent schemas* as depicted in Figure 2.

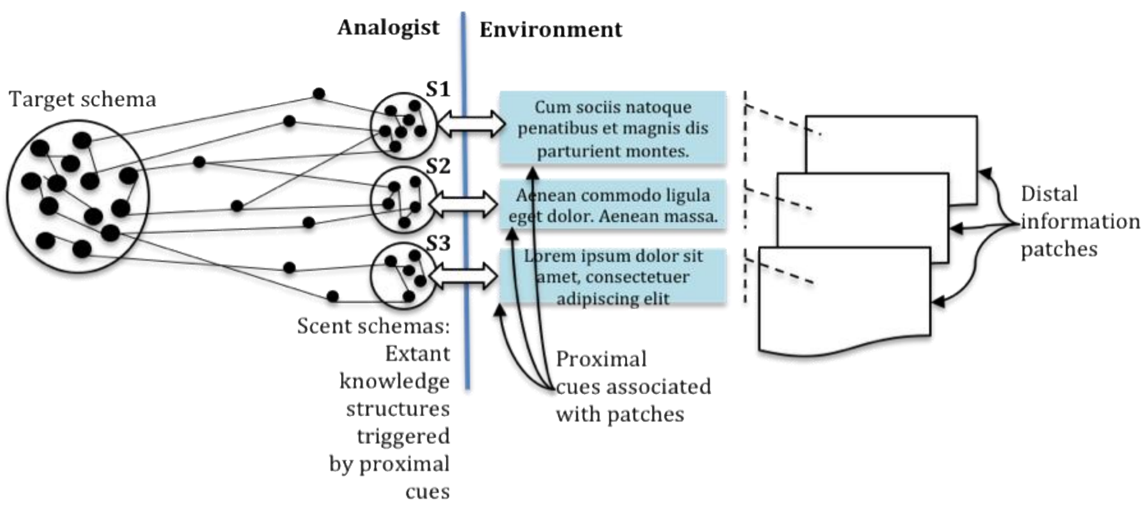


Figure 2: Scent perception in PRISM.

Given the target schema and scent schemas, PRISM computes the analogical similarity between the target and scent schemas in four stages, in a manner very similar to the original

ARCS model. In order to understand what those stages are, we have to make some minimal commitments about the knowledge representation of schemas and what those dots in Figure 2 mean. Let us assume that the conceptual structures representing these dots consist of propositions in predicate calculus. For instance, Table 1 provides a very simple illustration of a target schema (T1) consisting of two propositions (P1-1 and P1-2), and two scent schemas (S1 and S2) consisting of two propositions each (S1-1, S1-2, and S2-1 and S2-2, respectively). Let us also assume that the concepts *A* and *M* are semantically similar, and likewise concepts *B* and *N* are semantically similar; for instance:

- *A(a, b)* could represent *Regulate(kidney, potassium_ions)*;
- *M(m, n)* could represent *Control_Production(pituitary, estrogen)*;
- *B(b, a)* could represent *Is_Secreted_By(erythropoietin, kidney)*; and
- *N(n, m)* could represent *Is_Released_By(hypothalamic_hormones, pituitary)*.

Let us further assume that not all dots are equally important in the current context and that *A(a, b)* is more important than others.

Table 1. Example Target and Scent schemas (adapted from Thagard et. al., 1990, p 275).

Target Schema	Scent Schema 1	Scent Schema 2
P1	S1	S2
P1-1: A(a, b)	S1-1: M(m, n)	S2-1: M(n, m)
P1-2: B(b, a)	S1-2: N(n, m)	S2-2: R(n, m)

Assume: A and M are semantically similar; B and N are semantically similar; A(a,b) is most important in this context (dictated by the pragmatics of the context).

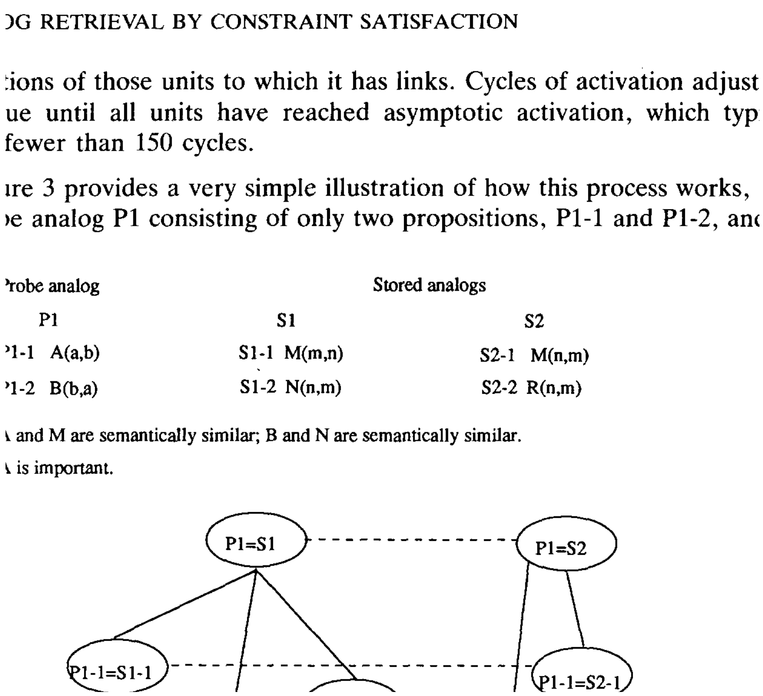
Network setup

Similar to how it is described in the original ARCS model, using information about the semantic similarity of predicates, the model creates a constraint network representing possible correspondences between concepts, predicates, relationships, and schemas. The network corresponding to knowledge in Table 1 is depicted in Figure 3. This network is a connectionist network. Units representing correspondences are created and links between units are set up to indicate correspondences between the target and scents that support each other.

The most important units are the ones that hypothesize that a scent schema is analogous to the target schema. Such units have names of the form TARGET=SCENT. (Here, "=" means "corresponds to," not identity). If the target is P1 and the scent is S1, then the unit created to represent a correspondence between them will be P1=S1. If P1-1 is a proposition in P1 that corresponds to proposition S1-1 in scent S1, then the unit P1-1=S1-1 which hypothesizes a correspondence between the propositions will have an excitatory link with the unit P1=S1. Moreover, units are created putting in correspondence the predicate and arguments of P1-1 with the predicate and arguments of S1-1, and these units receive excitatory links with the unit P1-1=S1-1.

Excitatory links are also set up from a special semantic unit to predicate-predicate units based on the degree of semantic similarity of the predicates (in Figure 3, there are excitatory links from semantic unit to (A=M) and (B=N) because they are semantically similar). Similarly, excitatory links are also set up from the special pragmatic unit to predicate-predicate units that are considered more important than others (in Figure 3, there are excitatory links from pragmatic unit to (A=M) because predicate A is assumed to be more important than others). The special semantic and pragmatic units are units whose activation level is always kept at the maximum value of 1. Hence such a unit serves to pump activation to all units that are linked to it.

Inhibitory links are constructed between units representing incompatible hypotheses, for example, between P1=S1 and P1=S2. These make utility calculation competitive, in that choosing one scent will tend to suppress choosing of an alternative. For more details about setting the network refer to the original ARCS work (Thagard *et. al.* 1990).



ɔG RETRIEVAL BY CONSTRAINT SATISFACTION

:ions of those units to which it has links. Cycles of activation adjust ue until all units have reached asymptotic activation, which typ fewer than 150 cycles.

ıre 3 provides a very simple illustration of how this process works, ›e analog P1 consisting of only two propositions, P1-1 and P1-2, anc

›robe analog	Stored analogs	
P1	S1	S2
›1-1 A(a,b)	S1-1 M(m,n)	S2-1 M(n,m)
›1-2 B(b,a)	S1-2 N(n,m)	S2-2 R(n,m)

ʌ and M are semantically similar; B and N are semantically similar.

ʌ is important.

Figure 3: Constraint satisfaction network for computing analogical similarity between Target and Scent schemas shown in Table 1. Ellipses represent possible correspondences, solid lines indicate excitatory links, and dotted lines indicate inhibitory links (from Thagard *et. al.* (1990), pp. 275).

Running the network

The constraint network constructed above is run by setting the activation of all units to a minimal initial (random) level, except for the special semantic and pragmatic units for which activation is clamped at 1. Then the activation of each unit is updated by considering the activations of those units to which it has links. Cycles of activation adjustment continue until all units have reached asymptotic activation. The equation used for updating activation suggested in ARCS model is: the activation of unit *j* on cycle *t + 1* is given by:

$$a_i(t+1) = a_i(t)(1-d) + enet_i(\max - a_i(t)) + inet_i(a_i(t) - \min)$$

Here *d* is a decay parameter, $enet_j$ is the net excitatory input, and $inet_j$ is the net inhibitory input (a negative number), with minimum activation *min* = -1 and maximum activation *max* = 1. Inputs are determined by the equations:

$$enet_j = \sum_i w_{ij} o_i(t) \quad \text{for } w_{ij} > 0\text{; and}$$

$$inet_j = \sum_i w_{ij} o_i(t) \quad \text{for } w_{ij} < 0.$$

Here, $o_i(t)$ is the output of unit *i* on cycle *t*, set by:

$$o_i(t) = \max(a_i(t), 0).$$

Updating the constraint network continues until all units have reaches asymptote, that is, a cycle is reached at which the activation change of each unit is less than a specified value, typically a low number (e.g., 0.001). For more fine-grained details about setting up the activation network, running such a network, computational complexity, etc. refer to Thagard *et. al.* (1990).

Analogical similarity and scent of an information patch

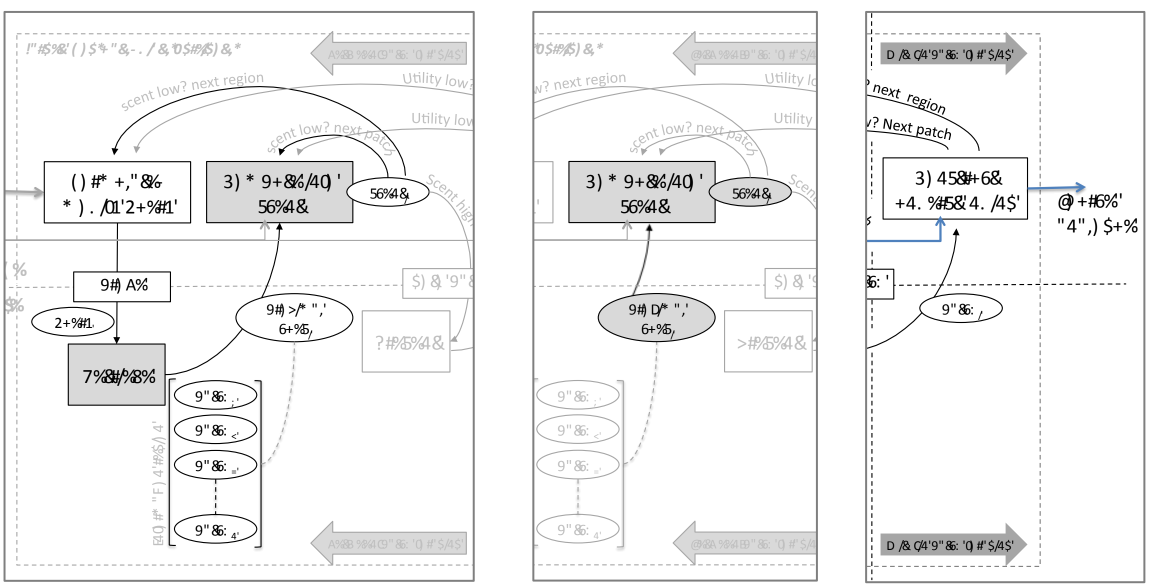
When the network settles, the analogical similarity between the target schema, *T*, and a particular scent schema, S_i, is equal to the activation value of the unit $T{=}S_i$ in the constraint network. The higher the activation accumulated by the unit $T{=}S_i$ the more similar is the scent schema, S_i, to the target, *T*. The information scent of a particular information patch, IP_i, which is associated with a set of proximal cues, $\{C_{ij}\}$, is equal to the analogical similarity between the scent schema, S_i, obtained from $\{C_{ij}\}$, and the target schema, *T*.

6. Explaining the Challenges of Online Bio-inspiration Seeking using IAR

In section 4, we discussed our empirical studies of online bio-inspiration seeking and identified some of the challenges associated with this task. Briefly, these challenges include (i) low rate of encountering relevant information resources, (ii) high rate of recognition errors, and (iii) significant difficulty in comprehending the encountered information resources. The IAR model can be used to reason forwards from deliberate changes in the information environment to its observable effects on the online bio-inspiration seeking process of designers, or backwards from observed bio-inspiration seeking effects to the factors in the information environment causing those effects. Reasoning backwards, the IAR model provides causal explanations for the three observed challenges associated with the online bio-inspiration seeking process.

The *low rate of finding relevant information* issue, where designers often go for long periods without finding a relevant information resource, can be localized to the loop highlighted in the

IAR model shown in Figure 4(a). If this loop is executed too many times, then the number of information regions foraged will be high. From the IAR model, we can infer that as the number of information regions increase, the period increases or the frequency decreases. One reason for why foragers have to loop back is because the current information region does not contain patches that produce strong scents. This can be attributed to the retrieval or the access mechanism in the information environment. In current common online information environments, keyword-based method of indexing and accessing of information resources is customarily employed, which support access to information resources based on literal similarity (word-for-word matching) while ignoring semantic-, structural- and pragmatic-similarity – the three pressures governing the process of analogical retrieval. This method does not support access to information resources based on the right kinds of things from a designer's perspective. As a result, each attempt at access can contain a large number of spurious information resources that are superficially related to the target problem. This resulting average low yield of information regions can result in an increase in the average number of information regions foraged and an increase in the average between-patch foraging time, resulting in the increase in the period between finding two useful information patches. Therefore, this issue can be traced to the current keyword-based methods of indexing and accessing information resources in online information environments, which support access to information resources based on literal similarity (word-for-word matching) while ignoring semantic-, structural- and pragmatic-similarity – the three pressures governing the process of analogical retrieval.



(a) (b) (c)

Figure 4: Challenges of online bio-inspiration seeking. Localizing the issue of (a) low rate of finding relevant information resources, (b) high rate of recognition errors, and (c) difficulty in comprehending information resources.

The recognition error issue can be localized to the information scent computation sub-process highlighted in IAR model shown in Figure 4(b). This issue is attributable to the nature of proximal cues that one encounters in customary online information environments – specifically, their lack of affordance for accurately perceiving the information scent of the resources they represent. Accurately perceiving the scent of an information resource in the context of interactive analogical retrieval requires accurately judging the deeper analogical similarity between that target problem or situation and the source information that can be gleaned from the cues. But recall that the design of proximal cues customarily contains small snippets of information. With deep background information in biology, this may be enough for domain (biology) experts to infer the missing concepts and relationships necessary to construct rich scent schemas. However, designers who are coming from engineering and who may not have the necessary background knowledge are more likely to be dealing with sparse scent schemas. According to the PRISM model presented in the last chapter, richer scent schemas afford computing the deep analogical similarity more accurately compared to sparse schemas. Therefore, the information scent computed by experts will be different from novices for the same given proximal cues. In light of this, novice designers are likely to make relevancy decisions based on superficial similarity as opposed to deep analogical similarity. This can lead to the rejection of information resources that contain structurally similar source analogues (false negatives) and/or selection of information resources that contain superficially or literally similar sources (false positives).

The issue of conceptual understanding can be localized to within-patch foraging process shown in Figure 4(c). This issue can be attributed to the fact that existing biological information resources (especially scholarly articles) are usually created by and for biologists. They often do not contain the right kind of explanations for the uninitiated. The explanations, for instance, may not be at the right level of abstraction for non-biologists. The explanations may also leave a lot of information implicit, which constitute gaps in knowledge for non-biologists, requiring them to first develop the required expertise as part of the search process. The problem of retrieval is therefore often intertwined with the problem of learning in the context of seeking bio-inspiration. Scaffolding this process of learning appropriately can therefore significantly improve the efficiency of the interactive analogical retrieval process.

7. Conclusion

Analogy is ubiquitous in human cognition. To understand how analogy functions in authentic interdisciplinary intellectual practices, cognitive theories need to take into account not only the internal cognitive mechanisms of the individual, but also the role of the environment in mediating analogical reasoning. One of the important questions related to understanding the distributed cognition of analogy is how people retrieve source analogues via their interactions with the external environment when their knowledge of source domains are severely limited. This work studies the phenomenon of interactive analogical retrieval both empirically and theoretically.

Our empirical investigation of interactive analogical retrieval was carried out in the real-world context of biologically inspired design. Although the practice of biologically inspired design involves creative use of analogies to biological systems in order to develop solutions for complex design problems, its practitioners significantly rely on online information environments in order to obtain their biological sources of inspiration. Our *in situ* studies of designers engaged in biologically inspired design helped us document the observable characteristics of interactive analogical retrieval in the context of biologically inspired design and the various challenges associated with its operation in the field. Furthermore, our findings form these studies provided a foundation to develop an empirically-grounded theoretical model of interactive analogical retrieval.

The theoretical work presented in this article develops a high-level information-processing account of interactive analogical retrieval. This account embodies central tenets of both ARCS, a conventional cognitive model of analogical retrieval (Thagard *et. al.*, 1990), and Information Foraging Theory, a theory of human information-seeking behavior in online information environments (Pirolli, 2007). It claims that seeking source analogues in online information environments involves a foraging mechanism that is driven by navigation using information scent perception, but the scent perception mechanism is itself dependent on the affordance provided by the environment for dealing with the three pressures of analogical retrieval: semantic-, structural-, and pragmatic-similarity.

Furthermore, the utility of this model was demonstrated through its ability to explain and identify the causes underlying the three observed challenges of interactive analogical retrieval in the context of bio-inspiration seeking: (i) low rate of finding relevant information resources, (ii) high rate of recognition errors, and (iii) significant difficulty in comprehending the encountered information resources.

References

Bar-Cohen, Y. (Ed.). (2011). *Biomimetics: Nature-based innovation*. Boca Ranton, FL: CRC Press.

Benyus, J. (1997). *Biomimicry: Innovation inspired by nature*. William Morrow.

Carbonell J. (1986). Derivational analogy: A theory of reconstructive problem solving and expertise acquisition. In Michalski, Carbonell & Mitchell (Eds.), *Machine Learning*, *3*, Los Altos, CA: Morgan Kaufmann.

Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. Dordrecht: Springer.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41(1)*, 1-63.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19(2)*, 141-205.

French, M. (1994) *Evolution and invention: Design in nature and engineering*, 2nd edition, Cambridge: Cambridge University Press.

French, R. M., & Hofstadter, D. R. (1991). Tabletop: A stochastic emergent model of analogy-making. *The Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*. Chicago, IL: Lawrence Erlbaum.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *23*, 155-170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. New York: Cambridge University Press.

Goel, A.K. (1997). Design, analogy, and creativity. *IEEE Expert*, *12(3)*, 62-70.

Goel, A., & Bhatta, S. (2004). Use of design patterns in analogy-based design. *Advanced Engineering Informatics*, *18(2)*, 85 –94.

Helms, M., Vattam, S., & Goel, A. K. (2009). Biologically inspired design: Products and processes, *Design Studies*, *30(5)*, 606-622.

Hofstadter, D.H. (1995). *Fluid concepts and creative analogies*. New York: Basic Books.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13(3)*, 295-355.

Kolodner, J. L. (1994). *Case-based reasoning*. San Mateo, CA: Morgan Kaufmann.

Nersessian, N.J., & Chandrasekharan, S. (2009). Hybrid analogies in conceptual innovation in science. *Cognitive Systems Research Journal*, *Special Issue: Integrating Cognitive Abilities*. *10*, 178-188.

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford: Oxford University Press.

Schank, R., (1983) *Dynamic memory: A theory of reminding in computers and people*. Cambridge: Cambridge University Press.

Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*, *46*, 259-310.

Turner, R. (1994). *Adaptive reasoning for real-world problems: A schema-based approach*. Hillsdale, New Jersey Hove, UK: Lawrence Erlbaum.

Vattam, S. (2012) *Interactive analogical retrieval: Practice, theory and technology*. Doctoral Dissertation, Georgia Institute of Technology, Atlanta, GA, USA.

Vattam, S., Helms, M., & Goel, A. (2010). A content account of creative analogies in biologically inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *24*, 467-481

Vattam, S., Helms, M., & Goel, A. (2008). Compound analogical design: Interaction between problem decomposition and analogical transfer in biologically inspired design. *Proceedings of the Third International Conference on Design Computing and Cognition* (pp. 377-396). Berlin: Springer.

Vincent, J., & Man, D. (2002). Systematic technology transfer from biology to engineering. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *360(1791)*, 159-173.

Wills, L.M. & Kolodner, J.L. (1994). Towards more creative case-based design systems. *Proceedings of the Twelfth National Conference on Artificial Intelligence* (pp. 50–55). Menlo Park, CA: AAAI Press.

Yen, J., & Weissburg, M. (2007). Perspectives on biologically inspired design: Introduction to the collected contributions. *Journal of Bioinspiration and Biomimetics, 2(4).*

Yen, J., Helms, M., Vattam, S., & Goel, A. (2010). Evaluating biological systems for their potential in engineering design. *Proceedings of the International Conference on Bionics Engineering*. Zhuhai, China.