On the Limitations of Memory Based Reasoning

Pádraig Cunningham
Department of Computer Science
University of Dublin
Trinity College, Dublin 2, IRELAND
(E-mail: Padraig.Cunningham@cs.tcd.ie)

Barry Smyth, Tony Veale
Hitachi Dublin Laboratory
Trinity College Dublin
College Green Dublin 2, IRELAND
(E-mail: {barry,tony}@hdl.ie)

Memory-Based Reasoning (MBR) represents a radical new departure in AI research. Whereas work in symbolic AI is based on inference and knowledge representation MBR depends on using a large memory of examples as a reasoning base. The MBR methodology is empirical so a typical system does not contain an explicit domain model. This means that MBR systems are quick to set up so the methodology shows considerable promise for knowledge based systems development. Indeed some impressive full scale systems have been demonstrated. In this paper we argue that despite this initial success there are considerable limitations to what can be achieved with MBR. We believe that the absence of a domain model means that MBR will not succeed in complex applications. We illustrate problems in natural language processing and planning that will require access to domain theories in their solution. Our conclusion is that the memory oriented philosophy of MBR has advantages but, for truly intelligent systems, this philosophy is better realised in the CBR paradigm where it can be integrated with a strong domain theory.

1. Introduction

This paper is prompted by concern about some of the claims made in research on Memory-Based Reasoning (MBR). The motivation behind the MBR movement is that the advent of massively parallel computing radically changes the way AI should be done. The argument is that intelligent performance can be based on a large empirical database of examples and in particular the requirement that the symbolic AI paradigm has for a strong domain model is avoided (Stanfill & Waltz, 1986; 1992), (Kitano, 1993), (Creecy et al., 1992). In this paper we will attempt to illustrate the unavoidable requirement of domain models and inference mechanisms for more complex tasks. We will use examples from planning and natural language processing to illustrate situations where empirical analysis is not adequate and inference based on a strong domain model is required.

MBR uses similarity metrics to find examples in memory that are similar to the problem in hand and then uses the actions associated with these examples to deal with the current situation. This is similar to the methodology used in Case-Based Reasoning (CBR) with the important difference that CBR systems incorporate some domain theory to guide the case matching and adaptation processes (see Kolodner, (1991); Veloso & Carbonell,(1991); Smyth & Cunningham, (1992), for descriptions of representative CBR systems). Thus MBR is a significant departure from the physical symbol system hypothesis of traditional AI. Instead it is memory centred, depending on a large content addressable memory of examples quickly accessible on parallel hardware. This means that MBR systems are easier to set up than symbolic AI systems as the knowledge engineering task is largely avoided (see Creecy et al., (1992) for example).

Advocates of the MBR idea have argued for the advantages of MBR over techniques based on knowledge representation and inference; the contention being that MBR techniques take better advantage of massively parallel architectures. This argument is a performance argument, whereas we believe that the crucial drawback with MBR is the lack of domain semantics. It is clear that MBR is a powerful methodology for some tasks for example; phrase translation (Kitano & Higuchi, 1991), classification (Creecy et al., 1992), or information retrieval (Stanfill & Waltz, 1992). However, there is reason to doubt that this impressive performance on what are essentially 'entry level' intelligent tasks will be manifest in tasks of greater complexity. In this paper we will describe problems in natural language processing and planning/design where it is clear that MBR will not succeed without an adequate domain model that captures the causal interactions between domain elements.

Our conclusion is that the memory oriented philosophy of MBR has advantages but, for truly intelligent systems, this philosophy is better realised in the CBR paradigm where it can be integrated with a strong domain theory.

Before elaborating on these arguments a brief introduction to MBR will be presented in the next section.

2. MBR and Massively Parallel AI

Before examining the MBR methodology it is worth reflecting on the problems with the conventional AI paradigm that MBR wishes to avoid. The case is presented [in Kitano '93]. 'The traditional AI approach' is based on the physical symbol system hypothesis. He argues that it has the following characteristics:-

- Formal Representations

- Rule driven inferencing

- Strong Methods: domain theory must be understood

- Hand-Crafted Knowledge Base

This approach has had considerable success with 'toy systems' but has had problems scaling up to larger applications. Kitano uses the diagram in Figure 1 to contrast the approach of traditional AI with that in MBR.
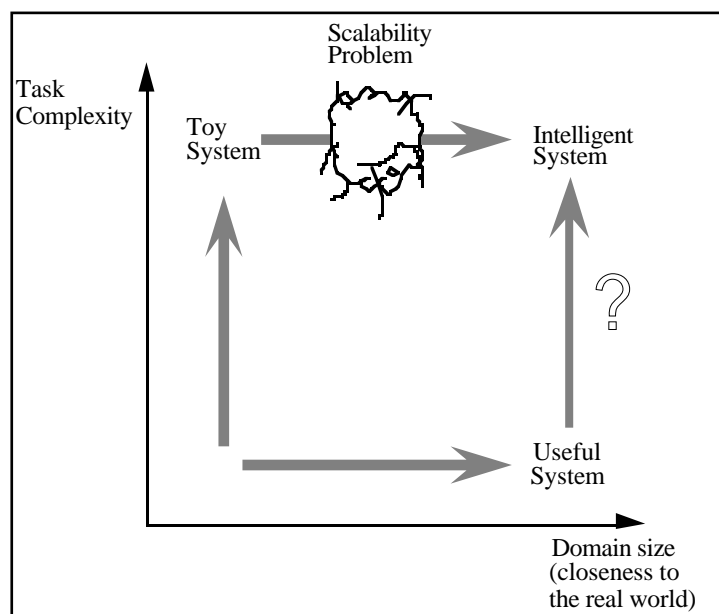
(based on (Kitano, 1993))

Figure 1: Approaches for building intelligent systems

The techniques of traditional AI have been successful in developing 'Toy Systems' that exhibit competence in limited domains. There is a widely recognised problem in scaling up this competence to a more broadly based intelligence. The approach in MBR is different. MBR focuses on performance from the start, concentrating on useful full scale systems for simple tasks. This performance on simple tasks has been illustrated in useful systems (see Creecy et al., (1992) for instance) however it is not clear that this methodology can be extended to cover more complex tasks.

The basic idea in MBR is to found intelligence on a large memory of examples. Nearest match computation is used to find the items in memory most similar to a current situation and then the actions associated with these items are used to deal with the current situation. Kitano emphasises three characteristics that are essential to the success of the MBR idea. These are; massive power, massive data resources and sophisticated modelling. He emphasises the importance of this modelling acknowledging that results will be poor if the modelling is not appropriate (Kitano, 1993). For the simple tasks tackled to date this sophisticated modelling involves structuring examples in a manner that allows useful remindings to be generated using statistical techniques. We believe that for more complex tasks this sophisticated modelling will be every bit as complex as the knowledge engineering that MBR wishes to avoid.

One of the major advantages of MBR is that it is a technique that takes full advantage of massively parallel hardware. Proponents of MBR argue that this is in contrast to logic based or rules based techniques used in

traditional AI. While it is true that MBR can take full advantage of parallelisation there has been considerable success in transferring the symbolic reasoning paradigm to parallel systems [Evett et al. '93]. Evett and Co. have shown that on parallel hardware recognition queries can be performed on a large knowledge base in $O(d)$ time where d is the depth of the knowledge base. This is in contrast to $O(B^d)$ (where B is the average branching factor in the network), the worst case performance on serial machines. This research is important in that it illustrates that AI on massively parallel hardware can be based on symbolic inferencing.

Some examples of areas of application of MBR to date are as follows:-

- Pronouncing English words (MBRTalk)

- Text retrieval (CMDRS)

- Classifying free text questionnaire responses (PACE)

- Speech to speech translation (ΦDMDIALOG)

It is worth looking at the details of one of these systems in order to get some understanding of the characteristics of MBR. The PACE system (Creecy et al., 1992) is a typical MBR system for classifying US Census Bureau long forms. These forms contain free-text responses to questions on occupation, company, duties and industry type. These responses have to be classified into predetermined categories; 232 industry categories, and 504 categories in the case of the occupation response. PACE uses a database of 132,000 manually classified returns as its reasoning base. The responses under consideration are compared with these using nearest neighbour techniques to determine the best classification. Similarity is normally determined based on features and in this task the features are the words in the responses. One of the main contributions of the PACE work has been the analysis of the reasoning base that determines the features that are most predictive in classification. Evidently, these features are weighted in the nearest neighbour classification. PACE runs on a CM2 Connection Machine so it meets the three criteria of, massive power, massive data resources and sophisticated modelling for MBR emphasised by Kitano. It is also a wholly empirical system.

In terms of the structure presented in Figure 1 this is a 'Useful System'; it tackles a full scale but simple problem. In the next few sections we will argue that this architecture will not be adequate for addressing more complex tasks. In particular we will argue that a system lacking a domain model will not be able to detect useful similarity in **retrieval** and will not be able to perform the required **adaptations** on retrieved solutions. We will conclude that domain knowledge must be available for effective retrieval and adaptation as is the case in CBR.

3 MBR in NLP: A Case of Performance Vs. Competence?

Inasmuch as strong MBR proponents seek to overturn the apple cart of traditional rule-based AI, they necessarily also part company with established formal linguistics on the complementary issues of competence and performance in language comprehension, and in particular, translation. Formal linguistics traditionally seeks to account for performance data, the actual use of language by a native speaker, in terms of a rule-based grammar of linguistic competence which generatively models the language ability of the speaker. In contrast, proponents of MBR-based translation claim that a rich example-base, in effect a highly-structured corpus of performance data, tied to a sophisticated statistical model of matching and retrieval, provides a sufficient basis for translation *without* the need for an explicit model of competence.

To commence the argument against this latter view, it is useful to first highlight those aspects of translation in which a performance-based approach reaps strong results. One such boon occurs in the treatment of language *aberrations* such as metaphor and metonymy, in which conceptual transpositions are performed, for reasons of eloquence or conciseness, in violation of default semantic constraints (see Veale & Keane, 1992). For example "boiling the kettle" is quite a different action to "boiling an egg", the former usage employing a metonymic shorthand for "the water in the kettle". Likewise, "to play Mozart" and "to play the Jupiter symphony" both use the verb "play" in isomorphic syntactic constructs, but with differing conceptual intent. The conventional rule-based approach to metonymy is to bolster the model of language competence with additional rules of construal (see Jackendoff, 1991), or metonymic coherence rules (see Fass, 1988). In contrast, the MBR approach does not place a dependence upon explicit *fixit* rules but instead relies upon a preponderance of performance data to provide a nearest-neighbour match. When translating to another language, the MBR system is thus not required to either resolve the metonymy in advance, or determine whether the metonymy will hold in the target language; it is enough (and indeed, from the point of preserving the style of the original statement, often preferable) to retrieve a similar usage from memory and modify the corresponding translation.

3

By choosing to emphasise linguistic performance over competence, an MBR treatment of language is essentially statistical rather than semantic, the aspiration being that given enough performance data and a sufficiently sophisticated stochastic model, an implicit model of competence will emerge. As argued above, this often is indeed the case, but it is vital to note here that such implicit competence is inherently dependent upon the *chunk* size employed by the statistical model. As the sentences processed by a translation system will be of a varying (and indefinite) size, the past examples stored in memory will have to be partitioned, or chunked, at some level of detail if the obvious combinatorial nightmare is to be avoided, and the utility of the example-base is to be maximised. Consider the following example:

(1) Neville Mariner played Mozart in a concert tour that has gathered great reviews

This sentence is simply too long to expect the case base to contain a complete match. More realistically, the sentence needs to be chunked into several semantically viable components, each of which can be matched and modified independently. But while an MBR model demonstrates a semantic competence comparable to a rule-based system on an *intra*-chunk basis, its reliance upon a *reduction to statistics* philosophy leaves it poorly positioned to reason at an *inter*-chunk level. Consider the following sentence, which is isomorphic with (1) above:

(2) Tom Hulce played Mozart in a film that has gathered many awards

In example (2) the verb "play" is used in quite a different sense to that employed in (1) however, unless the system possesses (and utilises) prior knowledge that Tom Hulce is an actor, this distinction can only become apparent at an inter-chunk level of processing. Ultimately, without a deep conceptual analysis of each individual chunk, performed against a strong conceptual model of the world, a translation system is unable to adequately characterise how different chunks inter-relate and combine to produce a coherent whole. Because of a reliance on statistics over semantics, an MBR system is unable to meet either of these requirements.

4   MBR & Planning

Before focusing on the role MBR has to play in planning it is worth considering some of the early work on the notion of "plan reuse". The basic idea in MBR of the replacement of a domain model with a extensive but shallow example-base is in direct contrast with the lessons that have been learnt from this early research. One of the first instances of plan-reuse was the work of Fikes, Hart, and Nilsson, (1972). To improve planning efficiency, the STRIPS planner was extended to facilitate the storage and retrieval of plan segments (Macrops). The nature of these stored segments is comparable to the type of examples stored in an MBR system's memory in that they were shallow packets of plan exemplars. It was soon recognised however that there were inherent problems with retrieving a suitable Macrop. Without a complex description language, Macrops were selected on the basis of surface features resulting in unforeseen instantiation problems due to bad interactions. As we shall see, we can expect similar problems within the MBR framework.

While there is a general acceptance of the advantages of "reuse" for many planning tasks, we will argue that it has the effect of redistributing rather than reducing domain knowledge requirements. In a reuse-based planning system a domain model is required to fulfil not only the needs of the adaptation task (which is similar in nature to traditional first-principles planning) but also the needs of the retrieval task.

We will consider these issues using examples taken from Déjà Vu, a case-based reasoning system for the design of Plant-Control software.[1] The problem domain of Déjà Vu has already been introduced in (Smyth & Cunningham, 1992) and so will only be described in outline here. Plant-Control software is concerned with the control of autonomous vehicles within a factory environment. An important class of Plant-Control tasks is aimed at the control of vehicles during the loading and unloading of metal coils in a steel mill. Déjà Vu's cases are software modules for controlling vehicles and other devices performing such tasks. An example of the type of code is shown in the Solution section in Figure 2.

---

[1] This software design task is essentially a planning problem.

Advance*B1*Spool	
Indices	
Machine:	Buggy*1
Content:	Spool
Action:	Advance
Speed:	2-Speed
Direction:	Forward
Source:	Skid
Dest.:	TR
:	:

Solution

Position Check — Stop(TR position) — Decelerate (200mm before TR)

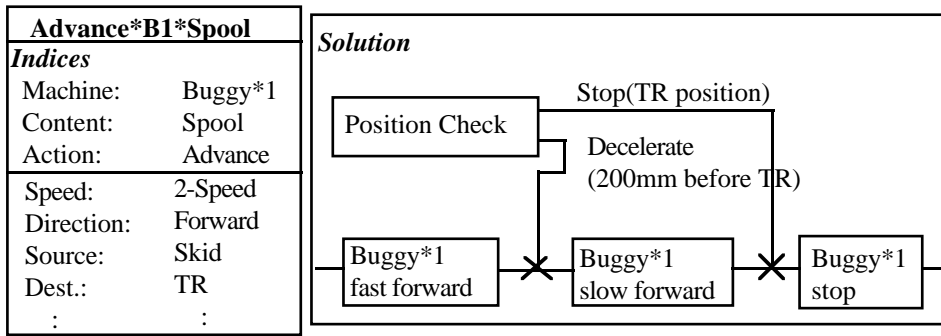Buggy*1 fast forward ✕ Buggy*1 slow forward ✕ Buggy*1 stop

Figure 2. An example case from Déjà Vu.

The code is expressed in this network representation that is compilable into and executable form. This sample case controls the movement of a buggy carrying an empty spool. Buggy*1 is a two speed buggy, so stopping is a two stage process with the buggy switching to its slower speed 200mm from its destination.[2]

4.1 Retrieval

It seems clear that if MBR systems are to enjoy their stated advantage of ease of set-up then the stored examples must be readily described in terms of shallow, easily observable, surface features. The analytical expense of deep thematic feature descriptors outlaws their use in a pure MBR system. However, surface features are not sufficiently powerful to capture the causal dependencies that exist between plan steps. Consequently, the retrieval mechanism has the potential to generate *false remindings*. That is, it is likely that example plans will be selected whose measurable similarity does not accurately reflect their *true* similarity to the target specification. In computing an accurate measurement of plan similarity one must consider not only surface features (such as the goals), but also deeper causal features that capture dependency configurations within plan structures. The result is the retrieval of a base example which exhibits strong surface commonalties to the target but is at best, difficult to adapt and at worst completely unsuitable.

For example, in Figure 2 we have presented the two-speed motion scenario where a buggy travels from one floor position to another. A behaviourally very similar task is that of controlling the motion of a lifter (a speed controlled lifting platform). This lifter is used to adjust the height of the load for loading and unloading. Like the buggy, the lifter can be a one or two speed device and a component case for a two speed lifter is shown in Figure 3. It can be seen that the solution for the two speed lifter has the same structure as the buggy case described above so it should be possible to reuse the solution from one in designing the other. However, this is problematic because these two cases do not share important surface features. So if this lifter case were a target case, the useful Advance*B1*Spool case may not be retrieved using similarity based on surface features.
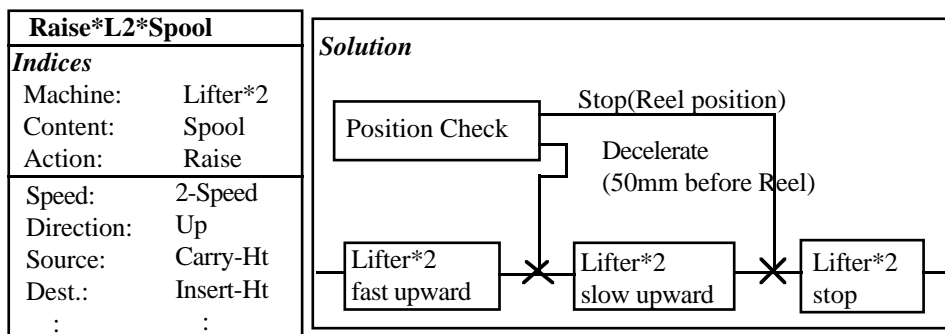
Raise*L2*Spool	
Indices	
Machine:	Lifter*2
Content:	Spool
Action:	Raise
Speed:	2-Speed
Direction:	Up
Source:	Carry-Ht
Dest.:	Insert-Ht
:	:

Solution

Position Check — Stop(Reel position) — Decelerate (50mm before Reel)

Lifter*2 fast upward ✕ Lifter*2 slow upward ✕ Lifter*2 stop

Figure 3. A two-speed lifter case.

---

[2] It should be noted that Déjà Vu uses a hierarchical reasoning technique and cases are arranged into partonomies that represent complex software designs. As such, the case here is a sub-component of a number of more sophisticated designs.

The current example from Déjà Vu has attempted to describe two cases in terms of surface features, indicating the ACTION, MACHINE, and CONTENT of the cases. The problem is that the observed difference in ACTION types means that the cases are superficially different, even though both behaviourally and structurally the cases are very similar.

It is difficult to see how the MBR paradigm could cope with such problems without moving towards more knowledge intensive and computationally expensive similarity-based retrieval models; for example, using abstract features or feature transformation techniques.

4.2 Adaptation

The shortcomings of pure MBR do not stop with retrieval. Probably the most compelling and obvious argument against the shallow knowledge ideal of MBR has to do with the post-retrieval task; an example has been retrieved, it differs from the target in terms of a number of criteria, how can it can be adapted to provide the desired target solution? Without an elaborate "fixing" of representation it should be apparent that this adaptation process will not be adequate in an MBR system tackling a complex task.

Without domain knowledge of any form, adaptation is driven solely on the basis of the mappings generated during the matching stage of retrieval. These mappings are nothing more than correspondences between "similar" base and target features. Differences are captured in the form of base and target features that have failed to be associated with a matching partner. The limited inferencing power of this similarity/dissimilarity information can be used only to guide a straightforward substitution of target features with their matching base features. The hope presumably being that such a superficial modification of the base example will yield the desired target solution. This will of course only work in simple domains.

Let us consider the problem of designing a single speed motion module in Déjà Vu as specified by Figure 4. The retrieval of the Advance*B1*Spool case means that the adaptation process must transform the two-speed structure into a single speed structure. The essential transformation is the removal of the "fast forward" speed control and the "deceleration" check.
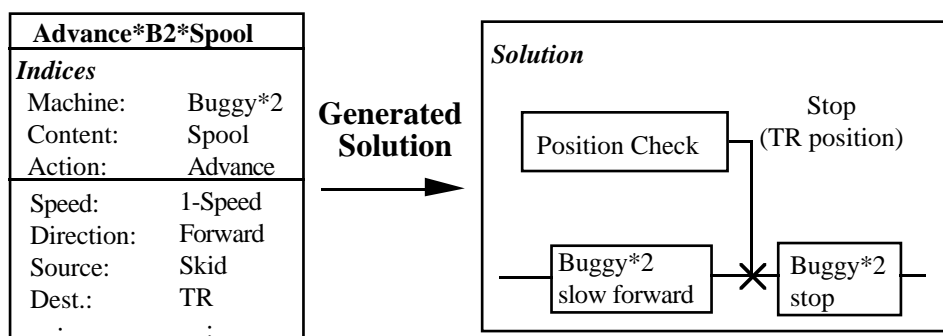

Figure 4. Designing a single speed motion solution.

This sort of modification is impossible without having domain knowledge available during the adaptation process. This is further illustrated with many more complex examples from Déjà Vu which rely of far more sophisticated adaptation; for example, arbitrary speed changes may effect a vehicle's fuel supply or particular routes while working for one vehicle may not work for another, due to size restrictions perhaps.

5   Conclusion

It this paper we have presented some examples of tasks for which we believe the pure MBR methodology will not be appropriate. For complex tasks in NLP and in planning an MBR system will not always be able to retrieve useful examples from memory nor will it be able to adapt solutions that are retrieved. In the first case, abstract features and domain knowledge are required to assess similarity. In the second case, a model of causal interactions is required in manipulating retrieved solutions.

These problems are not manifest in MBR systems presented to date because the comparatively simple domains chosen do not manifest the complex inter-dependencies that occur in more complex tasks.

Despite these criticisms we believe that MBR (or CBR systems without much domain knowledge) will have a significant impact in AI applications. For many tasks MBR does have tremendous advantages over traditional

AI techniques—primarily because it does overcome many knowledge engineering problems. However MBR on its own is not an architecture for general intelligence. The idea of reasoning from examples stored in memory is valid but for more complex tasks it needs to be augmented with domain knowledge—like in CBR.

References

Creecy R.H., Masand B.M., Smith S.J., Waltz D.L., (1992) "Trading MIPS and Memory for Knowledge Engineering: Automatic Classification of Census Returns on a Massively Parallel Supercomputer", *Communications of the ACM,* pp48-64, Vol. 35, No. 8, August 1992.

Evett M.P., Andersen W.A., Hendler J. A., (1993) "Massively Parallel Support for Efficient Knowledge Representation", In *Proceedings of IJCAI-93.*

Fass D., (1988). "An Account of Coherence, Semantic Relations, Metonymy, and Lexical Ambiguity Resolution", *Lexical Ambiguity Resolution: Perspectives from Psycholinguistics Neuropsychology and Artificial Intelligence* Small, S. I, G. W. Cottrell & M. K. Tanenhaus, eds., Morgan Kaufmann, San Mateo, CA.

Fikes, R. E., Hart, P. E., Nilsson, N. J. (1972). Learning and Executing Generalised Robot Plans, *Artificial Intelligence, 2*, 251-288

Jackendoff, R. (1991). Parts and Boundaries, *Lexical and Conceptual Semantics*, B. Levin and S. Pinker, eds. Elsevier Science Publishers, Amsterdam.

Kitano H., (1993) "Challenges of Massive Parallelism", Proceedings of the 13th. International Joint Conference on Artificial Intelligence, Chambery France, Morgan Kaufmann, pp813-834.

Kitano H., Higuchi T., (1991) "Massively Parallel Memory-Based Parsing", Proceedings of the 12th. International Joint Conference on Artificial Intelligence, Sydney Australia, Morgan Kaufmann, pp918-924.

Kolodner J.L, (1991), "Improving Human Decision Making Through Case-Based Decision Aiding", *AI Magazine*, Vol. 12, No. 2, Summer 1991, pp52-68.

Stanfill C., Waltz D., (1992) "Statistical Methods, Artificial Intelligence, and Information Retrieval", in *Text-Based Intelligent Systems*, P. Jacobs ed., pp215-225, Lawrence Earlbaum Associates, Hillsdale, New Jersey.

Stanfill C., Waltz D., (1986) "Toward Memory-Based Reasoning", *Communications of the ACM,* pp1213-1228, Vol. 29, No. 12, December 1986.

Smyth B., Cunningham P., (1992) "Déjà Vu: A Hierarchical Case-Based Reasoning System for Software Design", in *Proceedings of 10th. European Conference on Artificial Intelligence,* Vienna, Austria, ed. Bernd Neumann, Wiley & Son, pp587-589, 1992.

Veloso M., Carbonell J.G., (1991), "Learning by analogical replay in PRODIGY: first results", *European Working Session on Learning,* Y. Kodratoff, ed., pp375-389, Porto, Portugal, Springer Verlag.

Veale T., Keane M.T., (1992), "Conceptual Scaffolding: A Spatially Founded Meaning Representation for Metaphor Comprehension", *Computational Intelligence, 8,* 494-519.