A cognitively based approach to computer integration for design systems

Michael Ramscar, EdCAAD, Department of Architecture (and
Department of Artificial Intelligence), University of Edinburgh,
Edinburgh, UK

John Lee, EdCAAD, Department of Architecture (and Human
Communication Research Centre), University of Edinburgh, Edinburgh,
UK

Helen Pain, Department of Artificial Intelligence, University of
Edinburgh, Edinburgh, UK.

The kind of relationship a designer will have with a design assisting tool is dependent upon the kind of dialogue possible between the two, and any dialogue will be conditioned by the degree of mutual understanding existing between designer and assistant. In the first part of this paper we examine the methodology, currently prevalent in the computer-aided building design domain, that is supposed to address this problem: product modelling. We argue that product modelling is inherently incapable of fulfilling the goals that it sets itself, primarily because it is an attempt to move a methodology from, and appropriate to, a closed domain to an unbounded domain. In doing so, it fails to address the basic problem: it fails to provide a medium for dialogue. Dialogue entails mutual understanding; to support a dialogue between tools and designer, it is not sufficient for the designer to learn about the tool, the tool's designers must develop a strategy for enabling the tool to understand the designer. The second part of this paper comprises the theoretical background for a pragmatic strategy for classifying representations held by design agents by mapping the internal structures of differing representations of objects onto one another. Research in cognitive science has shown that such structure mappings can form the cognitive basis for certain kinds of classification decisions. We briefly describe ongoing work aimed at applying these insights in order to enable machines to classify, and thereby interpret, nonspecific design representations. Copyright © 1996 Published by Elsevier Science Ltd.

Keywords: analogical reasoning, artificial intelligence, design cognition, product modelling, communication

ELSEVIER
0142-694X/96 $15.00 *Design Studies* **17** (1996) 465–483 0142-694X(96)00025-7

1 Introduction

Our interest is in making the various tools that can assist in the design of safer and more efficient buildings available to designers in a form amenable to design practice. As a result of this, one of our primary concerns is integration. Unless the tools of assistance and the tools of design are integrated in some form of common environment, the utility of the former will forever be unable to assist in the use of the latter. The major challenge to be faced in integration in a multi-actor domain such as building design and construction is that of handling the number of different views held by actors—be they designers, engineers or computer programs—of the objects in the design domain: the problem of ascertaining whether or not the particular view one agent has of a given domain object corresponds to a representation in another agent such that communication of the object of representation (the representantum) can take place. The problem is a general one: an object can be any recurring class of phenomena, from concrete objects such as fish and bicycles to abstract concepts such as logic and truth. Since there is nothing inherent in a given object that necessarily governs the way in which it is represented (even though it seems equally necessary to assume that the properties of objects will in some way influence the content of these representations) it is unlikely that a presented representation will correspond identically to that present in the system.

A system lacking some kind of solution to this problem will suffer a number of limitations to its functionality. By constraining the user to using only those representations acceptable to the system, it will for instance: restrict designers to a predetermined subset of possible design solutions, forcing the designer to adopt the system view of the domain, obstructing and frustrating the designer's intentions; and be severely limited in its ability to accept input from external applications, preventing the functionality embodied within the system from being integrated with external applications which do not conform to the system's particular domain view.

The basic task of a computer system in multi-actor integration is to recognize similarities between the new representation and knowledge already present within the system (however encoded), such that the two representations can be seen as depictions of the same object. This is in essence a categorization task: the system must classify the two representations (internal and presented) as being different representations of the same thing (members of the set 'representations of *X*'). Thus the system must be capable of handling multiple representations and mediating between them. The deep problems this poses are not immediately apparent. After all, *A* and *B* are both representations of *X* if they represent *X*, so if one can determine independently that they both represent *X*, then one can proceed safe in the knowledge that both *A* and *B* represent the same thing: in this case, *X*.

The problem lies in determining individually whether *A* and *B* are both representations of *X*. Practically (and epistemically) the representantum, *X*, can play no direct part in this test. Since we can only perceive or reason about the representantum by representing it, *X* as a representantum drops out of the picture. Hence, any test along the lines of that described above degenerates into a comparison of *A* and *B* with *C*, where *C* is merely another representation of *X* (we shall assume, in this instance, that a *C* with suprarepresentational properties, such as a Platonic form, is unavailable). It follows then that categorization must be determined by information present within representations.

There are two basic assumptions that are universal to cognitive research into categorization[1]: firstly that concepts which are used in the representation of categories can be decomposed into smaller compositional elements or attributes; and secondly that similarities between categories and amongst category members can be derived from common distinctive elements. Since category members can only be reasoned about via representations, it follows that any approach to the problem of classifying objects in this way will have to be in some way pragmatic in nature: a given object *X* can only play a part in any reasoning through being represented, meaning that determining whether *A* and *B* are both representations of *X* will be contingent upon common features of representations of *X*, rather than any direct and arbitrating reference to *X* (though what counts as a commonality is difficult to determine: see Sections 2 and 3 below). Since no representation of *X* will have any more epistemic authority than another, the success of a given categorization method will be determined by its performance rather than an appeal to any external formal criterion.

One difficulty that research into categorization *per se* has encountered is the sheer enormity of the categorization problem: categorization plays an important part in perception, reasoning, language, motor performance etc.[2], not to mention the intimate relationship between categorization, semantic meaning and language. The very ubiquity of categorization leads to severe difficulties in isolating any 'single categorization process' (indeed, it is exceedingly unlikely that human categorization relies upon any single process). Yet, as we shall show in Section 3, the categorization problem is so unconstrained in itself that it seems that only an understanding of the kind of cognitive constraints which make human categorization judgements

[1] **Medin, D L and Barsalou L W** 'Categorisation process and categorical perception in **S Harnad** (Ed) *Categorical perception*, CUP, Cambridge, UK (1987) pp 455–490

[2] **Harnad, S** *Categorical perception*, CUP, Cambridge, UK (1987)

possible will enable us to support conceptual transactions—information exchange—in computer systems. Human concepts evolve dynamically, and, as we show in Section 2, cannot be supported merely by fixing rigid definitions upon them. This problem is particularly acute in design, where conceptual evolution, through the exploration of the possibilities of the ways that artefacts can be seen or represented, or even of what they can be, is an inherent, even defining, part of the process.

One area of research that is shedding light on the kind of cognitive constraints employed in human categorization is analogy. Analogy is an area in which decisions to classify representations as importantly similar is fundamental. Moreover, in stark contrast to work on categorization *per se* where most research is still very theoretical in nature, analogy research has yielded a number of computational models of the process by which the elements within representations are mapped onto one another in order to facilitate categorization decisions. Our intention, described in Section 5, is to explore the extent to which the cognitive principles isolated by research into analogy (detailed in Section 4) can be generalized to yield a pragmatically successful method for certain categorization tasks in the building design domain.

2 *Product modelling and the problems inherent in predefined domains*

A common feature of all information processing CAAD systems is their need for a model of the building concepts to be manipulated within the system[3,4]. This need for a model becomes a problem for design systems if the model is fixed, as it inevitably has needed to be in CAAD systems. These 'domain concept models' (or as they now tend to be called *product models*) are intended to completely define a 'design domain'. A comparison between these models and language grammars is often made[5]. Language grammars can be divided into two distinct subsets: those for closed 'artificial' languages with fixed bonds between syntax and semantics, where the language is defined in terms of its own grammar, i.e. programming languages; and those for natural languages, where the relationship between syntax and semantics is more complex, and where the languages exist and are used in spite of the absence of a defining grammar. Whilst complete grammars can be produced for closed world languages in virtue of their definition, the best results achieved in respect of natural languages are grammars that can deal with limited subsets of the language in certain constrained conditions. No complete—exhaustive and unambiguous—grammar for a natural language has yet been devised, or even seems possible.

The problem of defining a product data model is indeed analogous. If a domain is defined in terms of a model, then *de facto*, the model will be a

[3] **Bijl, A, Stone, D and Rosenthal, D** 'Integrated CAAD systems', UK Department of Environment Research Report, EdCAAD, University of Edinburgh, Scotland (1979)

[4] **Ramscar, M J A** Static models and dynamic designs: an empirical impasse vs. an inductive solution', in **R J Scherer** (Ed) *Product and process modelling in the building industry*, A A Balkema, Rotterdam, (1994) pp 69–76

[5] **Bjork, B -C** 'Intelligent front-ends and product models' *AI in Engineering*, Vol 6 (1991) No 1 pp 46–56

generic definition of that domain. On the other hand, in order to define a product data model for a real world domain such as building, a number of major problems must be addressed. The first and foremost of these is that there is no clear way of validating a product model: a consequence of the relationship between model and the modelled domain/object is that there is no way of proving that a product model relating to a real-world domain is complete. If W is the set of possible propositions describing some real-world domain, and we induce W' as a product model of that domain, we would require some further model W'', which would stand in the same relation to W' as W' stands to W in order to validate W. E.g.

$(W \leftarrow W') \leftrightarrow (W' \equiv \text{some determining model } W'')$ [*W' determines W because W' is equivalent to W'' which determines W, and therefore by implication W*].

Unfortunately, since we do not know W'' a priori, and since it seems likely that such predefinitive models of real-world domains do not exist, we are instead forced to generate W'' by empirical means (by induction from instances). This process runs into the same difficulties faced in defining W': without a further model W''', one will be unable to validate W'', which means that in turn one will be unable to validate W'.

$$(W \leftarrow W') \leftrightarrow$$
$$(((W' \equiv W'') \leftrightarrow ((W'' \equiv W''') \leftrightarrow (W''' \equiv W'''') \ldots \textit{ad infinitum}$$

This stands in direct contrast to the closed-world model. In this case, the defining grammar G acts as the validator for the induced defining model M' of domain M

$$(M \leftarrow M') \leftrightarrow (M' \equiv G)$$

thereby enabling the creation of a proof for M'. M' is a valid model of M if and only if M' is equivalent to G.

The impossibility of proving the completeness of a product model, and the corresponding problem that without a proof of completeness it is likely

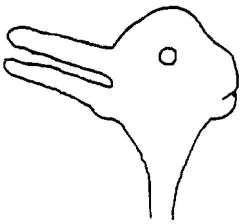

Figure 1 A CAD geometry product data model would only contain information structures concerning the relative position of the two lines which comprise the depiction, whereas a real world 'animal product model' would need to define information structures to accommodate rabbit and duck data (drawing from Wittgenstein[6])

that any product model is incomplete, are not simply minor logical inconveniences. Instead, they are inevitable consequences of real-world 'domains', where concepts are rarely stable for long. Conceptual exploration is an inherent part of the design, whether engineering or aesthetic: the Forth Road Bridge could never have been designed using a system that constrained the concept of a bridge to the subconcepts used in the design of the Forth Rail Bridge (Figure 2). A fixed conceptual model will prevent a system from supporting design, not enable it to offer design support.

2.1 *Some product modelling history*

The problems faced in creating a real world design product model for a domain such as architecture can be illustrated by contrasting this endeavour with an apparently similar area in which some progress has been made. A number of projects (CAD*I, Esprit project 322; CADEX, Esprit project 2195) have concerned themselves with the exchange of the geometric data produced by CAD programs.

Since CAD programs are used in the design of real-world objects such as buildings, the definition of a product data model to facilitate CAD geometry data exchange would appear to be a somewhat similar task to that of defining a product model to facilitate building data exchange. After all, if one can design a product model for a geometric description of a real-world object such as a building, it would seem to follow that one can define a product model for any other data entities associated with that real-world object. That this is not so is illustrative of the problems faced in modelling real-world objects, as opposed to closed-world models in which real-world objects can be represented.

The production of a data model for integrating the geometric output of a number of CAD tools is, in fact, a closed-world problem, the representation

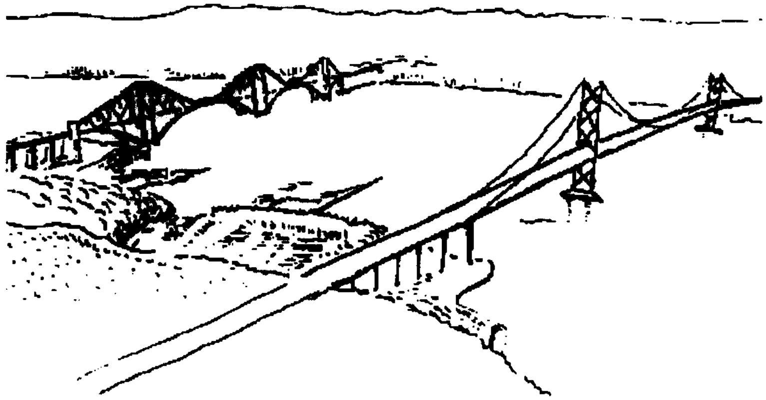

Figure 2 The Forth Road and Rail Bridges, an illustration of the empirical impasse. A conceptual model of the data structures involved in producing the Rail Bridge to the left would be unable to encompass the later design for the Road Bridge on the right. (Picture by Aart Bijl)

formalism of each tool being a closed language, with its syntax and operational semantics being circumscribed by a defining grammar G. Thus, to return to our earlier argument, a model M' for a drawing tool M is valid if it is equivalent to the grammar defining the tool's representational formalism, G.

$$(M \leftarrow M') \leftrightarrow (M' \equiv G)$$

Thus a product model M for a number of tools Tx can be validated if it can be demonstrated that M is equivalent to the grammars for those tools Gx

$$(Tx \leftarrow M) \leftrightarrow (M \equiv Gx).$$

This is because although real-world objects can be represented within closed-world geometric systems, they are uninterpreted within those systems. A 'building' is represented within a geometric system numerically rather than semantically, thus a product data model for a geometrical modelling system needs only to capture the semantics of the representational formalism(s) used within that system, rather than the semantics of whatever it is that is being represented within the formalism. Thus the model need only contain the structures and objects of the closed-world domain (geometric data, represented numerically), rather than any pertaining to a real-world domain (i.e. buildings). To take a familiar example, a closed goemetric model would need to represent the necessary numerical information that determines the relative position of the two lines making up the duck/rabbit in Figure 1: it would not contain any information pertinent to the determining of what it was in the real world that those lines were supposed to depict; drafting systems are free of the burdens of separating rabbits from ducks.

The products addressed by CAD*I, CADEX and similar projects are geometrical modellers, not the objects which they model. Returning to the earlier natural language grammar analogy: a CAD geometry data model might be likened to a data model for the integration of a number of word processing programs whereas a product model for, say, office buildings would be more akin to a model of the English detective novel (in the unlikely event of someone developing such a model, it would limit new books to the conceptual content of the idealized 'novel model'; it goes without saying that authors would be unlikely to accept such a situation). Predefined integration models are incompatible with the requirements of integrated design systems, since they will necessarily prevent design being done.

2.2 The empirical impasse

As the previous section shows, where the problem of real-world product model incompleteness is recognized—albeit sometimes unconsciously—there is a tendency to shy away from its implications: authors begin by talking about design, and end up discussing data exchange amongst building performance evaluation tools (BPEs)[7] or else talk about modelling in isolation from its applications.

More commonly, however, authors simply advocate product modelling in its purest form[8,9], with the requirement for a complete domain model to facilitate product data exchange. We have shown above the inherent problems involved in the production of real-world domain models.

There is no complete solution to this problem. Any computer interaction between designers on one hand, and design assisting tools on the other, will involve some form of inductive solution: the data and concepts current in the domain at the time a system (be it a closed proprietary system, or an open-system product model based approach) is developed (and embodied within BPE programs etc.) will have to be related to any new data structures or concepts embodied within new instances of design. And inductive processes are always in some way incomplete[10,11]. However, the inductive option offered by product models tends towards specification rather than interpretation, fixing in advance the scope and contents of the designs the system can accommodate (we might term this an *encapsulated* induction). As we have shown above, this is a reasonable approach towards the integration of bounded domains, but it cannot solve the problems posed by building design. It leads to an empirical impasse: in order to act as an integration standard, a product model must be complete, yet in virtue of the way in which it is—and must be—defined, such completeness is unachievable.

Product modelling fails because it attempts to put boundaries around domains. As we shall see when we discuss categorization below, these boundaries are impossible to draw. One cannot provide necessary and sufficient conditions for saying that something is, or is not an X, yet people seem very good at making categorical distinctions. The road bridge on the right in Figure 2 is quite clearly a bridge, despite the fact that in some ways it is quite fundamentally different to the earlier rail bridge. What is needed therefore is a pragmatic approach that captures at least some of those human skills used in making categorization judgements. In the following sections we examine some of those human judgmental skills, and suggest a way in which they can be incorporated within a computational strategy.

6 **Wittgenstein, L** (trans. Anscombe, E) *Philosophical investigations* Blackwell, Oxford (1953)

7 **Wright, A J, Lockley, S R and Wiltshire, T J** 'Sharing data between applications programs in building design: Product models and OO-programming *Building and Environment* Vol 27 (1992) pp 163–171

8 **Froese, T M and Paulson, B C Jr** 'Integrating project management systems through shared object-oriented project models' *Applications of artificial intelligence in engineering VII* **D E Grierson, G Rzevski and R A Adey** (Eds) Elsevier Applied Science, London (1992) pp 69–85

9 **van Nederveen, S, Bakkeren, W and Luiten, B** 'Information models for integrated design' *CAAD Futures 93* **U Fleming and S Van Wyk** (Eds) Elsevier Science, Oxford, (1993) pp 375–390

10 **Goodman, N** 'The new riddle of induction' *Fact fiction and forecast*, Harvard University Press, Cambridge, MA (1983)

11 **Holland, J H, Holyoak, K J, Nisbett, R E and Thagard, P R** *Induction: processes of inference, learning and discovery*, MIT Press, Cambridge, MA (1986)

3 Categorization and the representation of categories

3.1 Categorization

As we suggested in our introductory notes above, if one rejects prescriptive approaches such as product modelling—and since product modelling threatens more than it offers, it appears it must be rejected as an integration strategy for creative domains—then one is faced with a categorization task. One needs a system that will classify representations, handling multiple representations and mediating between them. Integration becomes a matter of classifying the views of the various agents (people and programs) using a system, rather than merely checking to see whether they conform to an overarching model (and rejecting them if they do not). This presents a new set of very different and equally difficult problems. We shall show below that where these problems differ from the prescriptive approach is that the new set of problems are not intractable.

Categorization poses an enormous challenge, and its importance can hardly be overstated. The categories formed by people represent only a minute fraction of the categories they could possibly form. As the number of unclassified objects to be partitioned into categories increases, so the number of potential partitions grows exponentially: it is possible to partition 3 unclassified objects in 5 ways; 4 objects in 15 ways; 5 objects in 52 ways: 10 unclassified objects yield more than 50,000 possible partitions. Thus the principles and constraints which govern human category formation are of central importance to any study which attempts to model human intelligence or knowledge.

Throughout the history of Western thought, categories have traditionally been defined in terms of a set of necessary and sufficient conditions for category membership—this approach is embodied in product model style approaches to integration. For a number of reasons (Smith and Medin[12] provide a review of the various conceptual and evidential arguments against categorizational 'classicism') this approach has been rejected.

The focus of most current research into categorization has switched to exploring the view that categories possess 'prototypes' or 'best examples' allied to somewhat vague boundary conditions. Thus exemplars of natural categories can be located on a rough continuum that ranges from the prototypical instances to borderline cases whose category membership status is dubious.

12 Smith, E E and Medin, D *Categories and concepts* Harvard University Press, Cambridge, MA (1981)

3.2 The classical account

The classical view's largest failing has been the failure over time to provide classical style definitions for most natural kind concepts. For instance, consider an attempt to specify the defining properties of a leopard. The most typical property a leopard has is its spots, yet if spots are made a necessary aspect of being a leopard, then such an account couldn't deal with a leopard which was black. Moreover, the classical account cannot accommodate the fact that people can reliably order instances of a given concept according to how typical or representative they are of the concept[13]. For instance, robins and bluebirds are considered typical birds, starlings and owls are usually considered less typical, and penguins and flamingos are generally considered to be atypical. This perceived typicality of objects has been shown to be uncorrelated to the frequency or familiarity of subjects' experience of instances[14].

Importantly, these ratings are predictive of performance in categorization tasks. Subjects make faster categorization decisions the more typical an instance is considered to be (e.g. *is a robin a bird?* versus *is a penguin a bird?*;[15]). Typicality also affects memory retrieval (typical instances of classes are recalled before atypical ones[16]), vocabulary development (children master the naming of typical instances more quickly than atypical ones[16]), deductive reasoning (in determining the validity of incomplete deductive arguments, e.g. *all birds are x, therefore robins are x*, subjects' time performance correlates to typicality[17]), and inductive reasoning (if an instance *I* of a concept has property *P*, estimates of the likelihood of another instance *J* also having *P* are higher the more *I* is adjudged to be typical[18]).

If the classical view were correct, and every instance of something is classed according to a definition, then one would expect instances of concepts to be equal; the inequalities inherent in the above typicality effects show quite clearly that this is not the case. Rosch and Mervis[19] experimented with getting subjects to list the properties of instances of concepts. For any given concept, they discovered that most instance properties were non-necessary (e.g. *bird, fly*) and that typicality was governed by the frequency of appearance of non-necessary properties in instances. Clearly these findings are incompatible with an account based upon defining necessary and sufficient properties.

3.3 Prototypes

This failure of the efforts to define a formal relationship between representations and the objects they represent is an important motivation for the pragmatic approach advocated herein. Instead of concepts being defined in

[13] **Malt, B C and Smith, E E** 'Correlated properties in natural categories', *Journal of Verbal Learning and Verbal Behaviour* Vol 23 (1984) pp 250–269

[14] **Mervis, C B, Catlin, J and Rosch, E** 'Relationships among goodness-of-example, category norms and word frequency' *Bulletin of the Psychonomic Society* Vol 7 (1984) pp 268–284

[15] **Smith, E E, Shoben, E J and Rips, L J** 'Structure and process in semantic memory: a feature model for semantic decisions' *Psychological Review* Vol 81 (1974) pp 214–241

[16] **Rosch E** *Cognition and categorisation* Lawrence Earlbaum Assoc., Hillsdale, NJ (1978)

[17] **Cherniak, C** 'Prototypicality and deductive reasoning' *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol 23 (1984) pp 625–642

[18] **Rips, L J** 'Inductive judgements about natural categories' *Journal of Verbal Learning and Verbal Behaviour* Vol 14 (1975) pp 665–681

[19] **Rosch, E and Mervis, C B** 'Family resemblances: studies in the internal structure of categories' *Cognitive Psychology* Vol 3 (1975) pp 382–439

terms of necessary and sufficient conditions, a view of concepts as prototypes (or best exemplars: representations of central tendencies) has grown out of the discovery of the above typicality effects. On this account, the prototype of a category contains the characteristic attributes of its category's examplars (i.e. attributes that are highly probable across category members, but are neither necessary nor sufficient for the determination of category membership). Category membership is thus determined by how similar an entity is to a prototype: instances above some threshold of similarity are classified as category members; all other, dissimilar, entities are classified as nonmembers.

The prototype account can thus successfully explain (and gain experimental support from) the typicality effects described above. The more similar an entity is to the prototype, the quicker and easier its categorization; the more similar a nonmember to a prototype, the slower and more difficult its exclusion. Most, if not all, categories will not have any clear-cut boundaries, thus it does not follow from the argument that basic object categories follow clusters of attributes that these attribute clusters are necessarily discontinuous.

Rosch[16] makes the following observations about prototype theories of categorization

- Any idea of *a prototype per se* is best seen metaphorically; what is really referred to are judgements of degrees of prototypicality
- The theory does not comprise any particular processing model of categories
- It does not constitute a theory of representation of categories
- Though prototypes must be learned, they do not constitute any particular theory of category learning

Thus prototypes are best seen as constraints upon, rather than specifications for, representation and process models.

3.4 Categorization and similarity

As can be seen from the above, judgements of similarity are essential to categorization. However, one drawback to current work in categorization is that studies tend to leave the mechanisms by which entities are adjudged to be similar unexamined. In part, this has been a result of the programme that psychological research into categorization has followed

> It should be noted that the issues in categorization with which we are primarily concerned have to do with explaining the categories found in a culture and coded by the language of that culture at a particular point in time. When we speak of the

> formulation of categories, we mean their formulation in culture. This point is often misunderstood. The principles of categorization proposed are not as such intended to constitute a theory of the development of categories born in to a culture nor to constitute a model of how categories are processed (how categorisations are made) in the minds of adult speakers of a language. (Rosch[16], p 28)

Given the intimate relationship between categorization and judgements of similarity, this approach can lead to a certain amount of question begging. Judgements of similarity may well be essential to categorization, but is is difficult to see how any similarity judgement could be made unless some category concepts were already present. Some idea of a category is necessary to any similarity judgement (even if the category in question is only a loose one, such as 'things that are similar'; such a category is a necessary aspect of the concept of similarity). Whilst leaving the concept of similarity unanalysed is reasonable for the psychological programme outlined above, a similar lack of analysis may not be the best strategy if we are to attempt to model, or at least replicate some part of, the ability of humans to make categorization judgements.

At a conceptual level, at least, categorization decisions are supervenient upon judgements of similarity, rather than merely dependant upon them. Some category of similar things is necessary to similarity judgements, and in turn similarity judgements are necessary to categorization. Given such a supervenient relationship between these processes, it follows that an investigation into the mechanisms involved in similarity judgements is likely to simultaneously reveal much about the mechanisms involved in what might be termed categorical reasoning, and *vice versa*.

3.5 Similarity, representation and description

Goodman[20,21] warns of the insiduous nature of similarity (an insidiousness which might explain why studies of categorization tend to leave similarity—their main explanational tool—unanalysed). All of Goodman's warnings relate to the enterprise laid out in these pages; some are worth a moment's reflection here.

He notes that similarity cannot pick out inscriptions that are tokens of a common type; there are no clear necessary and sufficient conditions for determining how an inscription of the letter *a* must resemble other inscriptions for it to be an *a* (Figure 3). The best we can do is say that all *a*'s are alike in being *a*'s. However, as Goodman points out

> to say all *a*'s are alike in being *a*'s amounts simply to saying that all *a*'s are *a*'s. The words "alike in being" add nothing; similarity becomes entirely superfluous (Reference 20 (p 439))

20 Goodman, N 'Seven strictures on similarity' in *Problems and projects*, Bobbs-Merrill, Indianapolis, IN (1973) pp 437–446

21 Goodman, N *Languages of art* Hackett, Indianapolis, IN (1976)

a	*d*	A
m	*w*	M

Figure 3 (from Goodman[20])

—which is essentially how Rosch determines similarity: an *a* is an *a* if we classify it as such.

Moreover, similarity cannot be equated with, or measured in terms of, possession of common characteristics: as we saw in the discussion of categorization above, definition of essential properties is invidious. Indeed, Goodman makes a further, telling point

> Where the number of things in the universe is n, each two things have in common exactly 2^{n-2} properties out of the total of 2^n-1 properties; each thing has 2^{n-2} properties that the other does not, and there are $2^{n-2}-1$ properties that neither has. If the universe is infinite, all these properties become infinite and equal (Reference 20, pp 443–444)

The majority of the readers of this article will share with the authors the characteristic of being approximately the same distance as one another from Mars, Uranus, the Crab Nebula etc., as well as the properties of not being in any of these places. The problem with common characteristics is that they are insufficiently constrained: again, what is needed is an account in virtue of which it can be said why certain common characteristics seem to be considered more important than others.

These remarks by Goodman question the use to which similarity is put. If, in addition to saying that two things are similar, we say that they have a property in common, we do not add to the initial statement; we render it superfluous, since saying that two things are similar in having a property in common is to say no more than they have a certain property in common.

Thus the explanational emptiness of the claim that judgements of similarity play an important role in categorization is brought sharply into focus: categories are distinguished by commonalities of properties across category members; judgments of these commonalities are essential to categorization. Similarity alone, therefore, is superfluous to explanations of categorization. On the Rosch account detailed above, what determines category memberships is clusters of property commonalities. Similarity serves as a blank space in the account, since we have no explanation for how these commonalities are determined.

What is needed then is an account of similarity in terms of the mechanisms underpinning judgements of commonalities. As Goodman[20] shows, similarity is used disparately and ubiquitously, and it is unlikely that any one process can accomplish such a variety of tasks. Thus, in isolating a similarity mechanism, the determining of the limits of its use will be an important part of the investigation into its potential applications. At least one body of research has revealed much about the mechanisms of one way in which entities are adjudged to be similar: research into analogical reasoning. These mechanisms provide an insight into the underlying structure of some similarity decisions: the kind of insight that makes possible some forms of automated categorization task.

4 *Analogy*

Analogy is a central cognitive process. The use of existing information in order to explain novel concepts or generate new ideas occurs with great frequency over many levels of cognitive activity, from the interpretation of visual perceptions to the metaphors of everyday speech. Importantly, analogy is a subprocess of categorization: in making an analogical connection between two representations, we classify the representations as being meaningfully similar.

The analogical reasoning process can be broken down into two basic subprocesses (it can be, and invariably is, broken down into more).

- *Accessing*: i.e. the problem, mentioned above, of explaining how an analogy is accessed from memory. How is it that the stored understanding of two analogous processes in memory—say the workings of the solar system and the hydrogen atom—enable us to select one as illustrative of the other?
- *Mapping*: following on, how does the person to whom the analogue has been presented map his prior knowledge of the solar system onto his, perhaps somewhat hazy, knowledge of the hydrogen atom in order to gain a firmer understanding? What common properties must be mapped between two domains for them to be considered analogous?

4.1 *Analogical mapping*

If we characterize analogy as being a process whereby a partial similarity match is made between certain common features of different domains, then there must be an account of the factors that influence the deciding of which commonalities are mapped and which are disregarded. To take for instance an analogy between the water supply and electricity, a number of relations could be mapped: both are distributed in networks from central supply companies (all of which (in the UK) are newly privatized natural monopolies . . .); both can be used in cooking; degree of pressure determines flow rate etc.—what acount can be given for people's observable fluency in performing the correct analogical mappings?

Gentner's structure mapping theory[22–24] shows that the mapping and inference between two domains can be achieved by assigning correspondences between objects and attributes and then mapping predicates with identical names. In order to do this, Gentner assumes a predicate-like representation, distinguishing between *objects, object-attributes* and *relations.* Object-attributes are those predicates that have one argument and describe object properties, e.g. RED (lobster). Relations are divided into a hierarchy of orders, with those predicates with two or more arguments which are used to describe relations between objects, for example UPSETS (stomach, lobster) forming the lowest order, and those predicates describe different levels of relationships between relations e.g. CAUSE (UPSETS (stomach, lobster), DRINKS (alka-seltzer, dinner)) forming the higher orders.

The theory itself comprises two parts: *mapping rules*, and the *systematicity principle*. Mapping rules state that

- Attributes of objects are not mapped
- Relations between objects are preserved

Whilst the systematicity principle requires that

- Complex higher-order relations (e.g. CAUSE above) are mapped preferentially followed by relations that constitute the higher order arguments

which is intended to capture the notion that analogy conveys a system of connected knowledge, rather than an assortment of independent facts

> 'structure mapping stems in part from the observation that useful analogies, such as those used in science or education, involve rich, interconstraining systems of mappings between two domains, rather than a set of independent correspondences' (Clement and Gentner[24], pp 91–92)

In order to demonstrate that systematicity acts as a selection filter during analogical transference, Clement and Gentner performed three experiments that looked separately at two of the components of analogical mapping

- Matching existing information in the base and target
- Inferring new information about the target that follows from the analogy with the base domain

[22] **Gentner, D** 'Structure-mapping: a theoretical framework for analogy' *Cognitive Science*, Vol 7 (1983) pp 155–170

[23] **Gentner, D and Gentner, D R** 'Flowing water or teeming crowds: mental models of electicity' in *Mental Models* Lawrence Erlbaum, Hillside, NJ (1983) pp 99–129

[24] **Clement, C A and Gentner, D** 'Systematicity as a selection constraint in analogical mapping *Cognitive Science* Vol 15 (1991) pp 89–132

The first tested whether systematicity constrains the matching process. A novel set of analogies were created. It was suggested that if systematicity plays a part in the matching process, then subjects would show a corresponding preference for those matches that are embedded within a matching causal system rather than those where causal systems are unconnected. The second and third experiments explored the notion of systematicity acting as a constraint upon inferences carried over from the base domain to the target. It was expected that subjects would not just infer any base fact, but would rather select a fact that follows from a shared causal network.

Experimental results supported these predictions. In the first experiment, subjects preferred the matching fact that was embedded within a matching causal system, whereas the control group, who saw only the target domain showed no preference for the shared system fact. This is evidence that the systematicity principle does indeed constrain matching: 'analogical matching is not merely a feature-by-feature decision: analogical matching concerns systems of predicates, not individual predicates'[24]. Experiment 2 also provided support for systematicity, in that subjects rejected inferences that resulted from isolated correspondences in favour of those that were supported by a larger causal (i.e. systematic) network. Indeed, the experiment showed that some subjects explicitly sought such systematicity in the generation of their inferences. Experiment 3, in which subjects had to rely on their memory representations of the base domain, provided still further support for systematicity, though the nonavailability of the base representation during mapping and inference did result in poorer results than in experiments 1 and 2. However, the results of these experiments support the hypothesis that systematicity acts as a constraint in the selection process; that the choice of which lower-order relations to map is not determined just by the independent relations themselves, but by the interconnections amongst such relations.

4.2 The structure mapping engine

The structure mapping engine (SME[25]) is a computer implementation of the principles of the structure mapping theory, which simulates the process of interpreting and making predictions from an analogy. SME uses systematicity extensively: both in evaluating a candidate mapping, and in deriving inferences.

SME has been run with success on examples from a number of different domains. From a psychological point of view, the model has two principle features

- The same processes that are used to form mappings also generate inferences

[25] **Falkenhainer, B, Forbus, K D and Gentner, D** 'The structure mapping engine: an algorithm and examples' *Artificial Intelligence* Vol 41 (1989) pp 1–63

- Goals are not required in the formation of coherent matching structures, nor in the generation of inferences

From an experimental point of view, SME possesses a pleasing simplicity. By focusing upon only one factor in mapping, namely the most important one, it offers the possibility of gaining an insight into the influence of that factor in the mapping process. Whereas other models, e.g. Holyoak *et al.*'s ACME[26], try to model every possible constraint that might lead to a mapping, creating a muddied view of the role of individual factors, SME allows one to gauge the role of factors other than structure by measuring the gap between the purely structural machine interpretation and a more complex human one.

4.3 Accessing analogies

It seems clear from the evidence that structural systematicity can act as a determinant of similarity. Gentner's theory provides an implementable model of (at least some of) the cognitive constraints governing similarity that was missing from the accounts of categorization discussed previously. Unfortunately, taken on its own, systematicity can only tell us how to make similarity mappings: it cannot assist us in choosing a candidate through which we can map and interpret. If we consider the classification task, we might have a number of prototypical representations of concepts in memory: mapping can enable us to interpret a presented representation only if we have another mechanism to select the appropriate prototype representation from memory. This is the problem of accessing, or retrieval.

The question of accessing analogies is also closely bound up with judgements of similarity. Analogy is merely one of a number of ways by which two things might be adjudged similar. Accordingly, in general, accessing models for analogy are based upon the more general principle of similarity based retrieval.

Gentner and Forbus[27] distinguish three differing classes of similarity match

- *Literal similarity* matches include both common relational structure and common object descriptions
- *Surface matches* are based upon common object descriptions, with some shared first-order relations
- *Analogy* (as described earlier, a match based upon a common system of internal relations)

Thus this account, in keeping with earlier work, defines similarity in terms of degrees of correspondences between structured representations. The new feature, however, is that Gentner and Forbus argue that similarity based

---

[26] **Holyoak, K J and Thagard, P** 'Analogical mapping by constraint satisfaction' *Cognitive Science* Vol 13 (1989) pp 295–355

[27] **Gentner, D and Forbus, K** 'MAC/FAC: a model of similarity-based retrieval' in *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Lawrence Earlbaum, Hillsdale, NJ (1991) pp 504–509

access from long-term memory relies more on surface similarities and less upon structural commonalities than mapping. Gick and Holyoak[28] observe that people often fail to access potentially useful analogues, whilst Ross[29,30] showed that whilst people engaged in problem solving are often reminded of prior problems, these remindings are usually based on surface rather than structural similarities between solution principles.

Gentner and Landers[31] and Ratterman and Gentner[32] examined the role of structural and semantic factors in accessing. Gentner and Landers' experiment had two purposes: to test the prediction that shared systematic structure determines the subjective soundness of a match; and to see whether the accessibility of analogies (and other similarity matches) mirrored their inferential soundness.

Subjects were presented with a number of scenarios sharing a diverse range of commonalities with a base scenario, in order to test for the effects of these commonalities on inferencing and recall. Gentner and Landers found that surface commonalities significantly influence memory access. Yet when the same subjects were then asked to rate the inferential soundness of their matches, they discovered that shared systematic structure played a more important role, with less easily accessed analogies being judged more inferentially sound than more easily recalled scenarios that shared only object level attributes with the base.

One problem with the Gentner and Landers' study is that it did not eliminate the possibility that retrieval was in fact simply dependent upon the overall similarities between the matches. In order to clarify this, Ratterman and Gentner added a similarity rating task in order to test whether similarity could be predicted simply by similarity ratings. Thus Ratterman and Gentner were testing three parameters: accessibility (recall); inferential soundness; and the degree of similarity between base and target.

By repeating the Gentner and Landers study in this way, Ratterman and Gentner discovered

- Accessibility: literal similarity and mere appearance led to significantly more remindings than true analogy and false analogy, supporting the results of the Gentner and Landers tests
- Soundness rating: again the results of the Gentner and Landers tests were verified
- Similarity rating: with respect to the base scenario, subjects rated the scenarios with object matches and shared systematic structure significantly higher than the analogy matches, and both significantly higher

---

[28] **Gick, M L and Holyoak, K J** 'Analogical problem solving' *Cognitive Psychology* Vol 12 (1980) pp 306–355

[29] **Ross, B H** 'Remindings and their affects in learning a cognitive skill' *Cognitive Psychology* Vol 16 (1984) pp 371–416

[30] **Ross, B H** 'This is like that: the use of earlier problems and the separation of similarity of effects' *Journal of Experimental Psychology: Learning, Memory and Cognition* Vol 13 (1987) pp 629–639

[31] **Gentner, D and Landers, R** 'Analogical reminding: A good match is hard to find' in *Proceedings of the International Conference on Systems, Man and Cybernetics*, Tucson, AZ (1985) pp 607–613

[32] **Ratterman, M J and Gentner, D** 'Analogy and similarity: determinants of accessibility and inferential soundness', in *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, Lawrence Earlbaum, Hillsdale, NJ (1987) pp 23–35

> than matches with only object commonalities, which were rated equally lowly. Moreover, the pattern of similarity ratings mirrored those for soundness ratings.

Since the results of the recall task on the one hand, and the similarity and soundness ratings on the other, varied markedly, Ratterman and Gentner concluded that different aspects of similarity govern the different processes, with surface matches playing the major part in recall, and structure being most significant in judging soundness and similarity.

Accordingly, the Gentner *et al.* similarity-based transfer process is decomposed into two subprocesses that are qualitatively different. *Accessing* a similar (*base*) situation in long-term memory, based primarily upon surface similarity; and creating a *mapping* from base to target using structural commonalities.

4.4 Implementing retrieval

MAC/FAC (for 'many are called, few are chosen'; Gentner and Forbus[27]) uses a two-stage retrieval process based upon Gentner's retrieval theory. It comprises MAC, a crude, computationally cheap matching process used to select a limited number of candidates for more expensive matching using FAC, which is SME (above), in order to apply structural constraints to select the best match(es). Both the MAC and FAC stages consist of a *matcher*, which is applied to every input selection, and a *selector*, which uses the evaluation of the matcher to select which comparisons are produced as the output to that stage.

The MAC stage matcher is used to estimate how well FAC will evaluate comparisons, in order to filter down candidates into a number suitable for the more extensive (and computationally expensive) processing in FAC. Since MAC is only sensitive to predicate overlap, and FAC is structure sensitive, FAC will reject much of MAC's output. However, the filtering provided by MAC does cut down the number of matches FAC is required to do.

The chief criticism of the MAC stage focuses on the narrowness of its definition of semantic similarity. Holyoak and Thagard[33] note that this results in too many missed retrievals. Gentner and Ratterman[32] discovered that first-order structural relations do play a part in accessing, but this is not modelled in MAC/FAC. Although these are reasonable criticisms, in practice they ignore memory-size. Human memories are far larger than any cognitive simulation produced to date. In the case of the latter, the problem of irrelevant retrievals ('false positives') becomes critical. Thus, whilst the

33 Holyoak, K J and Thagard, P *Mental leaps* MIT Press, Cambridge, MA (1995)

MAC stage might not be the most psychologically plausible possible solution, it is the best practical one (it can be successfully run as a computer simulation).

5 *Structural systematicity as a determinant of categorical similarity*

5.1 *Breaking the circle of similarity*

The aim of our current work is to explore the possibility of transferring the knowledge of the mechanisms involved in judgements of similarity gained in studying analogy to the problem of categorization. The intention here is not to develop any theory of categorization, but rather to explore whether the insights into similarity judgements offered by Gentner *et al.* can fill the large blank that exists in the categorization account afforded by psychological work such as that of Rosch in such a way as to make possible the development of an automatic classification system.

It is classification decisions, rather than the nature of 'categories', that are important to us. The classical view of categorization is muddied by a naive realism that assumes that categories exist in some way independent of classification decisions, and that the goal of categorization research is to uncover these categories. This naive realism is invidious (as can be witnessed by the product modelling approach discussed earlier): indeed, the distinction between analogical reasoning and categorical decision making can only be made by appeal to it. If we accept that categorical decision-making involves reference to independently existing categories, whereas analogy somehow points to similarities between them, then it is easy to distinguish the two; if one rejects the idea that categories have some independent existence (whether from pragmatic or metaphysical reasons, one *ought* to reject this idea), then analogical reasoning simply becomes a subarea of categorical decision making (a category within a category).

As we argued above, the blank in Rosch *et al.*'s theory of category prototypes appears whenever the word 'similarity' occurs. Unless 'similarity' is fleshed out with some explanatory properties, the statement

> 'similarities between categories and amongst category members can be derived from common distinctive elements'. (Medin and Barsalou[1])

carries no more information than 'category members have category membership in common'. As Goodman[22] showed, the common distinctive elements referred to above are those we settle upon in order to decide similarity. Since any two given objects have the potential to share as many properties as any other two, what is needed is an explanation of what it is that makes some common properties distinctive. And what makes these common properties distinctive is that they are held in common: i.e. category members are similar in some respects; similarities are derived from objects being considered to be similar. This explanatory circularity is not unique to categorization research. Bareiss and King[34] have shown that identical criteria are widespread in determining similarities between cases in case-based reasoning.

Gentner *et al.*'s research into analogy offers a way of breaking this circle (at least for some similarity judgements). Systematicity has demonstrated[26] that structural correspondences can play a determinant role in some similarity judgements (or, in concrete Goodmanian terms, what we mean by similarity in some circumstances is structural correspondences). This offers up the possibility that for at least some category decisions, Medin and Barsalou's rather opaque 'similarities between categories and amongst category members can be derived from common distinctive elements'[1] can be recast more meaningfully as 'similarities between categories and amongst category members can be derived from common internal structural relations'. As Clement and Gentner[26] show, this is just what happens in decisions to classify two representations as analogical.

5.2 Analogies between representations: integration through classification

The hypothesis presented here is that structural correspondences will be useful in categorizing representations of certain kinds of structured descriptions in the manner described above. Our contention is that, just as structural correspondences can be deterministic of analogical classification decisions, they can also act as a determinant upon which to base at least some other kinds of classification decisions. Moreover, since the determination of structural correspondences is amenable to computer implementation, *vide* MAC/FAC, in areas where structural correspondence classification techniques are appropriate they offer up the possibility of approaching the problems of software integration and the man–machine interface through pragmatic rather than prescriptive means.

We are examining the use of structural correspondence mapping techniques as a basis for the classification of user/software representations—so-called agent views—within computer-aided design systems. Our approach is to attempt to develop a system where, rather than limit input to a set of precircumscribed representations, designers can represent objects as they see them, with the system then mapping structural correspondences between that representation and a prototypical representation, selected semantically from memory in order to classify it.

---

[34] **Bareiss, R and King, J** 'Similiarity assessments in case-based reasoning' in **K Hammond**, (Ed) Proceedings of the Case-Based Reasoning Workshop, San Mateo, DARPA, Morgan Kaufman (1989) pp 67–71

There is a strong parallel to be drawn between the approach proposed and the frame proposals of Minsky[35] and the prototype system proposed by Bobrow and Winograd[36]. Frames or prototypes contain a description of concepts, however, the description is considered neither necessary nor sufficient to determine the applicability of the concept; the frame serves as a *characterization* of the concept. A major problem these early frame systems encountered was determining which aspects of this characterization were appropriate for use at a given time (i.e. answering the question why this element of the frame should be used and that element ignored in a given process). Winograd and Flores[37] observe that

> if we look at the literature on frame systems [for an answer to this question] we find a mixture of hand-waving and silence. Simple rules don't work. If, for example, defaults are used precisely when there is no explicit (previously derived) information to the contrary, then we will assume that one holds even when a simple straightforward deduction might contradict it. If analogies are treated too simply, we attempt to carry over detailed properties of one object to another for which they are not appropriate. (p 117)

Where the approach advocated here differs from those criticized by Winograd and Flores (and, where it meets those criticisms) is that it proffers an answer to the question of how to treat analogies, and at what level of detail properties are to be carried over.

In our initial experiments, we are using simple electric circuits to provide an example of the kind of domain in which our approach might prove appropriate. Electric circuits offer a number of experimental benefits: it is possible to define an ideal understanding of the principles involved in a given electric circuit, however, in practice, because the mechanisms involved are essentially invisible, most explanation/understanding of simple electric circuits is by reference to analogies. A number of different analogies are employed in explaining different properties of electric circuits, and it is expected that the different underlying models employed by the divergent analogies used in explanation/understanding will result in a range of different representations of the same concepts.

There are two main analogies used in the explanation/understanding of electric circuits: the waterflow analogy and the crowds and passages analogy.

In the crowds analogy, electric current is characterized by masses of people rushing through passageways (wires). Thus current corresponds to the number of people pushing through a passage at a given time and voltage to how powerfully they push, etc. This model is particularly useful for

---

[35] **Minsky, M** 'A framework for representing knowledge', in **P Winston** (Ed) *The psychology of computer vision* McGraw Hill, New York (1975) pp 211–277

[36] **Bobrow, D and Winograd, T** 'An overview of KRL' *Cognitive Science* Vol 1 No 1 (1977) pp 3–46

[37] **Winograd, T and Flores, F** *Understanding computers and cognition* Ablex Publishing, Norwood, NJ (1986)

explaining how resistors function: a resistor is analogized to a barrier in a passageway containing only a narrow gateway. The crowd analogy does, however, lack a good analogue for battery behaviour.

In the waterflow analogy: pipes characterize wires; narrowings in pipes are analogous to resistors; water pressure to voltage; current to flow etc. The waterflow analogy is particularly good at explaining the behaviour of batteries, owing to the strong analogy between pumps and batteries.

We expect that explaining combinatorial circuit behaviour (the effects of different combinations of batteries and resistors in a circuit) to experimental subjects by use of combinations of the different analogies will result in a range of internal representations of circuit element functions and properties within subjects, and that correspondingly these differences will be apparent in the subject's own representations of combinatorial circuits. The intention being that this exercise will then present a prototype system, based upon the kind of structure mapping system developed to model analogical reasoning, with the challenge of classifying the various subject-produced representations into categories according to what kind of circuit it is that is represented.

By conducting our early experiments in a domain which guarantees a wide range of possible representations, we hope to obtain a good idea of the kind of problems we shall encounter—and the performance of our resultant solutions to these—more generally.

6 Conclusions

6.1 A new approach to integration

In the first part of this paper we demonstrated that a simplistic treatment of representations, and the information contained therein, is an inadequate response to the problem of integrating computer support with domains such as building design, where the particulars of representation necessarily cannot be determined in advance. We cannot know in advance the structure and content of a particular design representation; if we limit designers to exploring only a system designer's anticipation of this information, we set in stone arbitrary restrictions to which designs must adhere, and erect unnecessary barriers in the face of creativity. These problems are identical to those identified by Bijl *et al.*[3] in their survey of what was then the state-of-the-art in integrated computer-aided architectural design systems. That this is so is unsurprising: the theoretical basis of product modelling, and other similar integration strategies, is the same as that which underpinned those systems nearly 20 years ago.

In the second part of this paper we have outlined a methodology—incorporating some of the advances made in cognitive science in those intervening 20 years—for integrating CAAD tools into design practice whilst avoiding the pitfalls inherent in the prescriptive approach. The idea is to replicate the strong structural elements inherent in some cognitive classification processes within a system which interprets the content of representations by classifying them as examples of representational categories. I.e. the system interprets a designer's representation of X by seeing that it belongs to the same category of thing as its own representation of X. Whilst at present this work is in its infancy (confined currently to the design of electric circuits) we belive that it is only by re-exploring the methodologies underlying computer systems in the manner proposed here that 20 years of stalemate can be broken.

We argue that the relationship between designer and assisting tools must take the form of a dialogue. In the past, management of this dialogue has taken the form of specifying in advance what it is that the designer is allowed to say. This is hardly the way to go about assisting design; in the future, we shall have to understand what it is that the designer is telling us.

6.1 A suitable paradigm for design research?

The problems of product modelling stem from a one-way interdisciplinary transaction: the taking of a technique from a suitable domain (software integration) and attempting to apply it directly within another unsuitable domain (design integration). We believe that our research offers a better model for multidisciplinary research: whilst the techniques we have adapted from cognitive science offer the prospect of minimally prescriptive integrated design tools, the constraints provided by attempting to satisfy a particular concrete goal—design integration—have provided a new perspective on categorization and analogy, and these insights are being fed back into that discipline, furthering the understanding and cognitive modelling of these phenomena.

Acknowledgments

The authors would like to thank Richard White and Paul Brna for their invaluable help, and for their influence in shaping the ideas behind this work. Some parts of this paper were previously presented to the First European Conference on Process and Product Modelling, Dresden, October 1994; the feedback we received from attending practitioners is gratefully acknowledged. This research was funded in part by EPSRC research grant GRJ 76897.