ELSEVIER

Cognitive Science 27 (2003) 259–283

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Intrinsic cognitive models

Jonathan A. Waskan*

Department of Philosophy, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received 11 July 2001; received in revised form 17 May 2002; accepted 19 November 2002

Abstract

Theories concerning the structure, or format, of mental representation should (1) be formulated in mechanistic, rather than metaphorical terms; (2) do justice to several philosophical intuitions about mental representation; and (3) explain the human capacity to predict the consequences of worldly alterations (i.e., to think before we act). The hypothesis that thinking involves the application of syntax-sensitive inference rules to syntactically structured mental representations has been said to satisfy all three conditions. An alternative hypothesis is that thinking requires the construction and manipulation of the cognitive equivalent of scale models. A reading of this hypothesis is provided that satisfies condition (1) and which, even though it may not fully satisfy condition (2), turns out (in light of the frame problem) to be the only known way to satisfy condition (3).

Keywords: Philosophy; Artificial intelligence; Psychology; Representation; Philosophy of mind; Philosophy of computation; Causal reasoning; Knowledge representation; Computer simulation

1. Introduction[1]

The long tradition of philosophical discussion concerning the structure, or format, of mental representations has historically been framed in terms of a pair of competing metaphors: the logic metaphor (Boole, 1854/1951; Kant, 1787/1998; Leibniz, 1705/1997) and the picture, or image, metaphor (Aristotle, 4th Century B.C., 1987; Berkeley, 1710/1982; Locke, 1690/1964). The logic metaphor seems to explain the human capacity to think before we act, it does justice to several intuitions concerning the nature of mental states, and, with the advent of the modern programmable computer, it has become possible to give a more literal (i.e., mechanistic) reading

* Tel.: +1-217-244-2657/1090; fax: +1-217-244-8355.
E-mail address: waskan@uiuc.edu (J.A. Waskan).

0364-0213/03/$ – see front matter
doi:10.1016/S0364-0213(02)00119-2

of the hypothesis that human thought processes are effected by a 'mental logic' (Rips, 1983) or 'language of thought' (Fodor, 1975). The image metaphor has, on the other hand, been criticized for its inability to satisfy our intuitions regarding the nature of mental states, and, worse still, no one has been able to provide an acceptable non-metaphorical reading of this proposal. While a discussion of the former concern will play an instrumental role in what follows, it is the goal of this essay to remedy the latter. Indeed, here I show that it is possible to provide a non-metaphorical reading of not only the image metaphor, but of the scale model metaphor as well. I show, moreover, that (*pace* Block, 1981, 1990; Fodor, 2000; Pylyshyn, 1984; Sterelny, 1990) certain computational systems harbor representations which can be distinguished from sentential representations in the precise fashion that actual images and scale models are, that these representations exhibit immunity to the frame problem, and that, as such, they constitute our only viable computational models of forethought.

2. The mental logic hypothesis

One way of effecting the transition from explanatory metaphor to explanatory mechanism is to show that there exist, or might exist, physical systems that are similar to the system being investigated, that embody the chief characteristics of the explanatory metaphor, and that, thereby, inherit its desirable features, and perhaps its undesirable ones as well. Just such a mechanistic reformulation has, of course, already taken place with respect to the logic metaphor. The following overview of this process will clarify some of the distinguishing characteristics of sentential representations, and it will provide a template that might be followed when attempting the mechanistic reformulation of other explanatory metaphors.

2.1. Virtues and drawbacks of the mental logic hypothesis

It is often claimed that the cardinal virtue of the mental logic (ML) hypothesis is that it can explain—some even contend that it is the only way to explain—the truth-preserving character of thought sequences and, thereby, the human capacity to think before we act (Devitt & Sterelny, 1987; Fodor, 1975, 1987, 2000; Pylyshyn, 1984). The matter of forethought is of central import because it is implicated in what may very well be a distinguishing feature of human behavior: humans, perhaps unlike any other terrestrial creatures (Povinelli, 2000), seem capable of behaving in an appropriate manner in the face of even novel environmental conditions. While its ability to explain truth-preservation is widely regarded as the cardinal virtue of the ML hypothesis, the hypothesis is also lauded for its ability to explain several other putative features of human thought processes. It is claimed, for instance, that the ML hypothesis explains the productivity and systematicity of thought (Fodor, 1987; Fodor & Pylyshyn, 1988), our ability to understand words denoting that which has normative import (e.g., war criminal, ownership, etc.) or which otherwise resists depiction (e.g., economic inflation) (Fodor, 1975; Fodor, Fodor, & Garrett, 1975), our ability to think about genera (e.g., not only about specific triangles, but about triangles in general), and our ability to think about very specific states of affairs (e.g., the color of Fred's car) (Fodor, 1981).

Although the ML hypothesis seems to explain a great deal, it also has at least one major shortcoming—namely, the frame problem. McCarthy and Hayes (1969) are generally credited

with first recognizing (and naming) the frame problem, which has to do with the challenge of getting a representational system to predict what will change and what will stay the same following alterations to the state of the world (Bechtel, Abrahamsen, & Graham, 1998). A quite general way to characterize the nature of the problem confronting the ML hypothesis—which first came to light following early attempts to model forethought with the help of formalisms not unlike like those of predicate calculus (PC)—would be to say that, although the postulation of a mental logic seems to do a reasonable job of accounting for *representational* productivity (i.e., the capacity to represent countless distinct states of affairs), it does not account for *inferential* productivity (i.e., the capacity to predict the consequences of countless distinct alterations to a represented system).

The frame problem is actually comprised of at least two component problems. The first of these, the prediction problem (Janlert, 1996), stems from the fact that an immense, probably infinite (Congdon & Laird, 1997) number of inference rules, or *frame axioms*, would be required in order to effect the predictive inferences that underwrite everyday planning. The prediction problem, which is sufficiently worrisome by itself, is actually compounded by the other component of the frame problem, the qualification problem (McCarthy, 1986). It is compounded because, in order to embody what we know about the consequences of alterations to the world, not only would an infinite number of rules be required, but each rule would also have to be qualified in endless ways.

2.2. Mechanistic reformulation of the mental logic hypothesis

A watershed event in the history of the ML hypothesis was its maturation—thanks to the advent of the modern programmable computer—from an explanatory metaphor into an explanatory mechanism. Once there existed other mechanisms whose activities could, at a high (viz., algorithmic) level, be explained, quite literally, in terms of syntactically-structured representations and syntax-sensitive inference rules, it was a relatively straightforward affair to advance beyond the mere logic metaphor to the much stronger claim that thought is *literally* effected by a mental logic. Of particular relevance was the existence of computational systems (e.g., production systems and semantic networks) whose representations and rules embodied the principle tenets of the logic metaphor and, thereby, inherited its desirable features, not to mention its shortcomings. Because such systems literally engage in the relevant sort of processing, one could reasonably claim that brains—which are also characterized by a complex circuitry and fail to outwardly evidence the application of syntax-sensitive inference rules to syntactically-structured representations—might, at an equally high level, best be described as literally engaging in this sort of processing as well. Despite the shortcomings of the ML hypothesis, this mechanistic reformulation is an achievement which proponents of non-sentential cognitive images and models have hitherto been unable to match.

3. The scale model metaphor

Images and scale models fall under the more general heading of physically isomorphic models (PIMs), which are representations that possess some of the very same properties as that

which they represent (Palmer, 1978). Because forethought often requires the truth-preserving manipulation of representations of three-dimensional spatial and causal relationships, the PIMs that hold the most interest in the present context are scale models.[2] Like the ML hypothesis, the scale model metaphor has its own distinctive set of advantages and disadvantages.

3.1. Representational and inferential productivity

In order to understand how the scale model metaphor can account for representational productivity, we need to turn our attention from models themselves to the modeling media from which they are constructed. When we do, we see that there clearly are productive (or at least quasi-productive) media for the creation of scale models. A finite supply of Lego blocks can, for instance, be utilized in order to model virtually any edifice. There are, of course, many other modeling media that exhibit representational productivity (e.g., matchsticks and glue, clay, and papier-mâché). The world is, in fact, its own modeling medium.

Although the scale model metaphor for mental representation has been largely overlooked since the early successes in AI, the (somewhat less impressive) picture metaphor has lately reclaimed the attention of philosophers, psychologists, and computational modelers. Of particular interest is the fact that spatial representations can be used to generate predictions in a manner that obviates the need for rules (i.e., frame axioms) that specify the consequences of each possible alteration to a represented system (Haugeland, 1987; Janlert, 1996; Johnson-Laird, 1988; Lindsay, 1988). One might, for example, use a sheet of graph paper to represent the relative positions of Harry, Laura, and Carlene. Should one then wish to know what the relative locations of all of these individuals would be if Harry moved to a new position, one can simply delete the mark representing Harry and insert a new mark in the square corresponding to the new position.

Two-dimensional spatial media can also be used to represent the structure of objects, and collections of such representations can be used to predict the consequences of changes in their relative location and orientation. For instance, a cardboard cutout of my coffee table (as seen from above) can be conjoined with two-dimensional representations (of equal scale) of the rest of the items in my living room and a depiction of the room itself in order to predict the consequences of countless changes in the relative spatial locations and orientations of these items. A highly desirable feature of representations of this sort is that side effects of alterations to the representation mirror the side effects of alterations to the represented systems *automatically* (Haugeland, 1987)—without, that is to say, requiring their explicit specification. As a consequence, such representations exhibit, at least with regard to a limited set of dimensions, immunity to the prediction problem. They can, moreover, easily be scaled up to include representations of further objects. Systems that rely upon frame axioms seem, on the other hand, to have real problems with this kind of scalability because incremental additions to the represented system have an exponential effect on the number of rules that need to be built into the representing system. This fact, to which we will return below, leads Janlert (1996) to suggest that scalability might provide an indicator of whether or not a representational system suffers from the frame problem.

Mere images do fall a bit short of the mark when it comes to supporting the kinds of predictions that humans make on a fairly routine basis. A viable model of human forethought

must, that is to say, explain the capacity to predict the consequences of *causal* alterations in *three* spatial dimensions. With the scale-model metaphor for mental representation, these demands are easily met.

Scale models have, of course, long been a mainstay of design testing. Much like the representations underwriting forethought, scale models are used to predict the behavior of countless systems, both familiar and novel (e.g., new structures, devices, manufacturing processes, etc.).[3] Like two-dimensional spatial representations, albeit for a much wider range of represented properties, scale models do not suffer from the prediction problem. Nor do incremental additions to the represented system have an exponential effect on what needs to be built into the representation; Janlert's (1996) scalability condition is satisfied. Nor, for that matter, do scale models suffer from the qualification problem. This is because much of what is true of a modeled domain will be true of a scale model of that domain. That is to say, just like our own predictions, the predictions generated through the use of scale models are implicitly qualified in an open-ended number of ways, and so these qualifications need not be made explicit.

3.2. The frame problem: diagnosis

It has been said that the reason why PC-style representations suffer from the frame problem while images and scale models do not is that the former are extrinsic representations while the latter are intrinsic (Palmer, 1978; also see Haselager, 1997; Haugeland, 1987; Janlert, 1996). The intrinsic/extrinsic distinction was first introduced by Palmer, who offered it as a way of distinguishing between types of representation. Representations are said to be extrinsic, according to Palmer, when they must be arbitrarily constrained in order to respect the non-arbitrary, or inherent, constraints characterizing a given represented domain, while representations are intrinsic when they do not need to be arbitrarily constrained in order to respect (i.e., they inherently respect) the non-arbitrary, or inherent, constraints characterizing a represented domain. According to this analysis, the use of PC to predict the behavior of physical systems generally yields extrinsic representations. Scale models, on the other hand, constitute intrinsic representations because they do not require the imposition of arbitrary constraints in order to preserve truth with respect to what they represent.

The problem with this way of distinguishing between representational formats is that it relies too heavily upon the notion of inherent versus arbitrary constraints. Notice, for instance, that if one is merely interested in truth-preservation with regard to the taller-than relation, then, with only minor modifications, a version of PC can be devised—let us call it PC+—which is perfectly suitable. PC+, in other words, would not have to be arbitrarily constrained, and so, according to Palmer's analysis, the formulas of PC+ should be considered intrinsic representations of relative height. Yet if the presence of a few simple axioms makes PC+ intrinsic, then so much the worse for the intrinsic/extrinsic distinction. It does not clarify the difference between logic representations and scale models, nor does it supply a diagnosis for why the former seem to suffer from the frame problem while the latter do not.[4] Nevertheless, it seems clear that there is some difference between the way frame axiom systems and scale models support truth preservation and that this difference—the *real* intrinsic/extrinsic distinction—has something to do with their relative susceptibility to the frame problem.

Rather than relying upon the notion of arbitrary versus inherent constraints, we would be better served by distinguishing between logic representations and scale models in terms of whether or not they support predictions concerning particular alteration/consequence pairs on the basis of distinct data structures or, relatedly, in terms of whether or not the consequences of each type of alteration need to made explicit. In order to generate predictions concerning the consequences of particular alterations, the traditional frame axiom approach is to utilize inference rules whose antecedents specify the starting conditions and nature of the alteration and whose consequents specify the myriad consequences of the alterations. In other words, the frame axiom approach mandates that the information be made explicit. It is for this reason that the frame axiom approach suffers from the frame problem. After all, in order to embody what the average human knows about the consequences of worldly alterations, a frame axiom system would have to contain distinct rules specifying how each of countless objects, both familiar and novel, will behave relative to one another following each of a consequently infinite number of possible alterations.

In the case of scale models, on the other hand, no separate data structures are required in order to predict how particular objects will behave relative to other objects in light of particular alterations. With a suitable model of the relevant system in hand, the consequences of countless distinct alterations to the representation will automatically mirror the consequences of the countless alterations to the represented system. In other words, all of the relevant information is implicit in the representation and thus need not be made explicit. Moreover, unlike frame axiom systems, scale models scale up gracefully because when a scale model is augmented with a new item, the new representation that results will also implicitly contain all of the information needed to predict the consequences of alterations to the new system. By the same token, the utility of the approach is not restricted to individual systems that contain finite numbers of objects. With the scale modeling approach, there is no need for an antecedent and explicit specification of how each of countless objects, both familiar and novel, will behave relative to one another following each of a seemingly infinite number of possible alterations. The relevant information—as the use of scale models in design testing illustrates—will be implicit in the models that we construct. There is, in fact, a bit more to the story. As explained in Section 4.2, the fact the information is implicit in these representations is itself a by-product of the use of a representational medium that is primitively constrained in certain respects. For the moment, however, I shall simply highlight some further salient characteristics of both scale models and systems beset by the frame problem.

It is, to start with, worth bearing in mind that when it comes to predicting the behavior of physical systems (even simple ones), it will generally not suffice to utilize intrinsic representations of each property in isolation. After all, notes Palmer (1978), worldly constraints are generally interdependent, so in most cases what one requires is an intrinsic representation of complex *inter-dimensional constraints*. This, of course, is just what PIMs supply.

It is also worth noting that the atomic constituents of frame axioms are traditionally viewed as standing in something very close to a one-to-one correspondence relation with the terms comprising the corresponding natural language descriptions (Haselager, 1998). In fact, the susceptibility of frame axiom systems to the frame problem seems to be just one illustration of the limited inferential capabilities of systems that base their predictions on generalizations regarding particular objects (or object types) and relationships (or relationship types). A similar

problem crops up, for instance, in the case of associationistic models of forethought like those of Hobbes and later Empiricists.[5] As Leibniz (1705/1997) notes in his critique of associationistic psychology, statistical generalizations might lead you to expect that one kind of event will follow another, but since they don't tell you why, they are of little use when it comes to predicting the effects of other alterations to the same system or for anticipating exceptions to an observed regularity. Not surprisingly, Leibniz' critique applies to standard back-propagation networks as well. The problem, as Clark (1993) puts it, is that first-order connectionist systems seem unable to learn how to deal sensibly with structure-transforming generalizations. Imagine, for instance, a connectionist system that has learned (and can thereby predict) that a bucket containing a ball will fall from atop a pushed door. This bit of knowledge would be of little use to the system if it were asked to determine whether or not a bucket can be used to carry a ball through a door. In order to make this prediction, a new set of statistical regularities must be picked up on with a new, though probably overlapping, set of weights. In other words, what might reasonably be construed as a new data structure is required, for the requisite information is not implicit in the earlier set of weights. As such, feed-forward connectionist systems seem to suffer from the frame problem, at least when they are used to pick up on coarse-grained regularities concerning the consequences of alterations to items like buckets, balls, doors, etc.

The grain of analysis is, however, only part of the problem. To see why, notice that a mere appeal to microfeatures will not alleviate the frame problem for either frame axiom systems or feed-forward connectionist systems. A mere microfeature encoding of the parts of an object will, for instance, fail to capture information about the relative spatial arrangement of parts and the relationships that distinct objects bear to one another (Barsalou & Hale, 1993). While this information can be made explicit in the form of further features, the price is, once again, a failure to exhibit scalability (St. John & McClelland, 1990).

3.3. Further virtues and limitations

If one wishes to expand the scale model metaphor into something more than an account of forethought—that is, if one wishes to develop it into a univocal account of mental representation—one finds cause for both optimism and concern. The best way to see this is by comparing and contrasting the relative benefits and drawbacks of the logic and scale model metaphors.

To start with, while the logic metaphor seems to offer a plausible account of systematicity, it is not the only such account. To see why, one simply has to note that the world itself admits of certain systematic variations (e.g., not only can the cat be on the mat, but the mat can be on the cat). Instead of pushing the structure of language 'down', so to speak, into the thought medium, proponents of the scale model metaphor have every reason to complain that an equally viable account of systematicity can be supplied by pushing the structure of the world 'up'. To be sure, according to both the mental logic and scale model metaphors, the systematically related representations are made of the same parts, but the parallels between the two explanations of systematicity seem to end there.

There are other properties of human thought processes that are not so easily explained by the scale model metaphor (Fodor, 1981, 1987; Pylyshyn, 1984). For starters, as proponents of the ML hypothesis are fond of pointing out, it is less than obvious that the image metaphor—and the same applies to the scale model metaphor—will be able to explain our ability to think

about what might be called *non-concrete* domains (e.g., war criminal, ownership, and economic inflation). Likewise, in and of themselves, images and scale models seem ill-equipped to represent either genera or specifics (i.e., to single out particular properties of particular objects).

At least for present purposes, it is worth conceding that, on its own, the scale model metaphor has problems satisfying these intuitions about mentality, while the logic metaphor does not.[6] What I have so far been advocating, however, is not a scale model metaphor for mental representation *tout court*, but for a circumscribed class of *cognitive* representation. What this theory provides is the best and, in light of the frame problem, perhaps the *only* satisfactory account of the kind of truth-preserving representational manipulations that underwrite planning. Proponents of the ML hypothesis have thus greatly overstated their case with the claim that the techniques of formal logic provide the only known means of capturing the norms of reasoning (Fodor, 1987, 2000; Pylyshyn, 1984). This constitutes a significant beachhead for the scale model metaphor, for the (overstated) ability of the ML hypothesis to explain truth-preservation has always been *its* biggest draw. The ground that has been gained on behalf of the scale model metaphor can, however, only be held if a mechanistic reformulation of the proposal is forthcoming.

4. Mechanistic reformulation of image and scale model metaphors

Though the scale model metaphor provides a compelling explanation of forethought, nothing has yet been said by way of rendering intelligible the claim that a system like the brain might actually harbor representations of the appropriate sort. Indeed, while the ML hypothesis allowed for a straightforward mechanistic reformulation, it is, in the case of the scale model metaphor, less clear how one ought to proceed.

The problems facing the scale model metaphor are just the same well-worn problems that have long confronted the image metaphor. Past attempts to supply a more literal reading of the image metaphor have, specifically, generally failed for one of two reasons: they have either (a) failed to maintain compatibility with basic brain facts; or (b) failed to support a distinction between sentential and imagistic representations. For instance, while the claim that the brain harbors representations that are physically isomorphic with what they represent suffers from problem (a), attempts to supply a mechanistic reformulation of the image metaphor that closely parallels the mechanistic reformulation of the ML hypothesis have been said to suffer—and necessarily so—from problem (b). A version of the latter tack is opted for below, so a bit of elaboration is in order.

4.1. Computational mechanisms and levels of description

On the face of things, it would seem to be worth considering whether or not there are computational systems that can do for the image and scale model metaphors what production systems and semantic networks did for the logic metaphor. The persistent worry about this approach, however, is that all computational systems are driven by the application of syntax-sensitive inference rules to syntactically structured representations. Moreover, when a computational system is used to support predictions concerning how the world will change, it seems reasonable

to construe such rules as data structures used to predict—as explicit specifications of—the consequences of alterations. In short, any such computational system arguably relies upon extrinsic sentential representations (Block, 1981, 1990; Fodor, 2000; Pylyshyn, 1984; Sterelny, 1990). In and of itself, this would seem to bar a mechanistic reformulation of the image metaphor—and, by extension, the scale model metaphor—analogous to the one effected with regard to the logic metaphor. What is even more worrisome is that if the brain itself turns out to be a computational system, then the possibility of *any* such reformulation would seem to be ruled out. After all, claims Fodor (2000), "if . . . you propose to co-opt Turing's account of the nature of computation for use in a cognitive psychology of thought, you will have to assume that *thoughts themselves have syntactic structure*" (p. 13).

For all their intuitive appeal, these concerns are misguided. After all, even if the brain turns out, at some level, to implement the application of syntax-sensitive inference rules to syntactically-structured representations, this will no more entail that thoughts are sentential than does the binary nature of the processes typically used to implement frame axioms entail that frame axioms are binary. In fact, contrary to popular wisdom, there are good reasons for thinking that, at a high level of description, certain computational systems (e.g., the models of imagery proposed by Kosslyn, 1980 and Glasgow & Papadias, 1992) do harbor representations that are non-sentential and, indeed, imagistic.

In order to make good on this claim, we first need to co-opt some of the conceptual apparatus long wielded by proponents of the ML hypothesis. In particular, one thing that all proponents of this hypothesis agree upon is that the descriptions of cognitive processing they offer are pitched at a very high level. The level that most interests ML theorists is, moreover, "distinguished" by the fact that one finds, at that level, factual and counterfactual representations of the environment (e.g., the constituents of the molecular formulas represent various objects, properties, and relationships); while at the next level down one finds a specification of the primitive properties of the implementation base (Pylyshyn, 1984, p. 95). Proponents of the image and scale model metaphors can utilize these same insights when supplying a mechanistic reformulation of these hypotheses.

Notice, to start with, that when the same criterion of levels individuation (i.e., multiple realizability) is employed, we find that there are at least two levels at which a scale model can be described: the level of the modeling medium (i.e., the implementation base) and the level of the models themselves. That is to say, if we take a given model *type* to subsume those token models that respect a particular set of inter-dimensional worldly constraints, we see that there will generally be multiple modeling media that can be used to implement a given model type. For instance, one such model type is the sort that can be used to predict the consequences of various three-dimensional spatial alterations to the items in my living room. Each such model will constitute an intrinsic representation of the complex inter-dimensional constraints imposed by shape, size, orientation, and location, and this type can obviously be implemented on the basis of any of a variety of materials. There is, in other words, a multiple-realizability relationship between model types and the various media that can be used to implement them, and so descriptions of models can be said to be pitched at a higher level than descriptions of their implementing media. Moreover, the difference between the two sorts of description is that the higher-level descriptions are 'distinguished' by the fact that one finds, at that level, factual and counterfactual representations of the environment (i.e., representations of various

objects, properties, and relationships), while at the next level down one finds a specification of the primitive properties of the particular implementation base (e.g., constraints governing the manner in which Lego blocks can be conjoined).

When a representational system can be described at any of multiple, independent levels, the possibility opens up—indeed, the case for the mechanical realizability of the ML hypothesis depends on this—that there may be properties invoked in the context of implementation-level descriptions that need not, and should not, be invoked when supplying higher-level descriptions. This opens up at least the bare possibility of representational systems that are best described at an implementation level in terms of extrinsic, sentential representations but which, at a higher level, are best be understood in terms of intrinsic, non-sentential representations.[7] To see why this is more than a bare possibility, let us start by revisiting the argument that computational matrix representations (CMRs) of the sort devised by Kosslyn (1980) are necessarily extrinsic.

4.2. Intrinsic computational images

To be sure, the claim that the implementation base (i.e., the representational medium) for CMRs typically involves extrinsic representations is not without merit. This is because, unlike 'real' spatial matrices, the construction of computational matrices necessitates the imposition of processing constraints through a reliance upon rules governing, for instance, the use of memory registers (Kosslyn, 1980; Pylyshyn, 1984). Such rules might even be construed as extrinsic representations of the consequences of alterations, though it bears mentioning that the alteration/consequence pairs in this case concern changes in the coordinates of particular cell contents.

Although the implementation base for CMRs is arguably sentential and extrinsic, CMRs themselves constitute intrinsic representations of inter-dimensional constraints. To see why, notice, for starters, that once a medium for the construction and manipulation of CMRs has been created by imposing the relevant processing constraints, the representations constructed from the 'materials' supplied by such a medium exhibit immunity to the frame problem, at least with regard to certain two-dimensional spatial relationships. In other words, as Pylyshyn (1984) points out, with regard to two-dimensional spatial relationships, the consequences of alterations to the representation automatically mirror the consequences of the corresponding alteration to the represented system (p. 103). As it turns out, the effect is not restricted to two spatial dimensions or to simple changes in relative location. Representational media have, for instance, been created (Glasgow & Papadias, 1992)—again, arguably through a reliance upon extrinsic representations—which supply an implementation base for representations that exhibit inferential productivity with respect to three-dimensional alterations in both the location and orientation of a seemingly endless number of objects.[8] Such systems are, in fact, precisely the sort that Janlert (1996) seems to have had in mind when he suggested that the solution to the frame problem might be a kind of 'mental clay'.

In the case of CMRs, there is no need to incorporate distinct data structures specifying how each of countless distinct objects will behave relative to one another following each of the consequently infinite number of possible alterations to their relative location and/or orientation. In other words, in the case of CMRs, the information is implicit in the representation and so need not be made explicit. As such, the consequences of alterations to the representations

automatically mirror the consequences of the corresponding alterations to the world. Thus, although a description of the medium would involve talk of the rule-governed imposition of processing constraints (e.g., constraints on the use of memory registers), the representations implemented by that medium are, like the scale model of my living room described above, intrinsic representations of the complex inter-dimensional constraints imposed by the relative shape, size, orientation, and location of objects.

For his part, Pylyshyn (1984) seems happy to concede that the relevant information "is implicit in the data structure" (p. 103). He is, however, reticent to call CMRs intrinsic, reserving that term instead to refer to the primitive constraints governing a representation's implementation base (i.e., properties of the functional architecture of some real or virtual machine or the primitive properties of some formal notation). Yet, as Sterelny admonishes, "It obviously does *not* follow from the fact that a representational system is primitive that it is intrinsic: English could be hard-wired into my brain, but it is a paradigm of a non-intrinsic system" (Sterelny, 1990, p. 623). In fact, a given functional architecture may itself be nothing more than a program (e.g., a Java virtual machine) run on some other kind of machine. It is, therefore, hard to imagine what useful notion of 'intrinsic' Pylyshyn might have in mind when he claims that the primitive properties of a notation or virtual machine are intrinsic.

It is not at the level of the primitive operations of an implementation base that we find intrinsic representations, but at the level of the representations realized by a given, primitively constrained implementation base. Part of what licenses this claim is the fact that certain constraints will be inviolable *at the representation* level—and, relatedly, the fact that a great deal of information will be implicit—given that the representations have been realized through reliance upon a particular implementation base. As Mark Bickhard puts it (in correspondence), "Properties and regularities are only going to be 'intrinsic' at one level of description if they are built-in in the realizing level—or else they are ontologically 'built-in' as in the case of strictly spatial relationships in physical scale models." While scale models are intrinsic for the latter reason, CMRs are intrinsic for the former. Building certain constraints in at the implementation level—the level of the medium—has, in the case of CMRs, the effect of guaranteeing that the representations realized by that medium will respect complex inter-dimensional constraints. As such, the consequences of many types of alteration follow automatically and so need not be specified explicitly.

Thus, as in the case of scale models, we find that there are at least two levels of description applicable when talking about CMRs: the level of CMRs themselves and the level of the implementation base. Moreover, not only do we find intrinsic representations at the former level and (arguably) extrinsic representations at the latter, but, just as frame axioms "are intended to represent something quite different from the expressions at the implementation . . . level" (Pylyshyn, 1984, p. 94), so too are CMRs intended to represent something quite different from the expressions at the implementation level. That is to say, the former are 'distinguished' by the fact that they are representations of objects and their relationships, while the latter represent such things as the numerical coordinates of filled and empty cells and the constraints governing the manner in which the coordinates of cell contents are permitted to change. In fact, at the representation level we find that a given representation (e.g., a representation of the structure of my coffee table) transcends any particular set of sentences found at the implementation level; after all, coordinates can and do change without changing the (non-relational) properties of the representation itself.

4.3. Intrinsic computational models

For those interested in modeling forethought, CMRs are a step in the right direction, but a full-scale solution to the frame problem—that is, the sort that can account for the kind of inferential productivity that underlies the effective behavior exhibited by humans in the face of various environmental contingencies—requires intrinsic representations of interacting three-dimensional spatial and causal constraints. Computational systems that harbor representations of this sort can be found in sectors of computer science that seem, as yet, a bit far removed from cognitive science proper. Specifically, virtual reality models (VRMs), devised primarily for entertainment purposes, and finite element models (FEMs), devised for engineering purposes, constitute intrinsic computational representations of interacting three-dimensional spatial and causal constraints.

4.3.1. Virtual reality models: Ray Dream 5.02

Much like CMRs, VRMs generally involve coordinate specifications (viz., in an *x*, *y*, *z* coordinate system) for primitive modeling elements. Rather than the filled and empty cells of a matrix, however, the coins of the realm in VR modeling are two-dimensional polygons. Coordinate specifications are given for the vertices of polygons, and the surfaces of objects are represented in terms of the collective arrangement of (usually) many polygons—thus forming a productive representational medium known as *polymesh* (Watt, 1993). While much VR modeling research has been focused on the interactions amongst the surface features of objects and various sorts of illumination, VR modeling media have also been created that are capable of predicting how countless objects, both familiar and novel, will behave relative to one another following each of a seemingly infinite number of possible alterations. In other words, VR modeling media have been created that can be used to generate representations that exhibit immunity to the frame problem with regard to three-dimensional spatial alterations and a wide range of causal interactions. To demonstrate that VRMs exhibit this degree of immunity to the frame problem, a set of models was created using an off-the-shelf program called Ray Dream Studio 5.02.

4.3.1.1. Representational and inferential productivity. The first model is based on a problem that most humans have little trouble solving, but which, somewhat surprisingly, chimps seem to find rather challenging (Povinelli, 2000). The goal is to pick the implement which, when pulled, will bring the banana within reach (Fig. 1). If the scale model metaphor for forethought is correct, humans construct the cognitive equivalent of a scale model of the problem and use this model to predict the consequences of pulling on each of the implements. To show that VRMs, like scale models, exhibit the requisite powers of truth preservation, a model of the problem was created and the locations of each of the implements was made to change over time—that is, each was moved from the back of the table to the front. One would hope to find that the difference between moving the toothless rake and moving the T-bar (inverted rake) is that in the latter case the banana comes along for the ride. This, in fact, is precisely what transpired. Importantly, in this case the outcomes of these alterations to the representation mirrored what would happen in light of the corresponding alterations to the represented system, and without requiring any rules framed with respect to the properties of bananas, toothless rakes, or T-bars.

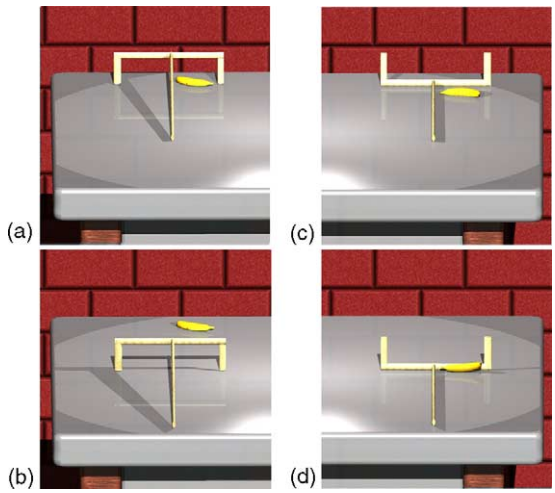

Fig. 1. Truth preservation: effects of moving implements. Virtual toothless rake and banana (a) before move and (b) after move. Virtual inverted rake and banana (c) before move and (d) after move.

That is to say, the VRM exhibited powers of truth-preservation similar to those exhibited by scale models. Of course, if VRMs truly have the same predictive powers as scale models, they will exhibit immunity to the frame problem.

If you will recall, one facet of the frame problem is the qualification problem. Unlike frame axiom representations, scale models do not suffer from the qualification problem because the predictions they license are implicitly qualified in countless ways. For instance, pulling on a scale model of the T-bar will cause a scale model of the banana to move within reach, provided that, among other things, there is not a hole in the scale model of the table. VRMs also implicitly admit of such qualifications. To demonstrate that this is so, the model just described was altered in one simple respect: a hole was put in the table between the T-bar and the opening to the enclosure. Once again, the results were highly promising. Instead of the banana being carried along to the edge of the container, it fell through the hole (Fig. 2). Like our own predictions and the predictions generated through the use of scale models, predictions generated on the basis of VRMs are implicitly qualified.

The other main facet of the frame problem is the prediction problem. Scale models are immune to this affliction, as are CMRs (at least when it comes to predicting the consequences of changes in spatial relationships). VRMs mark a major advance over the CMRs considered above in that they exhibit immunity to the prediction problem with regard to both three-dimensional spatial and causal relationships. To show this, a model of the door, bucket, ball system (described in Section 3.2) was constructed and altered in various ways. The starting condition for the first alteration has the bucket resting atop the door and the ball positioned over the bucket. The only direct manipulation to the ensuing chain of events is that the door is opened rather abruptly. What we should like to find in this case is that the bucket and the ball fall to the floor, and this is exactly what transpired (Fig. 3). As in the case of the previous model, we find that the side effects followed automatically, and without requiring any rules framed with respect to the properties of doors, buckets, and balls.

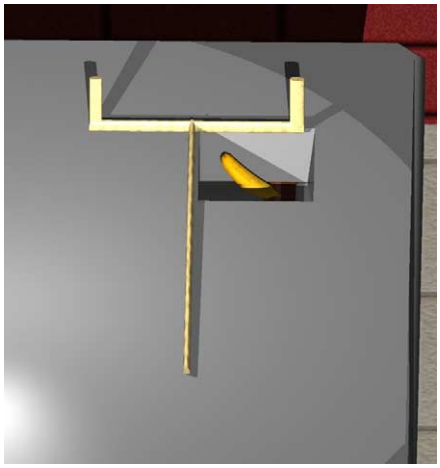

Fig. 2. Qualified predictions: banana falls through hole in table.

In a new scenario, the bucket is turned upside-down and placed over the ball. The bucket is then moved through the doorway and is subsequently raised. Were this alteration carried out with respect to either the actual door-bucket-ball setup or a scale model of this setup, we should expect to find the ball (or the scale model of the ball) underneath the bucket. This is also what we find in the case of the VRM (Fig. 4).

While it would be a straightforward matter to show that this same model can be used to predict the consequences of any number of further alterations, another way to demonstrate that VRMs are immune to the frame problem is to show that they satisfy the scalability criterion. If you will recall, Janlert (1996) suggested that a system not beset by the frame problem admits—unlike frame axiom representations—of incremental additions to the represented system without having an exponential effect in terms of what needs to be added to the representation. Scale models (not to mention CMRs) satisfy this scalability criterion, and so do VRMs (more will be said below concerning *why* this is so). As a simple demonstration of scalability, a board was added

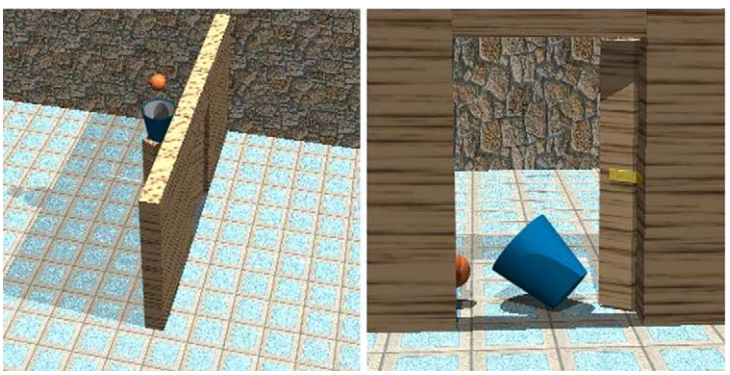

Fig. 3. Predicting alterations: effects of opening door, (left) starting position of items and (right) position of items after door supporting bucket/ball opened.

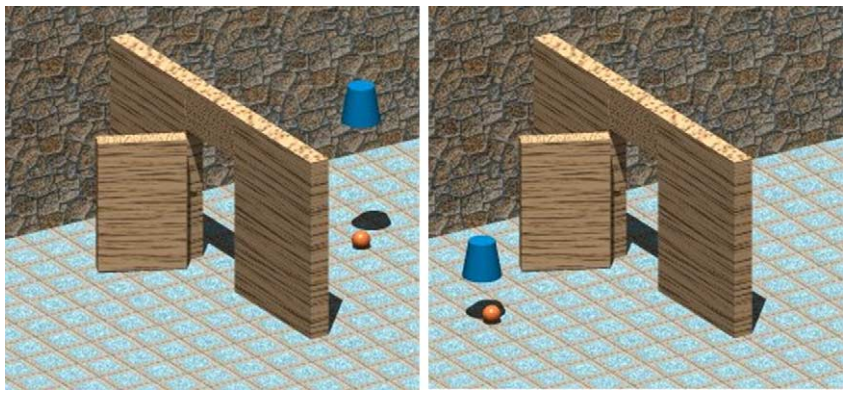

Fig. 4. Predicting alterations: using bucket to move ball through doorway, (left) before and (right) after.

to the door-bucket-ball model. To ensure that its presence was relevant, the board was placed broadside across the doorway (on the same side of the wall as the ball and bucket) and the bucket was used to throw the ball on a relatively low trajectory through the doorway. Once again, what we would expect to happen in both the world and a scale model thereof took place in the VRM—that is, the ball bounced off of the board instead of rolling through the door (Fig. 5).

As you can see, Ray Dream models exhibit an impressive degree of immunity to the frame problem. Just as in the case of scale models, a VRM that represents the structure and relative sizes of the objects comprising some system can be used to predict the consequences of countless alterations to that system. Moreover, just as in the case of scale models, there is no need to incorporate separate data structures corresponding to each alteration/consequence pair. These models do not, for instance, rely upon rules specifying that a ball inside of an upright bucket will tend to move wherever the bucket moves; that a bucket set atop a pushed door will

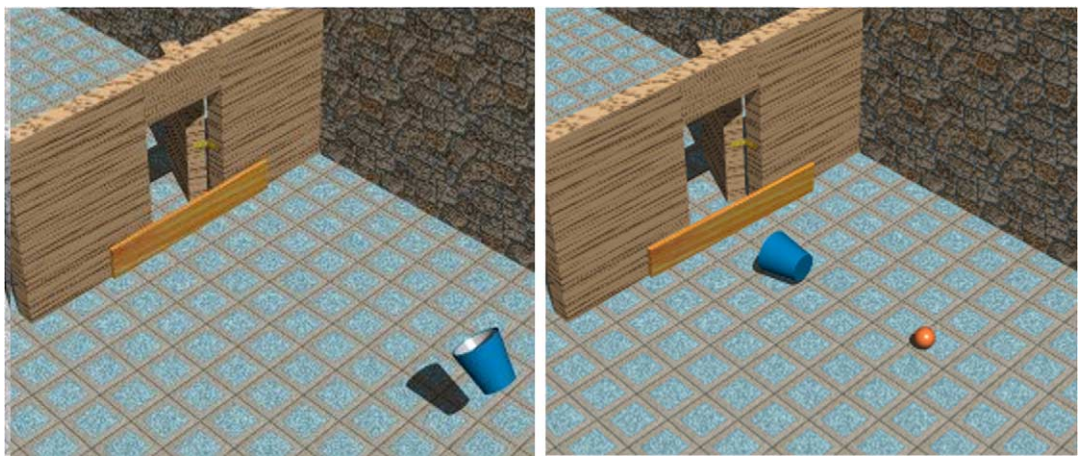

Fig. 5. Satisfying Janlert's (1996) scalability criterion: a board is added to the model and the bucket is used to throw ball toward doorway, (left) before and (right) after.

tend to fall to the floor; that T-bars can be used to drag bananas, and so forth. Instead, like scale models, the side effects of alterations to the VRM automatically mirror the side effects of alterations to the represented system. VRMs implicitly contain all of the relevant information, so it need not be made explicit. In other words, while there is a case to be made that the medium used to implement VRMs involves extrinsic representations, the VRMs themselves constitute intrinsic representations of complex inter-dimensional constraints. For this reason, it seems not unreasonable to view VRMs as supplying a promising model of forethought. There are, however, a few complications worth mentioning.

4.3.1.2. The psychological plausibility of VRMs. To start with, one expects that when two objects are stacked atop one another in a positive gravity environment they will eventually come to rest. With Ray Dream models, however, when objects are placed atop one another they never settle entirely.[9] Instead there is always a small degree of oscillation. This was particularly apparent when the bucket containing the ball was set atop the door. The two objects never quite settled, and when they interacted on the way to the floor their motions sometimes appeared a bit jerky and artificial. It should be borne in mind, however, that in every trial the outcome of pushing the door open was basically the same: both the ball and the bucket fell on the floor.

Another worry about the Ray Dream models is that the outcomes of collisions are not determined by such factors as mass, momentum, or degree of rigidity/springiness. For instance, the simple bouncing behavior of the ball in the second model was not a consequence of a specification of such underlying factors as the storage and release of energy due to compression. Instead, there is a primitive rebound setting that determines how bouncy the ball is. Although this may seem like a shortcoming, there is (somewhat surprisingly) a case to be made that something similar occurs when physics-naïve individuals predict the outcomes of collisions.[10] For instance, in a seminal study conducted by Chi, Glaser and Rees (1982), novices and experts were asked to categorize a set of physics problems. Novices were found to categorize problems on the basis of their surface features, while experts categorized them on the basis of the underlying physical principles they exemplified. Even more to the point, DiSessa (1983) examined the manner in which physics-naïve individuals understand the nature of bouncing behavior, and the results were similar. DiSessa discovered, for instance, that one physics-naïve individual, M., lacked an accurate understanding of the underlying basis for bouncing behavior. This property seemed, for M., to be a primitive that she discovered through experience and in terms of which she subsequently explained and predicted the behavior of objects in the world. In fact, even physics experts were found in certain cases to rely upon high-level primitives that effectively simulate the consequences of low-level principles.

This, again, is not unlike how the Ray Dream modeling medium supports predictions regarding physical interactions. Objects in the model do not undergo compression, though the primitive constraints of the medium guarantee that they behave in many ways as if they did. As a result, these models (like those harbored by physics-naïve individuals) do a reasonable job of generating the kinds of predictions required in order to respond appropriately in the face of various environmental contingencies—for example, those that include T-bars, bananas, buckets, doors, balls, and so on.[11]

Construed as a model of human forethought, then, Ray Dream seems consistent with empirical findings concerning how humans predict the behavior of physical systems. Indeed,

according to the picture of forethought painted by naïve physics researchers, individuals create cognitive models of the world and 'run' these models in order to predict and explain the behavior of various physical systems (Chi et al., 1982; De Kleer & Brown, 1983; DiSessa, 1983; Larkin, 1983; Norman, 1983; Schwartz, 1999). There is, moreover, widespread support amongst such researchers for the claim that these internal models of the world are non-sentential.

On the other hand, while the VRMs implemented by the Ray Dream modeling medium do enjoy a certain amount of psychological plausibility, the fact that objects are invariably represented as rigid means that the predictive powers of these VRMs are limited in ways that seem psychologically unrealistic. To be sure, many behaviors that result from the deformable nature of certain bodies (e.g., when balls bounce, fly off on a particular trajectory when struck, or are wedged in the bottom of the bucket) can be simulated with Ray Dream, but many others (e.g., what happens when an inflatable ball is stabbed with a knife) cannot be. There are, however, other computational modeling media that harbor representations—called FEMs—of the appropriate sort.

4.3.2. FEMs

For decades, the methods of finite element modeling have been under development for use in the engineering disciplines (for an overview, see Adams & Askenazi, 1999). Like VRMs, FEMs are constructed from polymesh—that is, objects are represented in terms of a number of polygons (called elements) whose vertices (called nodes) are specified in terms of their coordinates. One major difference between FEMs and Ray Dream models is that the relative positions of the nodes comprising an object are not fixed in the former, but can change in ways that enable one to model the behavior of deformable bodies.

To provide a simple illustration of finite element modeling techniques, a two-dimensional finite element model of a fairly thin sheet of material (which is fixed in place by four supports at its base) was constructed with a program called PlastFEM (Fig. 6). As with other FEMs, how the material behaves under various conditions can be modeled by simulating the application of forces to specific nodes and running the model in order to see how the consequences play out. For example, how the sheet behaves in light of a causal interaction with a sharp object was modeled by applying a force to a single node. Likewise, how the sheet holds up in the face of a collision, of equal total magnitude, with a blunt object (in this case from a different direction) was determined by distributing the same total force across a larger set of nodes.

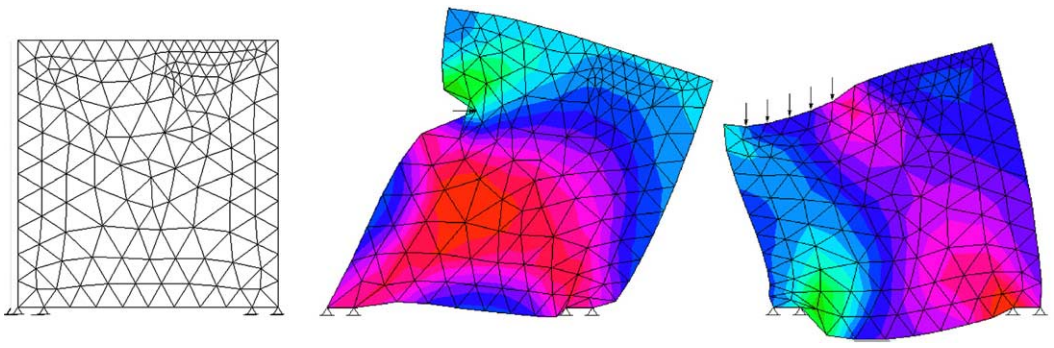

Fig. 6. Modelling the effects of loads applied to a sheet of material. Arrows indicate forces applied to nodes.

While the models created with PlastFEM exhibit inferential productivity with regard to the effects of loads applied to two-dimensional deformable bodies (of any shape), the same basic techniques can be scaled up in order to model the effects of collisions between, and numerous other factors affecting the behavior of, three-dimensional deformable bodies. For instance, acceleration and rotation can be modeled by distributing forces, in the appropriate directions, to some or all of the nodes comprising a model, ambient pressure can be modeled in terms of a load applied to the entire surface of an object, and the effects of thermal expansion and contraction can also be modeled by applying forces to nodes (Barton & Rajan, 2000). Nor does the power of finite element modeling method end there. As Barton and Rajan (2000) explain, this "[o]ne method can solve a wide variety of problems, including problems in solid mechanics, fluid mechanics, heat transfer and acoustics, to name a few." Moreover, the tricks for modeling each sort of problem have been integrated into general-purpose modeling systems such as MSC.visualNastran and LS-DYNA. Not surprisingly, such systems are widely used in the testing of prototypes for airbags, circuit breakers, pyrotechnic devices, and countless other novel mechanisms.[12] They are also used in order to determine whether or not particular theories—concerning, among other things, spinal chords, neurons, and tectonic plates—actually explain observed phenomena. FEMs can be used, in other words, in order to predict the consequences of countless alterations to countless novel systems.

As with scale models and VRMs, there is no need in the case of FEMs to incorporate separate data structures that represent the consequences of each possible alteration to a set of items. Instead, like scale models, the side effects of alterations to the FEM will automatically mirror the side effects of alterations to the represented system. Because all of the relevant information is implicit in the FEMs that we construct, it need not be made explicit. FEMs, in other words, constitute intrinsic representations of the complex inter-dimensional constraints imposed by size, shape, location, orientation, velocity, and numerous physical forces. FEMs, in short, exhibit full-blown immunity to the frame problem.

One potential worry about FEMs, however, is that—due in part to a tremendous short-term memory capacity and in part to the fact that the principles built into their realization bases are inspired by our best scientific characterizations of the principles underlying the behavior of macro-sized objects—the predictions generated through the use of FEMs are often far more accurate than those made 'in-the-head' by humans. From a psychological modeling standpoint, then, it may be an advantage of Ray Dream models that they rely upon inaccurate, though oftentimes useful, characterizations of the physical principles underlying the behavior of everyday objects. An advantage of FEMs, on the other hand, is that—due to the fact that the relative positions of nodes comprising the representation of an object are not fixed—they support predictions concerning the behavior of deformable bodies. The truth with regard to human forethought may, therefore, ultimately lie somewhere in-between Ray Dream models and FEMs.

5. The intrinsic cognitive models hypothesis

As noted earlier, a mechanistic reformulation of the logic metaphor was supplied by showing that there exist computational systems that embody, at a high level of description, the central

characteristics of the metaphor and, thereby, inherit its explanatory virtues as well as its limitations. A similar set of claims can now be made with regard to the image and scale model metaphors.

As in the case of images and scale models, in the case of CMRs, VRMs and FEMs there is a distinction to be made between the images and models themselves and the media from which they are constructed, and talk of the former is pitched at a higher level than talk of the latter. To be sure, the medium used to implement a given computational image or model can be described in terms of syntactically structured representations and syntax-sensitive inference rules. In the case of VRMs and FEMs, such representations and rules specify the coordinates of polygon vertices (nodes) and impose constraints on the manner in which the coordinates of nodes are permitted to change. As such, it may even be appropriate to describe these media as relying upon extrinsic, sentential representations, though it bears emphasizing that the representations at issue are mathematical formalisms whose variables take on continuous numerical values and which bear only a superficial resemblance to traditional frame axioms.

Unlike the frame axiom approach and like the scale modeling approach, the methods of virtual reality modeling and finite element modeling offer a tractable solution to the frame problem because they pair down the basic ontology to a comparatively simple set of building blocks and permissible building block behaviors. Instead of explicitly specifying how objects will behave relative to one another in light of countless possible alterations, a primitively constrained computational modeling medium can be used to construct representations of the objects of interest, and the consequences of alterations to these representations will automatically mirror the consequences of the corresponding alterations to the represented system. Just like scale models, these representations implicitly contain all of the information needed to predict the consequences of countless alterations to their represented systems, so the information need not be represented explicitly with the help of countless distinct data structures. Nor is the utility of the approach restricted to individual systems that contain finite numbers of objects. That is, unlike the frame axiom approach, the computational modeling approach described here does not necessitate an antecedent and explicit specification of how each of countless objects will behave relative to one another in light of the consequently infinite number of possible alterations. Instead, just as in the case of scale models, the relevant information will be implicit in the models that we construct. It is for this reason that VRMs and FEMs satisfy Janlert's (1996) scalability criterion.

One common reaction to these claims is to suggest alternative strategies for dealing with the frame problem. From where I sit, however, the proof is in the pudding. What I am offering here is not mere speculation concerning how the frame problem *might be* solved, but rather a diagnosis for how it *has been* solved. This, predictably enough, gives rise to another common reaction, which is to deny that the frame problem has been solved at all. It has, in fact, been suggested that even scale models suffer from frame problem. Claims of this sort are, it would seem, taken to be justified by the fact that certain questions have been left unanswered. For instance, it seems clear that merely having a high-fidelity model of some system in hand provides little or no indication of just what alterations (of the tremendous number possible) are, in light of a particular goal, worth considering. Nor, for that matter, has a mechanical procedure been supplied that can enable the determination of just how many (or, more appropriately, how few—see footnote 3) of a system's properties should be represented in any given

instance. Worries of this sort have, however, very little to do with the frame problem (i.e., the prediction-plus-qualification problem) that has long beset ML-inspired approaches to modeling the truth-preserving machinery that (*in part*) underwrites human planning. Indeed, such indiscriminate use of 'frame problem' (e.g., to refer to any and all difficulties encountered when attempting to explain or model human planning) threatens to strip the term of all theoretical interest. It is, by the same token, asking far too much of *any* mere model of forethought that it solve all of the problems having to do with human planning. In order that the proper perspective on these issues might be maintained, I would like to suggest that the following consideration be kept in mind whenever arguments against the present hypothesis are entertained: humans clearly possess whatever cognitive resources are required in order to construct, manipulate, and read predictions off of external models. In light of this simple fact, it becomes hard to imaging what *a priori* justification could exist for claiming that humans would be incapable of constructing, manipulating, or reading predictions off of *internal* models.

To return to the matter at hand, although it may be appropriate to describe the media for the construction of VRMs and FEMs in terms of extrinsic sentential representations, the representations implemented by such media are—like scale models and unlike frame axiom representations—intrinsic representations of complex inter-dimensional constraints. We can make sense of this claim because, just as in the case of scale models, descriptions of these computational models are pitched at a higher level than are the descriptions of their respective media. It is, moreover, at the higher, 'distinguished' level of description (Section 4.1), the level of the models themselves, that we find representations of objects and their myriad relationships. At the lower, implementation level, on the other hand, we find extrinsic representations that specify nodal coordinates and impose constraints on the manner in which those coordinates are permitted to change.

An added, and no less important, finding is that the computational systems considered here exhibit, in addition to representational and inferential productivity, virtually all of the features that distinguish images and scale models from sentential representations. To retrace the points covered in Section 3.3, we find that, like scale models, computational models provide an elegant account of systematicity. This is because, like scale models, computational representations such as VRMs and FEMs admit of certain systematic variations. And while the systematicity exhibited by VRMs and FEMs can be explained in terms of the rearrangement of parts, the manner of their rearrangement seems no more sentential than the rearrangement of the parts of a scale model.

Likewise, in terms of their capacity to supply a univocal model of mental representation, models such as CMRs, VRMs and FEMs face the very same set of challenges that the image and scale model metaphors faced. For instance, as was the case with the image and scale model metaphors, problems arise for CMRs, VRMs and FEMs when we consider their suitability for representing non-concrete states of affairs. Moreover, the same line of reasoning that led to the conclusion that thoughts are unlike images and scale models in that the latter are capable of representing both genera and specifics applies with equal force to *computational* images and models.[13,14]

The fact that the precise limitations of the image and scale model metaphors are inherited by these computational systems would seem to present a rather uncomfortable dilemma for die-hard proponents of the ML hypothesis. After all, many such individuals contend that

non-sentential representations such as images and scale models are distinct from sentential representation in that the former, but not the latter, are (by themselves) limited in their ability to represent non-concrete domains, genera, and the assignment of particular properties to particular objects. At the same time, however, these very individuals (e.g., Fodor, 2000; Pylyshyn, 1984; Sterelny, 1990) claim that computational systems only harbor sentential representations. It would seem that one of these two claims must be abandoned. As the intuitions behind the former are difficult to counteract, and the intuitions behind the latter have here been undermined; it appears that the appropriate course of action is to abandon the claim that computational systems harbor only sentential representations and concede that, at a high level of description, some computational systems harbor non-sentential images and models.

To parallel the arguments made by proponents of the ML hypothesis, it can now be claimed that while brains, like computers, are characterized by a complex circuitry and fail to outwardly evidence the harboring or manipulation of non-sentential intrinsic models of complex inter-dimensional worldly constraints, perhaps, at a suitably high level of description, they too can literally be said to harbor and manipulate such representations. Unlike past attempts to mechanistically reformulate the image and scale model metaphors (Section 4), this intrinsic cognitive models (ICM) hypothesis is both sufficiently distinct from the ML hypothesis and compatible—for the same reasons that the ML hypothesis is compatible—with basic brain facts.

There are, admittedly, some properties that have been lost in the transition from explanatory metaphor to explanatory mechanism. One is physical isomorphism. This, of course, is fortunate given that the brain surely does not harbor PIMs of doors, buckets, and balls. Another difference between PIMs and such computational models as CMRs, VRMs and FEMs has to do with the nature of constraints governing the behavior of their primitive modeling elements. While the constraints governing the behavior of physical building blocks are fixed by the laws of physics, the constraints governing the behavior of computational building blocks are primitive but not nomological. In some ways the difference is irrelevant. After all, *given* that the representations have been realized through the use of a particular primitively constrained modeling medium, there will be certain constraints on the behavior of the representations that are (as explained in Section 4.2) inviolable and, relatedly, a great deal of information will be implicit. On the other hand, in the case of PIMs, if one wishes to know how an object would behave were it made from a different material, one will (generally) need to construct an entirely new model using that other material. In the case of computational models, however, there are certain properties of the building blocks that can be modified simply by changing the values of variables in the equations describing how those building blocks behave. One can, in effect, change what an object is made of without having to construct that object anew, though the option of constructing the object anew does remain open. Presuming, then, that the ICM hypothesis is correct, it is an open question whether or not humans create representations of the world anew when, for instance, they discover that their default assumptions were incorrect.

One outstanding worry about non-sentential representations that bears further scrutiny is Pylyshyn's (1981, 1984) infamous cognitive penetrability criterion. Pylyshyn claims that if our cognitive representations of spatial and causal properties are influenced by our beliefs in logically coherent ways, then this will provide sufficient warrant for concluding that the representations involved are sentential in character—for logical coherence, argues Pylyshyn,

is only explicable through the postulation of a mental logic. For instance, suppose it turns out that our predictions concerning the behavior of the system depicted in Fig. 3 differ—and they almost certainly will—depending upon whether we are told that the ball is a volley ball (Scenario 1) or a bowling ball (Scenario 2). If we adopt the cognitive penetrability criterion, we will be forced to conclude that this difference can only be accounted for by sentential cognitive representations. Adopting this criterion, however, also commits us to the absurd claim that scale models are sentential. After all, if I construct a different scale model of the setup depicted in Fig. 3 depending upon whether I believe Scenario 1 or 2, then the predictions generated by those models will clearly be sensitive to my beliefs in logically coherent ways. This consideration, I take it, suffices to rid us of the dubious cognitive penetrability criterion.

The foregoing mechanistic reformulation of the image and scale model metaphors clearly has favorable ramifications for empirical research (both behavioral and computational) concerning the possibility that humans harbor and manipulate non-sentential cognitive representations. For the first time, the proposal has been given the kind of non-metaphorical reading required in order to achieve the same level of scientific respectability so long enjoyed by the ML hypothesis. It also provides a way to temper the concern that a sentential model can always be constructed that generates the same behavioral predictions as a non-sentential one (Anderson, 1978; Palmer, 1978). To be sure, one can always (at least after the fact) create a sentential model that generates a set of behavioral predictions, but theorists are also often interested in the broader question of whether or not a given model can explain some specified function performed by a particular component of the cognitive system. If performance of a given function (e.g., forethought) requires immunity to the frame problem, then we have good *a priori* reasons for believing that no sentential model will suffice.

In closing, let me simply remark that it should come as no surprise that some insight into the workings of the human mind should come from the consideration of VRMs and FEMs. After all, the point of creating scale models and, more recently, computational models has always been to generate predictions concerning the behavior of some target system, and a similar capacity for predictive inference may well be one of our most distinctive cognitive powers. Indeed, if the ICM hypothesis is correct, then—at a suitably high level of description—there are few relevant differences between scale models, computational models (e.g., VRMs and FEMs), and cognitive models.

Notes

1. For further information regarding this manuscript, please visit the on-line annex of *Cognitive Science*.
2. Craik (1952), Block (1990), and Janlert (1996) seem to have come closest to appreciating this point; though Craik, for his part, failed to distinguish between mere isomorphism and physical isomorphism, while Block incorrectly maintains that non-sentential cognitive models are incompatible with the computational theory of mind.
3. In the latter case (and perhaps also in the former), what one builds into one's model will depend upon what kinds of properties one is interested in tracking and what the

consequences are of failing to anticipate the relevance of an untracked property. For instance, if one is interested in the optimal arrangement of items in one's living room, a two-dimensional model may well suffice; and the consequences of failing to anticipate the relevance of a property (e.g., height) amount to only a minor inconvenience. If one is interested in testing the design of a new type of spacecraft, on the other hand, it makes a good deal more sense to model the system, as is commonly done, down to the last detail, for the unforeseen importance of any given property might have dire consequences. When constructing a scale model, then, the important question is not how much should be built into a model, but rather how much one can afford to leave out.

4. Thanks to Mark Bickhard for pointing out the shortcomings of this version of the intrinsic/extrinsic distinction.
5. Dennett (1988) makes a related point.
6. If the scale model metaphor is to have any hope of overcoming these limitations, additional cognitive resources (e.g., capacities for mapping between different representations, attentional selection, pattern recognition, etc.) will surely have to be invoked.
7. Johnson-Laird (1983) is one of the few theorists to have recognized that the existence of distinct levels of computational description might be relevant to the discussion of computational models of imagery.
8. In such cases, the representational medium is constituted by ordered memory registers and control processes.
9. This is the case, specifically, when objects are assigned physical properties through the 'apply physical effects' command.
10. Though this might seem less surprising if the designers of Ray Dream happen to be physics-naïve. It is more likely, however, they are physics savvy and recognize a computation-sparing shortcut when they see one.
11. Although Hayes (1995) has famously suggested that AI researchers incorporate the principles of naïve-physics in their models, the present model-based approach seems at rather at odds with his prescription of an expert-systems style axiomatization.
12. For some illustrations of this point, see: http://www.sri.com/poulter/fe_modeling/impacts.html, http://www.ncac.gwu.edu/archives/animation/index.html, and http://www.csm.ornl.gov/viz.html.
13. There seems to be an unavoidable degree of specificity exhibited by scale models and VRMs that gives rise to both of these concerns. Interestingly enough, this unavoidable specificity may be a necessary accompaniment to avoidance of the frame problem. Stenning and Oberlander (1995) make a similar claim, and even contend that this kind of specificity suffices to distinguish imagistic from sentential representations. By itself, however, a mere appeal to specificity does not provide a sufficient basis for distinguishing between sentential and imagistic representations. Stenning and Oberlander contend, for instance, that a tightly constrained PC-style notation is imagistic. In the absence of a distinction between levels of description, however, critics of mental imagery can simply charge that such representations at best function like images (see also Section 3.1).
14. Just as in the case of the scale model metaphor (footnote 6), if there is to be any hope of overcoming these limitations, an (I think unproblematic) appeal to extra-representational cognitive resources will almost certainly be required.

Acknowledgments

Comments from the following individuals and groups helped to shape the thoughts expressed in this article: Bill Bechtel, Jesse Prinz, Mark Bickhard, Ron Chrisley, Andy Clark, Susan Stuart, Gary Ebbs, Patrick Maher, Dick Schacht, Mark Rollins, Dave Balota, Desiree White, Brian Keeley, Daniel Dennett, Whit Schonbein, Tad Zawidski, The PNP Program at Washington University, the philosophy departments at Ohio University in Athens, California State University in Long Beach, William Paterson University, and the University of Illinois at Urbana-Champaign. Special thanks go to the creators of Ray Dream Studio 5.02 and PlastFEM.

References

Adams, V., & Askenazi, A. (1999). *Building better products with finite element analysis*. Santa Fe, NM: OnWord Press.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*(4), 249–277.

Aristotle (4th Century B.C./1987). On the soul. In J. L. Ackrill (Ed.), *A new Aristotle reader* (pp. 161–205). Princeton, NJ: Princeton University Press.

Barsalou, L. W., & Hale, C. R. (1993). Components of conceptual representations: From feature lists to recursive frames. In I. Van Mechelen, P. Theuns, & R. Michalski (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 98–144). London: Academic Press.

Barton, M., & Rajan, S. D. (2000). Finite element primer for engineers. Available at: http://ceaspub.eas.asu.edu/structures/FiniteElementAnalysis.htm. Accessed November 13, 2002.

Bechtel, W., Abrahamsen, A., & Graham, G. (1998). The life of cognitive science. In W. Bechtel, A. Abrahamsen, & G. Graham (Eds.), *A companion to cognitive science* (pp. 1–104). Cambridge, MA: Blackwell.

Berkeley, G. (1710/1982). *A treatise concerning the principles of human knowledge*. Indianapolis: Hackett.

Block, N. (1981). Introduction: What is the issue? In N. Block (Ed.), *Imagery* (pp. 1–18). Cambridge, MA: The MIT Press.

Block, N. (1990). Mental pictures and cognitive science. In W. G. Lycan (Ed.), *Mind and cognition* (pp. 577–606). Cambridge, MA: Blackwell.

Boole, G. (1854/1951). *An investigation of the laws of thought*. New York: Dover Publications.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–76). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, A. (1993). *Associative engines*. Cambridge, MA: The MIT Press.

Congdon, C. B., & Laird, J. E. (1997). *The Soar user's manual: Version 7.0.4*. University of Michigan.

Craik, K. J. W. (1952). *The nature of explanation*. Cambridge, MA: Cambridge University Press.

De Kleer, J., & Brown, J. S. (1983). Assumptions and ambiguities in mechanistic mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 155–190). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dennett, D. (1988). *Brainchildren*. Cambridge, MA: The MIT Press.

Devitt, M., & Sterelny, K. (1987). *Language and reality: An introduction to the philosophy of language*. Cambridge, MA: The MIT Press.

DiSessa, A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 14–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fodor, J. A. (1975). *The language of thought*. New York: Thomas Y. Crowell.

Fodor, J. A. (1981). Imagistic representation. In N. Block (Ed.), *Imagery* (pp. 63–86). Cambridge, MA: The MIT Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: The MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way*. Cambridge, MA: The MIT Press.
Fodor, J. A., Fodor, J. D., & Garrett, M. F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, *6*, 515–531.
Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1/2), 3–71.
Glasgow, J., & Papadias, D. (1992). Computational imagery. *Cognitive Science*, *16*(3), 355–394.
Haselager, W. F. G. (1997). *Cognitive science and folk psychology: The right frame of mind*. London: Sage Publications.
Haselager, W. F. G. (1998). Connectionism systematicity and the frame problem. *Minds and Machines*, *8*(2), 161–179.
Haugeland, J. (1987). An overview of the frame problem. In Z. W. Pylyshyn (Ed.), *Robot's dilemma* (pp. 77–93). Norwood, NJ: Ablex Publishing.
Hayes, P. J. (1995). The second naive physics manifesto. In G. F. Luger (Ed.), *Computation and intelligence* (pp. 567–585). Menlo Park, CA: AAAI Press.
Janlert, L. (1996). The frame problem: Freedom or stability? With pictures we can have both. In K. M. Ford & Z. W. Pylyshyn (Eds.), *The robot's dilemma revisited: The frame problem in artificial intelligence* (pp. 35–48). Norwood, NJ: Ablex Publishing.
Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
Johnson-Laird, P. N. (1988). How is meaning mentally represented? In U. Eco, M. Santambrogio, & P. Violi (Eds.), *Meaning and mental representation* (pp. 99–118). Bloomington, IN: Indiana University Press.
Kant, I. (1787/1998). *Critique of pure reason* (P. Guyer & A. Wood, Trans.). Cambridge, MA: Cambridge University Press.
Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
Larkin, J. H. (1983). The role of problem representation in physics. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 75–98). Hillsdale, NJ: Lawrence Erlbaum Associates.
Leibniz, G. W. (1705/1997). *New essays on human understanding*. Cambridge, MA: Cambridge University Press.
Lindsay, R. K. (1988). Images and inference. *Cognition*, *29*(3), 229–250.
Locke, J. (1690/1964). *An essay concerning human understanding*. Oxford, UK: Clarendon Press.
McCarthy, J. (1986). Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, *28*(1), 86–116.
McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence* (pp. 463–502). Edinburgh, UK: Edinburgh University Press.
Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Hillsdale, NJ: Lawrence Erlbaum Associates.
Palmer, S. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–302). Hillsdale, NJ: Lawrence Erlbaum Associates.
Povinelli, D. J. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works*. New York: Oxford.
Pylyshyn, Z. W. (1981). The imagery debate: Analog media versus tacit knowledge. In N. Block (Ed.), *Imagery* (pp. 151–206). Cambridge, MA: The MIT Press.
Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: The MIT Press.
Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, *90*(1), 38–71.
Schwartz, D. L. (1999). Physical imagery: Kinematic versus dynamic models. *Cognitive Psychology*, *38*(3), 433–464.
Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, *19*(1), 97–140.
Sterelny, K. (1990). The imagery debate. In W. G. Lycan (Ed.), *Mind and cognition* (pp. 607–626). Cambridge, MA: Blackwell.
St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence processing. *Artificial Intelligence*, *46*(1/2), 217–257.
Watt, A. (1993). *3D computer graphics* (2nd ed.). Harlow, UK: Addison-Wesley Longman Limited.