ICED13
INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED13
19-22 AUGUST 2013, SUNGKYUNKWAN UNIVERSITY, SEOUL, KOREA

GROUNDED KNOWLEDGE REPRESENTATIONS FOR BIOLOGICALLY INSPIRED DESIGN

Michael HELMS, Ashok GOEL
Georgia Institute of Technology, United States of America

ABSTRACT

Over the last decade or so, biologically inspired design has emerged as a major paradigm in engineering design. In our work on biologically inspired design we generate grounded descriptive accounts of design, which then scaffold explanatory models of biologically inspired design processes. In this paper we use Structure-Behavior-Function (SBF) representations as a "conceptual seed" to develop a knowledge representation called SR.BID that can capture complex problem-solution relationships in biologically inspired design. The evolution of SR.BID (for Structured Representations for Biologically Inspired Design) from SBF is grounded in empirical data gathered in a classroom biologically inspired design context. SR.BID empowers us to more deeply study the breadth of processes entailed by biologically inspired design including the use of biological analogies for both solution generation and problem formulation. This paper explains in detail the process of building the content account of SR.BID, and provides a glimpse into the utility of the representation..

Keywords: design cognition, biomemetics, design theory, analogical design, problem-solution coevolution

Contact:
Michael Helms
Georgia Institute of Technology
Computer Science
Atlanta
30308
United States of America
mhelms3@gatech.edu

1 INTRODUCTION

Over the last decade or so, biologically inspired design (also known as biomimicry, biomimetics, and bionics) has emerged as a major paradigm in engineering design research (Benyus 1997; Vincent & Mann 2002; Yen & Weissburg 2007; Chakrabarti & Shu 2010; Bar-Cohen 2011; Shu et al. 2011). As one might expect, different researchers have developed different theories of design ideation and concept generation in biologically inspired design, depending on the goal, context, scope, focus and methodology of their work. For example, Biomimicry 3.8 Institute has developed an ontology of functions of biological systems that supports its design spiral model for generating design solutions (http://www.asknature.org/). Vincent and colleagues have developed a detailed ontology of biological systems to support a TRIZ-like model of biologically inspired design (Vincent et al. 2006). Chakrabarti and colleagues have developed a detailed ontology of biological and engineering systems with guidelines to support design ideation in biologically inspired design (Sartori et al. 2010). Stone, McAdams and colleagues have used the extant function-flow ontology of engineering systems to support concept generation in biologically inspired design (Nagel et al. 2010). It is important to note that while these models may represent a view of best practices of biologically inspired design, they are normative. The evidence in support of these models is frequently measured in their effectiveness at increasing a desired outcome, such as the novelty of design ideas, in the context of some design task.

Our work on biologically inspired design has taken a slightly different approach, first conducting *in situ* studies of open-ended and temporally extended design episodes, next building descriptive accounts of the empirical observations, and constructing information-processing models to explain the observations. This approach also differs from our previous work on artificial intelligence models of analogical design (Goel, Bhatta & Stroulia 1997; Goel & Bhatta 2004) in at least three fundamental ways. Firstly, the descriptive accounts and information-processing models are strongly grounded in the empirical observations from the i*n situ* studies (Helms, Vattam & Goel 2009; Vattam, Helms & Goel 2010). Secondly, the information-processing models have had to explain unexpected observations, for example, that biologically inspired design is both solution-driven and problem-driven, that it often uses compound analogies, and that it typically entails problem-solution co-evolution (Helms 2011; Helms & Goel 2012). Thirdly, the efficacy of the models is measured first in their ability at explaining and predicting design observations, and ultimately in the ability to improve design creativity, whether through improving the design process, pedagogy or technology.

As much as this approach differs from our previous work, our emphasis on grounding design processes in knowledge contents remains constant. The first research question then becomes what is a good content account of biological and technological systems that can ground the observed processes of biologically inspired design? We operationalize "good" here by defining the content account as reliable, in that if offers a repeatable method of categorizing empirical data, and comprehensive, in that it can be robustly applied to all (or nearly all) of design processes of interest. In this paper, we are specifically interested in content accounts that can capture the processes of incremental evolution or coevolution of solutions and/or problems in biologically inspired design (Helms 2011; Helms & Goel 2012)., We use the Structure-Behavior-Function (SBF) knowledge representation (Goel, Rugaber & Vattam 2009) as a "conceptual seed" to develop a knowledge representation called SR.BID that can capture the problem-solution co-evolution. The evolution of SR.BID (for Structured Representations for Biologically Inspired Design) from SBF is grounded in empirical data gathered from observing design projects in a biologically inspired design class. SR.BID empowers us to more deeply study the breadth of processes entailed by biologically inspired design including the use of biological analogies for both solution generation and problem formulation. This paper explains in detail the process of building the SR.BID representation, and provides a glimpse into its utility.

2 METHODOLOGY

Since 2006, we have observed ME/ISyE/MSE/PTFe/BIOL 4740, an interdisciplinary, project-based undergraduate class taught jointly by biology and engineering faculty at Georgia Tech, in which mostly senior-level design students work in small interdisciplinary teams of 4-5. The open-ended and temporally extended design projects involve identification of a design problem of interest to the team and conceptualization of a biologically inspired solution to the identified problem (Yen et al. 2011). Our goal in this work is to construct a content account of problem-solution coevolution that is grounded in unstructured textual data generated by the design teams. Using a consistent scheme for coding the textual data, the content account must enable comprehensive and reliable classification of observed data.

To build our content account, we use a variation on the methodology of Grounded Theory (Glaser & Strauss 1967; Strauss & Corbin 1990). In the Grounded Theory methodology, a theory about any phenomenon is derived (solely) from data. In a recent variation, the theory is derived from data but the coding scheme is seeded with a predefined ontology (Lamp & Minton 2007). We use the SBF representation to seed the coding scheme and then derive a draft schema from data during initial coding. We refine the draft schema by training and refinement, and then test the coding schema using multiple coders and standard measures of coder agreement. Finally we apply the schema to a third data set, to validate its reliability. Our objective is to demonstrate that the grounded method used to generate the SR.BID content account is reliable.

3 DATA

In this work, we use the three sets of data from the design projects in the ME/ISyE/MSE/PTFe/BIOL 4740 class summarized in Table 1. The first set of data consists of the project submissions of one design team in 2008 that focused on solar energy capture for use in homes. The project was selected as a typical example of biologically inspired design. The data consists of 4 individual design problem description assignments, a team mid-term presentation, and the team final presentation. For individual problem description assignments each designer generated a 1-2 page text description of their interpretation of the design problem the team was working on. In the case of the mid-term and final presentations, the complete text descriptions in the slides used during the presentation were used as data. In both presentations, teams were required to describe their design problem. Only the text related to the definition of the problem was used. We shall refer to this as the **2008 data set**.

The second set of data consists of an individual assignment given to the students in 2010. This assignment asked the students to provide a short 1-2 page design problem description. A total of 38 assignments were collected in the third week of the class; one was eliminated as it belonged to a member of our research laboratory who was taking the class at the time. We shall refer to this as the **Week 3 2010 data set**. The third set of data consisted of an individual assignment given to the students in 2010 and collected during the eighth week of class. This assignment consisted of generating problem descriptions between one quarter of a page and one full page in length. A total of 32 assignments were collected in the eighth week of class; again the assignment from the member of our laboratory was eliminated. We shall refer to this as the **Week 8 2010 data set**. We selected the problem description assignments as they provide rich descriptions of both problems and solutions. Other collected data, such as quantitative design assessment or descriptions of biological solutions, tends to focus exclusively on either technological or biological solutions, to the exclusion of design problems and problem-solution relationships.

Table 1: Summary of Data

Data Set	*2008 Data*	*Week 3, 2010 Data*	*Week 8, 2010 Data*
Source	Class design project	Class homework	Class homework
Number and Description	4, 1-2 page individual assignments, 1 group mid-term presentation, 1 group final presentation	37, 1-2 page individual assignments	32, up to 1 page individual assignments

4 REVIEW OF STRUCTURE-BEHAVIOR-FUNCTION MODELS

Structure-Behavior-Function (SBF) is a family of functional models of complex systems (Goel, Bhatta & Stroulia 1997; Goel & Bhatta 2004; Prabhakar & Goel 1998). The basic SBF schema consists of three nested conceptual models of the *structure*, *behavior*, and *function* of complex systems (Goel, Rugaber & Vattam 2009). Briefly, the *structure* model consists of a set of elements, such as *substances* and *components*, and *connections* among them. *Elements* may have associated *properties* and *values*, while *connections* express relationships (e.g. hinged) between *elements*.

The *behavior* model consists of *states* and *transitions* between the *states*. *States* consist of a set of *elements* and a set of *property - value* relations for the *element*. Each *transition* is annotated by *causal explanations* for the *transition.* Since one kind of *causal explanation* pertains to a *function* of a *component, behaviors* act as indices to *functions* of *components.*

The *function* model consists of a *given* or *prerequisite state*, and one or more *makes* or *resultant states*. It also specifies one or more external *stimuli*. In addition, it specifies the *behavior* that accomplishes the *function.* Thus, functions act as indices to behaviors.

Table 2 represents the basic SBF ontology we initially used for coding our design documents. Note that the SBF coding scheme suggests the grain size at which the design documents should be analyzed. In this case, we used a coding structure comprised of up to several words at a time. For instance a function typically appears in a document as a verb-noun pair, such as "clean surface" or "generate lift." A component may appear as a word such as "leg", "muscle", or "wing," whereas a property-value pair may present as a short phrase such as "positioned at 32 degrees."

Table 2 Conceptual "seeds" of the basic SBF ontology

STRUCTURE	*BEHAVIOR*	*FUNCTION*
Element	State	Prerequisite State
--Component	--Element	Resultant State
--Substance	--Property & value	Stimuli
Property	Transition	
--Value	--Causal explanation	
Connection		

5 SR.BID INITIAL CODING

During initial coding, our goal was to align the SBF seed concepts with the data and add new conceptual categories as they emerged from the data. We used a single coder to map the problem description text data in the 2008 data set to the conceptual units found in the SBF ontology. In all, 2405 words were coded, chunked into 636 concepts. Of these, 66 (10.4%) were deemed unrelated to the design project. We analyzed the remaining 570 concepts, which exposed several new categories for inclusion in SR.BID, including: perceived *benefits* and *deficiencies* of solutions; *performance criteria*; *operational environment;* and *constraints*. The findings from our initial coding provide a base for developing a more complete and reliable coding schema for problem descriptions.

6 SR.BID REFINEMENT

Following the initial coding, we used two coders to refine and validate SR.BID using the Week 3 2010 data set. The first coder was the first author of this paper (Helms) and was well versed with the content and processes involved. The second coder was a third year undergraduate biology student new to the field of biologically inspired design, and without prior background knowledge in design or cognition, SBF or SR.BID. We allocated half of the data (17 problem statements, selected at random) to training and refinement and used the remaining to draw samples for testing and validation. Training and refinement occurred in sessions of 1-3 hours, one or two days a week, for approximately fifteen weeks. The entire training set was parsed into more than 1000 individual concepts, over a wide variety of problem types.

During training and refinement we refined SR.BID by generating sub-categories, which we added based on our experience and reflection over the data set. The generation of the sub-categories was incremental, done only after instances of a new sub-category became evident from the refinement and training data set. Some categories in the initial SBF schema were eliminated due to lack of supporting data at the observed grain size.

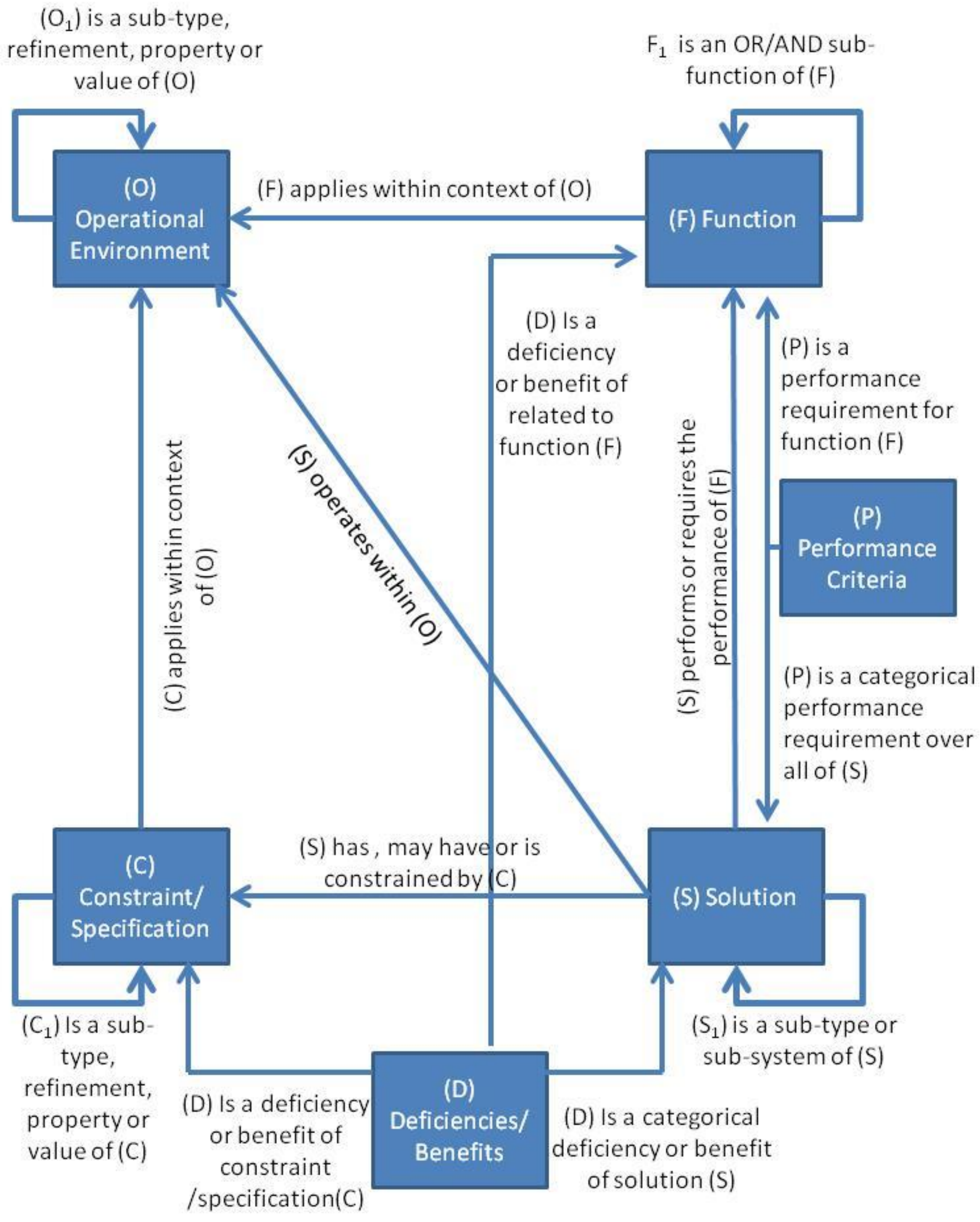

Figure 1: SR.BID concepts and their relationships

In addition to the categorization of each concept, relationships among concepts were also identified and coded. Through analysis of the 2008 data and incremental addition during the training and refinement step we derived a matrix of relationships that is illustrated in Figure 1.

After two passes on refinement and training data, a random sample of five was pulled from the remaining problems to be used for validation. Each coder independently coded each test sample. A total of 246 base concepts were identified as relevant by both coders; the coders were in agreement on 198 (80.5%) of them. The Cohen's Kappa measure of inter-coder reliability, which adjusts for chance agreement, was .778. Generally Cohen's Kappa values near 80% are deemed acceptable. Relationship concepts numbered 112; the two coders were in agreement on 84 (75.0%) of them. The Cohen's

Kappa value for relationships was .703, slightly less than is desirable. After initial comparison, the coders conducted a negotiation phase, in which they attempted to resolve coding discrepancies. As expected, post-negotiation agreement levels were significantly higher, 96.7% for concepts, and 98.0% for relationships with Cohen's Kappa values of .962 and .976 respectively.

7 SR.BID VALIDATION

To validate the conceptual soundness and potential usefulness of SR.BID, we applied it to the 2010 Week 8 data set, consisting of 31 problem statements. While previous tests confirmed that we could achieve acceptable results with a single coder, because of the difficulty of coding relationship concepts, we used a more conservative dual-coding strategy over the entire 2010 Week 8 data set. During dual-coding, each of the two coders is present during the session: while one coder takes the lead, the second coder may question coding decisions leading to discussion and negotiation until a code is agreed upon. This ensures reliability closer to the post-negotiated numbers shown in the previous test, at the expense of requiring the two coders to code all documents. Coding was conducted over 10 working sessions separated by at least 48 hours, lasting between 45 and 105 minutes each. To check for reliability, intra-coder reliability was examined after a waiting for 12 weeks, using a random sample of five problem statements. Waiting for 12 weeks ensured the coders distanced themselves from knowledge of their previous coding decisions.

Table 3 Non-weighted mean, standard deviation and frequency of the occurence of concepts in 31 problem descriptions

	Mean	*Standard Deviation*	*Frequency*
Functions	25.1%	9.1%	97%
Solutions	18.9%	11.3%	100%
Operating Environments	26.9%	15.7%	97%
Performance Criteria	5.3%	5.5%	61%
Deficiencies/Benefits	4.3%	5.4%	52%
Constraints/Specifications	5.6%	8.5%	42%

Table 4 Number and percentage of relationships between concept categories

Core Concept	*Related Concept*	*Number of Relationship Occurrences*	*Percent Relative to Number of Relationships*
Operating Environment	Operating Environment	13	3.1%
Constraint/Specification	Constraint/Specification	0	0.0%
Solution	Solution	39	9.2%
	Operating Environment	5	1.2%
	Function	109	25.8%
	Constraint/Specification	47	11.1%
Function	Function	41	9.7%
	Operating Environment	66	15.6%
Performance Criteria	Function	67	15.9%
	Solution	4	0.9%
Benefit	Function	8	1.9%
	Solution	2	0.5%
Deficiency	Function	16	3.8%
	Solution	5	1.2%

The five random problem statements consisted of 164 concepts, 17 of which were considered not relevant to the design problem. The codes were then compared between the remaining concepts. Of the remaining 147 concepts, the coders matched on 129, or 87.8%, of the category assignments. Relationship coding was likewise compared, and matched on 63 of 72, or 87.2%, relationships. As noted previously, levels above 80% are usually deemed acceptable. After coding, the 31 problem statements consisted of a total of 968 concepts, of which 112 were considered not relevant to the design problem. Of the 856 relevant concepts, 442 concepts were coded as relationships. The coders were unable to identify corresponding categories for 23 (2.7%) of the 856 concepts. Table 3 shows the non-weighted mean percentage occurrence of each category in a problem statement, the standard deviation, and the frequency with which the category occurs at all (e.g., while solutions concepts always are present, deficiencies/benefits occur in roughly half of the problem descriptions). Table 4 provides a percentage breakdown by core concept and the concept to which it is related.

8 EXAMPLE

Figure 2 illustrates the coding of the following somewhat short, coded problem description from our data set in the SR.BID schema. (Note that the problem description has been modified slightly from its original form for expository purposes.)

> ***The development of the electric car is a great thing for car owners and the environment, since tail pipe emissions can be reduced to zero, have less moving parts, and there have been huge developments in electric motors. However there is a problem in charging the battery. The time it takes to charge the battery is at least six hours. And there is limited range of the vehicle. There is a huge future for electric cars but electricity will still need to be generated to power them. The design problem is that it takes too long to charge.***

9 DISCUSSION

In building information-processing theories of design, knowledge representations typically are tied to design tasks: As we study new design tasks, we develop new knowledge representations appropriate to the new task. Thus, as we study problem-solution co-evolution in biologically inspired design, SR.BID representations are evolving out of SBF: the knowledge structures of SR.BID capture the descriptions of problem-solution relationships and the use of biological analogies for problem evolution. Given that SR.BID categorizes 97.7% of the relevant concepts in the third data set, we can say that within our design context and at the level of granularity of our analysis, the SR.BID content account appears to fully capture the design processes of interest.

We note that SR.BID allows us to capture problem descriptions more reliably than SBF. In the basic SBF ontology (Goel, Rugaber, & Vattam 2009), a system's interaction with its external environment is captured in terms of system's functions and external stimulus from the environment to the system. Prabhakar & Goel (1998) did extend the basic SBF ontology to accommodate the external and internal environments of a system but those ideas were not fully developed. Our analysis of design teams' problem descriptions indicated that in addition to functions, the designers focus on operational environment with great regularity (26.9%). This implies a rich connection between a system and its external environment. SR.BID provides the vocabulary for explicitly expressing these concepts and their relationships to other concepts in the design problem. Similarly, SR.BID provides a richer vocabulary for expressing the relationship between performance criteria, functions and solutions, against which the design of a system may be evaluated.

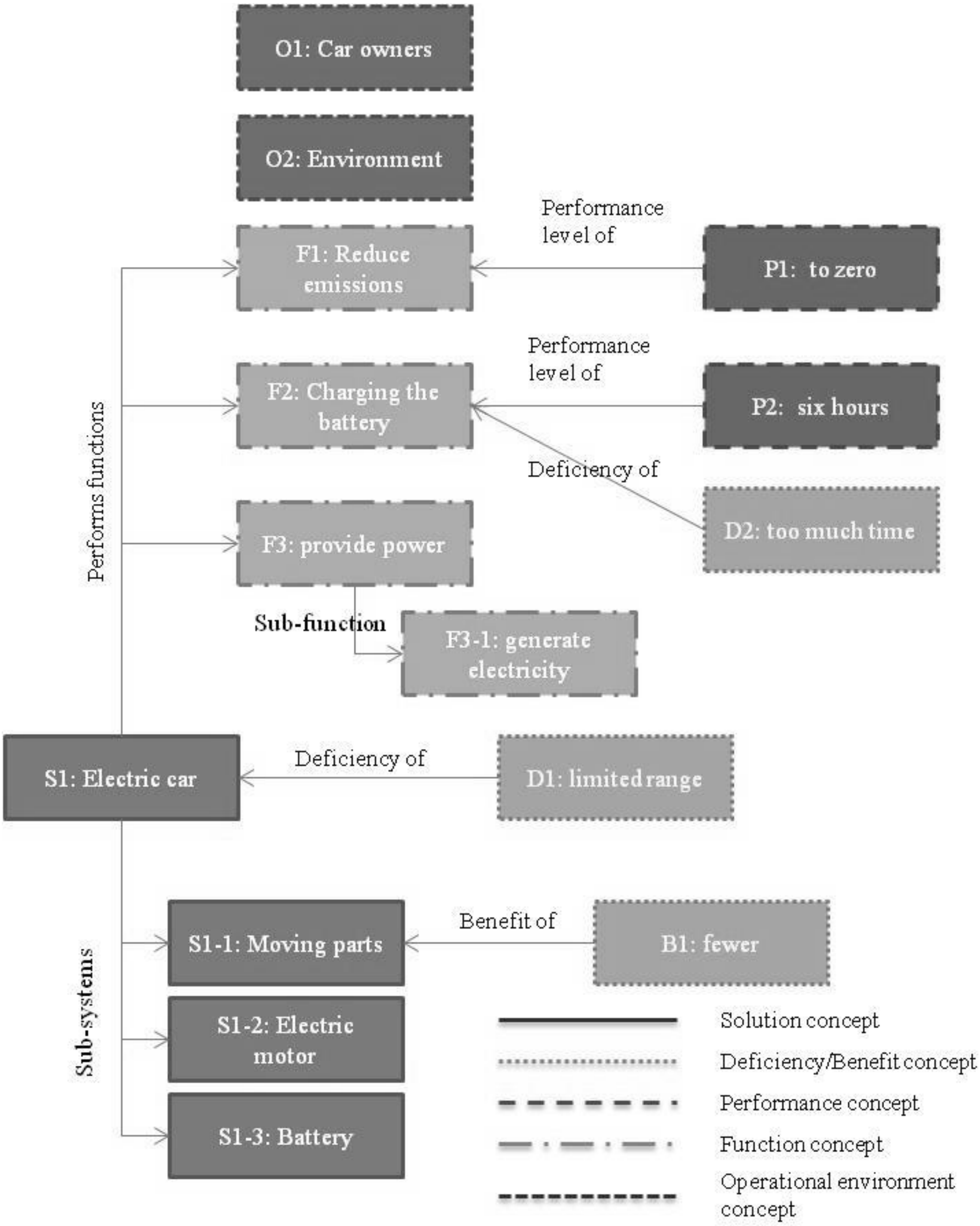

Figure 2: An illustrative example of coding using SR.BID schema

Our analysis shows that verbal descriptions of biologically inspired design in our study also always refer to biological analogies and/or other existing solutions. This may have to do with the way in which design problem formulation and re-formulation occurs. Beginning with a need, how might a designer begin to formulate the design problem for that need? One method might be to look to existing solutions that have been used to solve the need, or similar needs, in the past. The solution provides a base case, a plan, or a pattern from which the designer might abstract key concepts, such as functions, which provide the seeds necessary to begin framing the design problem. This has deep implications for biologically inspired design because it indicates that biological analogies may serve to help frame problems as well as solve them (Helms & Goel 2012).

10 USES OF SR.BID

Currently we are using SR.BID in four ways. Firstly, we are using it as a coding scheme to analyze the relationship between biological analogies and problem formulation. In a manner similar to that of this paper, we are studying the influence of biological analogies on problem formulations over time. In this way we may provide additional validation for the SR.BID schema. Secondly, we are using SR.BID as an assistive technology to help students formulate design problems, a task that has proven exceptionally difficult for students (Yen et al, 2011). In what we call the four-box method, students define their problems in terms of: Operational Environment, Function, Constraints/Specifications, and Performance Criteria. Thirdly, students currently lack a systematic method for evaluating analogies in design. Evaluation is ad-hoc, and suffers from confirmation bias effects. We use the same four-box method to evaluate analogies in biologically inspired design. Students compare their four-box problem representation against a four-box representation constructed for their analogue system, and then use this to frame a discussion of how their analogy is similar and dissimilar. Finally, where most search engines in biologically inspired design focus on indexing by function, we are using SR.BID to structure a database of design problems and biological systems to help facilitate search across the breadth of concepts found in a problem description. By improving the effectiveness of designers for specific learning goals or design tasks, the last three efforts provide validation of another kind for SR.BID.

While we find SR.BID a promising start on an ontology for design problems, it is important to acknowledge that the relationships and concepts identified thus far are grounded in, and limited by, the rich text data provided by designers in the context of a class in biologically inspired design. These static representations miss the intermediate, tacit, non-textual and process knowledge used to generate the problem formulation. Moreover, other tacit concepts and relationships may also exist, and we suspect they do, but they were not made transparent in our analysis. We expect that inclusion of multi-modal representations may be useful for extending the SR.BID schema.

11 CONCLUSION

Many current theories of biologically inspired design are normative and largely focused on enhancing designers' capability with respect to a cognitive task such as search, ideation, transfer, etc. We posit that it is also useful to develop descriptive and explanatory theories of biologically inspired design. Rigorous studies require knowledge constructs that can help analyze the fundamental processes in biologically inspired design such as problem-solution co-evolution. Thus, in this work, we use the SBF ontology as a seed for developing the SR.BID schema for representing problem descriptions in biologically inspired design. The conceptualization of SR.BID's knowledge constructs was data driven, and grounded in the textual descriptions of designing generated by the designers in our study. As measured by standard tests of coder reliability, the SR.BID knowledge constructs seem to provide comprehensive and reliable encoding of the verbal descriptions of interdisciplinary design teams engaged in biologically inspired design.

The SR.BID ontology allows us to express rich problem and solution descriptions that design teams construct in collaborative, temporally extended, open-ended biologically inspired design. It also enables us to capture the relationships between the problem and the solutions descriptions as well as systematically trace the influence of the problem on the solution and vice versa. In addition, it affords explanations for some of the fundamental processes of biologically inspired design including analogical problem inception and evolution.

ACKNOWLEDGEMENTS

We are grateful to the instructors and students of the ME/ISyE/MSE/PTFe/BIOL 4740 class from 2006 through 2012, especially Professor Jeannette Yen, the main instructor and coordinator of the class, and for the contributions of Marika Shahid, Swaroop Vattam, and Bryan Wiltgen. We thank the US National Science Foundation for its support of this research through an NSF CreativeIT Grant (#0855916) entitled "Computational Tools for Enhancing Creativity in Biologically Inspired Engineering Design" and NSF Division of Undergraduate Education Grant (#1022778) entitled "Biologically !nspired Design: A Novel Interdisciplinary Biology-Engineering Curriculum."

REFERENCES

Bar-Cohen, Y. (Editor, 2011) Biomimetics: Nature-based Innovation. CRC Press.

Benyus, J. (1997) Biomimicry: innovation inspired by nature. William Morrow.
Chakrabarti, A. and Shu, L. (2010) Biologically Inspired Design. AIEDAM, 24:453-454.
Glaser, B. and Strauss, A. (1967) The discovery of grounded theory: strategies for qualitative research. Chicago: Aldine.
Goel, A., Bhatta, S. and Stroulia, E. (1997) Kritik: An Early Case-Based Design System. In Issues and Applications of Case-Based Reasoning in Design, M. Maher and P. Pu (editors), Mahwah, NJ: Erlbaum, pages 87-132, 1997.
Goel A.K. and Bhatta S. (2004) Design patterns: a unit of analogical transfer in creative design. Advanced Engineering Informatics 18(2):85-94
Goel, A., Rugaber, S. and Vattam, S. (2009) Structure, behavior & function of complex systems: the SBF modeling language. AIEDAM, 23:23-35.
Helms, M. (2011) Solution based problem evolution and problem inception in biologically inspired design. Technical report, GVU, Georgia Institute of Technology. GIT-GVU-11-10.
Helms, M. and Goel, A. (2012) Analogical problem evolution in biologically inspired design. In Proc. Fifth International Conference on Design Computing and Cognition, College Station, Texas, July 2012, Springer.
Helms, M., Vattam, S. and Goel, A. (2009) Biologically inspired design: product and processes. Design Studies, 30(5):606-622.
Lamp, J. and Milton, S. (2007) Grounded theory as foundations for methods in applied ontology, Proceedings of the 4th International Conference on Qualitative Research in IT & IT in Qualitative Research (QualIT), Victoria University of Wellington.
Nagel J., Nagel, R., Stone, R. and McAdams, D. (2010) Function-based, biologically inspired concept generation. AIEDAM, 24:521-535.
Prabhakar, S. and Goel, A. (1998) Functional modeling for enabling adaptive design of devices for new environments, Artificial Intelligence in Engineering, 12: 417-444.
Shu, L., Ueda, K., Chiu, I. and Cheong, H. (2011) Biologically inspired design. CIRP Annals – Manufacturing Technology, 60:673-693.
Sartori, J., Pal, U. and Chakrabarti, A. (2010). A Methodology for supporting “transfer” in biomimetic design. AIEDAM, 24, pp 483506
Strauss, A. and Corbin, J. (1990) Basics of Qualitative Research: Grounded Theory Procedures and Techniques, Sage.
Vattam, S., Helms, M. and Goel, A. (2010) A content account of creative analogies in biologically inspired design. AIEDAM, 24: 467-481.
Vincent J., Bogatyreva O., Bogatyrev N., Bowyer A. and Pahl A. (2006) Biomimetics: its practice and theory. Journal of the Royal Society Interface, 3, 471–482.
Vincent, J. and Mann, D. (2002) Systematic technology transfer from biology to engineering. Phil. Trans. of the Royal Society of London, 360:156-173.
Yen, J., Weissburg, M., Helms, M. and Goel, A. (2011) Biologically inspired design: a tool for interdisciplinary education. In Biomimetics: Nature-Based Innovation, Y. Bar-Cohen (editor), Taylor & Francis.
Yen, J. and Weissburg, M. (2007) Perspectives on biologically inspired design: introduction to the collected contributions. Journal of Bioinspiration and Biomimetics, 2.