A Virtual Solution to the Frame Problem

Jonathan A. Waskan

Philosophy Department, William Paterson University, 300 Pompton Road, Wayne, NJ
07470-2103
waskanj@wpunj.edu

Abstract. We humans often respond effectively when faced with novel circumstances. This is because we are able to predict how particular alterations to the world will play out. Philosophers, psychologists, and computational modelers have long favored an account of this process that takes its inspiration from the truth-preserving powers of formal deduction techniques. There is, however, an alternative hypothesis that is better able to account for the human capacity to predict the consequences worldly alterations. This alternative takes its inspiration from the powers of truth preservation exhibited by scale models and leads to a determinate computational solution to the frame problem.

1. Thinking Ahead

Consider the following setup:
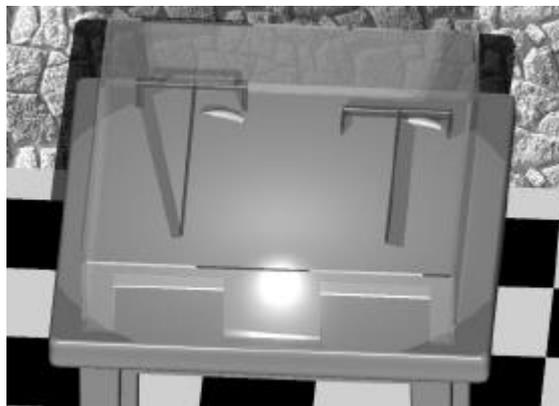
Fig. 1. A choice between two implements (a toothless rake and a T-bar) for removing food (a banana) from an enclosure. Problem based on Povinelli (1999).

I you happen to be hungry for a banana, it is obvious which implement you ought to reach for. A bit of forethought reveals that the toothless rake would pass over the banana, while pulling on the T-bar would cause the banana to move within reach.

The proposal that human behavior is frequently informed by forethought is intuitively quite plausible. That is, it certainly *seems* that we represent to ourselves the outcomes of possible actions—it certainly seems that we are capable of thinking ahead. Aristotle may have been the first to clearly formulate this proposal. As Aristotle puts it, "sometimes you calculate on the basis of images or thoughts in the

soul, as if seeing, and plan what is going to happen in relation to present affairs" (Aristotle, *De Anima III*, 431b).

Since about the middle of the twentieth century, beginning with the work of Wolfgang Köhler (1938) and Kenneth Craik (1952), this proposal has been taken to form the basis for a powerful explanatory hypothesis. What makes this hypothesis powerful is that it can explain why human behavior is often so appropriate, even in the face of even novel environmental conditions. In Köhler's terms, forethought gives one *insight* into a problem, and, he notes, "[w]hen insight refers to practical situations it seems to us itself a most practical gift, because its decisions tend to do justice to the nature of such situations . . . " (1938, p. 28-29).

There is still a good deal of controversy surrounding the question of whether or not chimps or other non-human primates are capable of insight. At some level, this controversy reflects the familiar concern that we are being too generous when we ascribe to lesser creatures the same high-level cognitive abilities that we humans seem to possess. Povinelli and colleagues (1999) have recently provided some compelling demonstrations that we ought to be leery of such ascriptions. In one experiment, they present a chimp with a setup very much like the one depicted in Fig. 1. On seeing footage of a chimp pulling ineffectively and repeatedly on the toothless rake, one is left with the distinct impression that chimps lack the same well-developed capacity for forethought that we humans seem to possess. Indeed, the marked disparity between chimps and humans in this regard suggests that the capacity to manipulate representations of the world may be one of the most profound differences between chimps and ourselves.

Be this as it may, it should at least be clear that the proposed capacity to manipulate representations is properly viewed as an hypothesis that can explain why human behavior is often so appropriate in the face of novel conditions. According to this hypothesis, we humans are, through the manipulation of representations, capable of predicting the consequences of alterations to the world. This model of the mechanisms underwriting (at least certain forms of) human behavior has several presuppositions, but it also leaves a great many questions unanswered.

1.1 Representing

One of the central presuppositions of the Köhler-Craik model is that we humans have a capacity to represent novel situations as they arise. The name given to this property of our representational system is 'productivity'. A representational system is productive to the extent that it can be used to represent an open-ended number of circumstances. For example, we can represent to ourselves the setup in Fig. 1, or, to take another example, the setup here (consisting of a doorway, a bucket, and a ball):
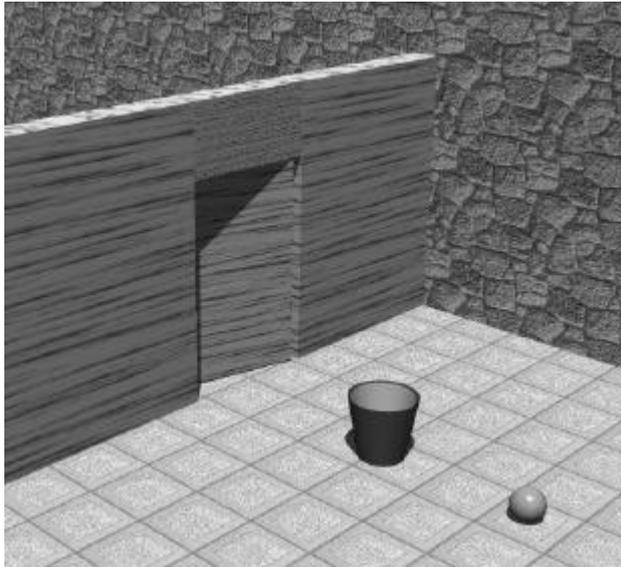
Fig. 2. A simple physical system: A doorway, a bucket, and a ball.

What is open to debate is the matter of how productivity is achieved by the human cognitive system.

One very popular proposal is that mental representations have a language-like structure. Languages are, of course, productive in the requisite sense. Natural language such as English or artificial deductive notations like Predicate Calculus can be used to represent any of an open-ended number of circumstances. Perhaps, the argument goes, the representational productivity exhibited by the human cognitive system can itself be attributed to a language-like or logic-like medium of representation.

According to this model, we humans represent the situations confronting us in working memory as a set of sentences or, more precisely, well-formed formulae. When presented with a slightly different arrangement of the items in Fig. 2, for example, the result might be a set of formulae encoded in one's short-term memory that are notational variants of the following:

```
There is a bucket, a ball, and a door.
The bucket is resting atop the door.
The ball is inside of the bucket.
```

1.2 Predicting

Also presupposed by the Köhler-Craik model is a means of using our representations to generate predictions. In order to use our representations to determine the effects of alterations to a given situation, the representations we construct must be amenable to truth-preserving manipulation. By 'truth-preserving', I mean that the consequences of alterations to our mental representations mirror the consequences of the corresponding alterations to the world.

We know, for example, that if the ball in Fig. 2 is placed inside of the bucket and the bucket is kept upright, then the location of the ball will henceforth change with that of the bucket. We also know what will happen if the ball is inside of the bucket, the bucket is set atop the door, and the door is subsequently pushed.

In terms of its capacity to explain truth-preservation, the logic metaphor for thought appears quite promising. According to the logic metaphor for thought, we represent the state of the world in short-term memory as wff, and inferences concerning the consequences of alterations to the world are made through the

application of inference rules. We might, for instance, have in our heads a notational variant of the following rule:

```
If
     the bucket is resting atop the door, and
     the ball is inside of the bucket, and
     the door is pushed
then
     the bucket and the ball will fall to the floor.
```

According to the logic metaphor for thought, we humans harbor a (presumably massive) set of such inference rules, and we draw upon these rules in order to predict the consequences of alterations to the world.

Because it can account for both representational productivity and truth-preservation, the logic metaphor for thought enjoys a strong base of support throughout cognitive science. In the field of artificial intelligence, for example, developers of production systems have traditionally adopted this framework in order to generate computational models of the processes that underwrite forethought. What is nice about the production system approach to cognitive modeling is that accords quite well with the tenets of the Köhler-Craik model. The latest versions of Soar, for example, represent the state of the world in terms of wff's held in a working memory, and inferences concerning the consequences of alterations to the world are effected by applying rules (called 'operators') to those formulae (Congdon & Laird, 1997). On this basis, the general Soar architecture can be used to create systems capable of thinking before they act.

There are also many cognitive psychologist who view the logic metaphor for thought as the best way to fill out the details of the Köhler-Craik model. Philip Johnson-Laird and Ruth Byrne (1991) have, for instance, proposed that a capacity for deductive inference is what underwrites the truth-preserving representational manipulations that give rise to our most basic planned behaviors. Likewise, Lance Rips (1990) takes an understanding of deductive competence to be necessary if we are to give "an account of how the inferences people draw manage to be truth preserving in a sufficiently large number of cases to make both science and practical affairs possible" (pp. 293). Rips, in fact, utilizes production systems in order to model the cognitive underpinnings of those inferences made by subjects under controlled conditions.

Amongst philosophers, the contention that mental representation and inference are effected by the cognitive equivalent of a formal deductive notation is known as the Language of Thought (LOT) hypothesis. Like computational modelers and psychologists, philosophers are attracted to the LOT hypothesis because of its capacity to explain both representational productivity and the truth-preserving manipulation of mental representations (Devitt and Sterelny 1987; Fodor 1975; Pylyshyn 1984).

1.3 Acting

A third presupposition of the Köhler-Craik model is that there is a means of selecting, amongst the many possible alterations to the world, a course of action that will lead to the fulfillment of one's desires. For example, for one who has the desire to move the ball from one side of the wall to the other, it is apparent that throwing the ball over the wall or using the bucket to roll it through the door will have the desired effect. There are also many alterations that will fail to bring one closer to this goal. Somehow we are able to determine which alterations to the world, of the many that are open to us, will lead to the fulfillment of our desires, and after selecting an appropriate alteration we are able to act upon it.

In contemporary production systems, operators embody knowledge of the consequences of particular alterations and the task of selecting and executing the appropriate operators is relegated to a set of productions. Operator-proposal productions determine which operators *can* be applied in a given situation. There are

obviously a great number of alterations that can be made to even a simple setup like the door, bucket, ball setup, so a production system designed for reasoning about this setup might incorporate a huge number of operators. One of the nice features of contemporary production systems is that they are able to learn which operators or sequence of operators led to a desired result under similar conditions in the past. This knowledge is incorporated into operator-comparison productions that determine which of the many operators that *can* be applied in a given situation will be likely to lead to the desired result. Finally, there are operator application productions that execute the appropriate alterations following the decision process. Execution of operators can be carried out either 'in the head' of the production system or with respect to the world itself.

1.4 The GOFAI–Mental Logic–LOT Model

What the present discussion suggests is that any viable theory concerning the nature of mental representation will have to account, minimally, for representational productivity and truth-preservation. To early AI researchers, and many since, the mental logic model of knowledge representation and inference has seemed quite promising. Again, formal notations like predicate calculus are highly productive and, through the application of the appropriate inference rules, they can be used to carry out the kind of truth-preserving representational manipulations that seem to inform our actions.

As a result of the early work in artificial intelligence (which philosophers often refer to as Good Old Fashioned Artificial Intelligence, or GOFAI)[1], many philosophers and psychologists have come to endorse the proposal that mental representation and inference are effected by the cognitive equivalent of a formal deductive notation. There are problems with this model, however. There is, moreover, an alternative model that does not suffer from these same problems.

2. Problems with the GOFAI–Mental Logic–LOT Model

Although the promise of being able to account for productivity and truth-preservation has been the main impetus behind the widespread acceptance of the logic metaphor for thought, it turns out that, at least where truth-preservation is concerned, the superficial promise of the logic metaphor is just that, superficial. Although early AI researchers were enticed by the productivity and inference powers of formal deductive notations, it did not take them long to realize that there is a serious problem with this approach. The problem with the logic metaphor for mental representation and inference has come to be known as the *frame problem* (McCarthy and Hayes 1969).

In order to guide behavior effectively, a representational system needs to be capable of tracking what will change and what will stay the same in light of particular alterations. If a logic-driven system like a production system is to accomplish this, a great deal of domain-specific knowledge will have to be encoded in the form of inference rules (e.g., production-system operators). One of the problems with this approach, the original 'frame problem' pointed out by McCarthy and Hayes, is that due to the huge number of non-changes that will need to be deduced, "the system will simply get lost in performing irrelevant deductions" (Haselager unpublished manuscript, pp. 5).

As it is conceived of today, *the* frame problem is actually comprised of a collection of problems. The general worry, and the reason for unifying these problems under a single heading, is that the kind of knowledge that we humans have concerning the consequences of alterations to the world cannot be formalized in the manner that many had initially hoped it could. Two of what I take to be the most

---

[1] I believe Haugeland (1985) can be credited with coining the phrase.

important of these problems are the qualification problem (McCarthy 1986) and the prediction problem (Janlert 1996).

2.1 The Qualification Problem

Let me start with the qualification problem. If you will recall, one of the key virtues of the logic metaphor for thought is that the consequences of alterations to a given physical system can be captured with the help of rules like this one:

```
If
     the bucket is resting atop the door, and
     the ball is inside of the bucket, and
     the door is pushed
then
     the bucket and the ball will fall to the floor.
```

Yet, as any philosopher of science will tell you, this rule does not adequately express what we know about the consequences of this particular alteration. In order to express what we know about the consequences of this alteration, the antecedent of this rule would need to be qualified in an open-ended number of ways. The rule would have to state that:

```
If
     the bucket is resting atop the door, and
     the ball is inside of the bucket, and
     the door is pushed, and
     it is not the case that the bucket is bolted to the
     top of the door, and
     it is not the case that there is a string
     connecting the bucket to the ceiling, and
     it is not the case that an atomic bomb will explode
     when the door is pushed, and
     so on . . .
then
     the bucket and the ball will fall to the floor.
```

Again, in order to capture what we know about the consequences of this alteration, the antecedent of this rule would need to be qualified in an *open-ended* number of ways. This is the qualification problem. The problem, more generally, is that our knowledge of the consequences of such alterations cannot be formalized in the manner that many had initially hoped it could. The same can be said in the case of the prediction problem.

2.2 The Prediction Problem

In order to understand the prediction problem, consider how many distinct rules would be needed in order to express what we know about the consequences of alterations to the door, bucket, ball setup (Fig. 2). For starters, there would need to be a rule specifying that if the ball is placed inside of the bucket and the bucket is kept upright, then the location of the ball will henceforth change with that of the bucket. In addition, there would need to be rules specifying what happens when the bucket is set atop the door and the ball is rolled through the door, what happens when the bucket with the ball in it is tipped over, what happens when the bucket is used to throw the ball at the door, and so on indefinitely. In order to embody what we know about the consequences of such alterations, there would have to be an open-ended number of rules. As with the qualification problem, the general problem is that our knowledge of the consequences of alterations to even a simple setup like the door, bucket, ball setup, cannot be formalized in the manner that many had initially hoped it could.

As if either problem were not worrisome enough by itself, keep in mind that the qualification problem actually compounds the prediction problem. In order to express what we know about the consequences of alterations to a given domain, not only would there have to be an open-ended number of rules, but the antecedents of each of these rules would need to be qualified in an open-ended number of ways. The frame problem is, as you can see, quite serious, and it is not surprising that books continue to be written about it.

So the point is this: The superficial promise of the logic metaphor for thought is just that, superficial. Although formal deduction techniques embody means of effecting truth-preserving representational manipulations, this is not, by itself, sufficient to account for our *open-ended knowledge* of the consequences of various alterations.

3. The Model Model

What might be considered an unfortunate consequence of the early work in AI is that it eclipsed another promising hypothesis about the nature of mental representation and inference. This hypothesis was given its first clear formulation by Kenneth Craik.

3.1 Craik's 'Hypothesis on the Nature of Human Thought'

In 1943, well before GOFAI had taken root, Kenneth Craik was struck by the powers of truth preservation embodied by devices like scale models. As Craik explains:

> If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise … and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it (Craik 1952, p. 61).

That is, rather than a logic metaphor for thought, Craik favored a scale-model metaphor. It seems clear that scale models provide an effective means of evaluating the outcome of alterations to a given system. For instance, with a scale model of the door, bucket, ball setup, one can predict what will happen when the ball is placed inside of the bucket and the bucket is set atop the door and the door is subsequently pushed.

3.2 The Productivity of Modeling Media

In order to understand how the scale model metaphor accounts 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 clearly that there are productive (or at least quasi-productive) media for the creation of scale models. Notice, for instance, that a finite supply of Lego blocks can be utilized in order to model virtually any possible edifice:

Fig. 3. A scale model of New York City constructed with Lego blocks. Image from: http://www.legolandca.com/images/postcards/thumb/kidsny.jpg

There are, of course, many other modeling media that exhibit productivity. For example, matchsticks and glue, clay, and papier-mâché can all be used productively in the service of creating representations. So, like the logic metaphor for thought, the scale model metaphor has the virtue of being able to account for both productivity and the truth-preserving manipulation of representations.

3.3 Avoiding the Prediction Problem with Spatial Images

Although the scale model metaphor for thought has been largely overlooked since the early successes in AI research, it's somewhat less impressive forerunner (i.e., the picture metaphor) has recently begun to regain the attention of philosophers, psychologists, and computational modelers. Specifically, researchers have begun to realize that spatial representations, such as pictures and maps, can be used to generate predictions in a manner that obviates the need for rules specifying the consequences of various alterations (see Haugeland 1987, Johnson-Laird 1988, Lindsay 1988, Janlert 1996). Because they do not owe their inference powers to such rules, pictures and maps can be used to predict the consequences of an open-ended number of alterations to the systems they represent. In other words, such representations do not suffer from the prediction problem.

Notice, for instance, that one can use a two-dimensional matrix like this one in order to represent the relative positions of two or more people and to predict how these relative positions will be affected when a person (or persons) changes locations:

	Alice					Betty
	Carol					

Fig. 4. Use of a spatial matrix to represent the relative positions of objects.

If a person moves from one location to another, this change can be tracked with our simple matrix and all of the relative positions will be updated automatically. For example, if Betty moves from her original location to a new one, we can keep track of this change by deleting our Betty representation from its old location and inserting it into the new one. Regardless of where we reinsert the Betty representation, the relative positions of *all* of the individuals will be automatically updated as a by-product—and without the need for rules specifying how the change in location will affect their relative positions.

This holds, moreover, for any number of individuals (provided, of course, that the matrix is comprised of enough cells to represent all of the individuals). It is worth noting that this is a property that Lars Erik Janlert has identified as an indicator of whether or not a representational system suffers from the frame problem. Janlert (1996, pp. 40) explains, "A sign that the frame problem is under proper control is that the representation can be incrementally extended: A conservative addition to the furniture of the world would involve only a conservative addition to the representation."

Not only can matrices be used to represent relative positions, but they can also be used to construct simple depictions of objects. For instance, the matrix here is being used to depict (albeit crudely) a rocket:

Fig. 5. Use of a spatial matrix to depict an object.

Once such a depiction has been constructed, it becomes possible to predict the consequences of further kinds of spatial alteration. For instance, this rocket representation can be turned around and the result will be that the nozzle is over the body of the rocket, rather than visa versa. This inference can be effected without the need for a rule specifying, for example, that if a rocket is upright and is rotated 180 degrees around a horizontal axis then the nozzle will be over the body.

Because the need for such rules is obviated, spatial matrix representations can be used to predict the consequences of an open-ended number of alterations to the systems they represent. In other words, spatial matrix representations do not suffer from the prediction problem, at least when it comes to predicting the consequences of two-dimensional spatial alterations.

The benefits of matrix representations are not restricted to two spatial dimensions, however. Recognizing this, some artificial intelligence researchers (e.g., Janlert, 1996) have begun to speculate that human thought might be effected with the help of a kind of 'mental clay'.

3.4 Avoiding the Frame Problem with Scale Models

Of course, if a theory of mental representation is to account for the truth-preserving powers that underwrite our ability to deal effectively with novel circumstances, such a theory will have to handle a full range of three-dimensional spatial *and* causal inferences. It would be nice, in addition, to know that the theory was not susceptible to either the prediction problem or the qualification problem.

Returning to the full-blown scale-model metaphor for thought, we see that these demands are easily met. One can, for instance, use a reasonably faithful scale model of the door, bucket, ball setup to predict the consequences of an open-ended number alterations to the actual system. One can use such a model to predict what happens when the bucket with the ball in it is set atop the door and the door is pushed, what happens when the bucket with the ball in it is tipped over, what happens when the bucket is used to throw the ball at the door, and so on indefinitely. The side effects of alterations to the representation will mirror the side effects of alterations to the represented system automatically and without the need for a prior and explicit specification of such side effects—as, for example, would be required if we were relying upon a logic in order to make such determinations. So scale models do not suffer from the prediction problem. Nor, for that matter, do they suffer from the qualification problem.

To see why, notice that much of what is true of a modeled domain will be true of a scale model of that domain. With regard to the door, bucket, ball setup, for instance, the following is true:

```
If
    the bucket is resting atop the door, and
    the ball is inside of the bucket, and
    the door is pushed, and
    it is not the case that the bucket is bolted to the
    top of the door, and
    it is not the case that there is a string
    connecting the bucket to the ceiling, and
    it is not the case that an atomic bomb will explode
    when the door is pushed, and
    so on . . .
then
    the bucket and the ball will fall to the floor.
```

The same is true of a scale model of this setup. That is, if both the scale model of the door is pushed and it is not the case that the scale model of the bucket is bolted to the top of the scale model of the door (and so on), then the scale models of the bucket and ball will fall to the floor. Just like our own predictions, the predictions generated through the use of scale models are implicitly qualified in an open-ended number of ways.

4. The Computational Solution to the Frame Problem

Because scale modeling media exhibit representational productivity, and because the scale models constructed from such media do not suffer from either the prediction problem or the qualification problem; the scale model metaphor looks quite promising in terms of its capacity to explain forethought in humans. While this finding should be of great interest to philosophers and psychologists, it does not, in and of itself, supply any serviceable tools to those who are attempting to design full-scale autonomous humanoids. As it turns out, however, the benefits of the scale model metaphor are inherited by certain computational systems that operate according to similar principles.

4.1 Avoiding the Prediction Problem with Computational Matrices

To see why this is so, it will help to start small. If you will recall, actual spatial matrices avoid the prediction problem with regard to predicting the consequences of changes in the spatial location of one or more represented items. The working memory component of Stephen Kosslyn's (1980) computational model of mental imagery functions in manner not unlike how an actual spatial matrix functions, and it enjoys many of the same benefits.[2]

In Kosslyn's model, representations are constructed by filling in the cells of a *computational* matrix. In order to implement a computational matrix, a computer's memory registers are ordered, either numerically or with the help of pointers, such that they functionally mimic the topology of an (x, y) co-ordinate system. There is, of course, nothing about memory registers per se mandating that they be used in this way. It is, rather, a primitive and (in some ways) arbitrary constraint imposed on processing. However, by constraining the use of memory registers in this way, one effectively creates a productive representational medium—one that, much like an actual spatial matrix, can be used to construct any of a wide variety of representations by filling in the appropriate cells. Also like a spatial matrix, the productivity of a computational matrix varies with the number of basic modeling elements—that is, with the number of cells.

In terms of the manner in which they support spatial inferences, computational matrix representations offer an alternative to the kind of high-level rules (i.e., operators) characterizing the production-system approach. As such, the use of computational matrix representations brings with it some definite advantages. As Zenon Pylyshyn explains,

> a matrix data structure seems to make available certain consequences with no apparent need for certain deductive steps involving reference to a knowledge of geometry . . . Further, when a particular object is moved to a new place, its spatial relationship to other places need not be recomputed (Pylyshyn 1984, p. 103).

In other words, when it comes to predicting the consequences of changes in relative spatial location, computational matrix representations are like actual spatial matrix representations in that they avoid the prediction problem. Kosslyn capitalizes on this fact with his model of visual imagery.

Glasgow and Papadias' (1992) have also modeled mental imagery with the help of computational matrix representations. Their model actually seems to mark a significant advance over Kosslyn's in that it incorporates the functional equivalent of *what* and *where* visual processing streams (Mishkin, Ungerleider, and Macko 1983) and can represent objects in three spatial dimensions. The *what* system of Glasgow and Papadias' model encodes the shapes of objects in terms of "patterns of filled cells isomorphic in surface area to the objects" (1992, pp. 370). The *where* system omits the details of object structure and instead represents the relative positions of objects.

---

[2] Kosslyn refers to this component as the 'visual buffer'.

Glasgow and Papadias' were clearly cognizant of the benefits of computational matrix representations. They explain:

> Consider, for example, changing the position of a country in [a] map of Europe. In a propositional representation we would have to consider all of the effects that this would have on the current state. Using the [computational matrix] to store the map, we need only delete the country from its previous position and insert it in the new one (Glasgow and Papadias 1992, pp. 376).

As with an actual spatial matrix, no matter where an object moves and no matter how many objects are represented, the relative positions of *all* of the objects will be updated automatically—and without the need for rules specifying how the change in location will affect their relative positions.

The capacity to predict the effects of changes in spatial location results from the imposition of primitive constraints on the use of memory registers. In order to support inferences concerning such spatial alterations as object rotation, it is necessary to impose even further primitive constraints on processing (Pylyshyn 1984, p. 204). By imposing such constraints at the level of the representational medium, however, one is able to avoid the need for rules framed with respect to the properties of specific objects. When the medium is itself constrained in the appropriate ways, there will be no need, for example, for a rule specifying that if a rocket is upright and is rotated 180 degrees around a horizontal axis then the nozzle will be over the body. The consequences of alterations to a particular representation are determined as a natural by-product of the primitive constraints governing the representational system.

The *what* component of Glasgow and Papadias' model, which is able to predict the consequences alterations such as rotation in *three* dimensions, actually brings to mind Janlert's proposed solution to the frame problem. Janlert's suggests, specifically, that in order to avoid the frame problem, "[o]ne possible approach is *mental clay*: a mass of small, uniform cells with modest capabilities of computation, interconnected to form the topology of granular space" (Janlert 1996, pp. 47).

Speaking in very general terms, it seems that the solution to the prediction problem is to exploit the constraints characterizing a given modeling medium. In the realm of non-computational modeling media, we typically take advantage of the constraints that a given medium *inherently* obeys, so there is no need to impose them. For example, clay is a representational medium that inherently respects the fact that two objects or two parts of the same object cannot occupy the same space. Any representations constructed with clay will likewise respect this constraint.

In the computational realm, however, the constraints that govern a modeling medium are the ones that we impose. It matters not, however, whether the constraints governing the medium are inherent or enforced. In either case, one has the guarantee that anything constructed from the materials supplied by that medium will obey the constraints *of* that medium. One is thus able to avoid the need for rules that are framed with respect to specific objects or properties thereof.[3]

4.2 The Limitations of Imagery Models

Much like an actual spatial matrix, Kosslyn's (1980) computational matrices can be used to represent a wide variety of two-dimensional spatial properties and relationships, and the representational productivity of a given matrix is a function of the number of basic modeling elements (i.e., the number of cells). Also like actual

---

[3] As computational systems, models like those of Kosslyn (1980) and Glasgow and Papadias (1992) ultimately effect whatever processing they do on the basis of the Turing-computable functions of some *formalism* or other. Many philosophers argue that this provides a sufficient basis for denying that such models harbor non-propositional representations (Block 1990; Devitt and Sterelny 1987; Pylyshyn 1984; Sterelny 1990). Because the point of the present conference is to highlight those state-of-the-art advances that might further the goal of creating autonomous humanoids, this point can be bracketed for the time being (though I know of at least one persuasive argument to the contrary (see Waskan 1999)).

spatial matrix representations, Kosslyn's computational matrix representations do not, with regard to the domain of two-dimensional spatial properties and relationships, suffer from the prediction problem. Glasgow and Papadias' (1992) computational matrix representations take matters a step further. They too are free from the prediction problem, but they are able to support inferences with regard to alterations over three spatial dimensions.

Although these constrained computational matrix representations do a very nice job of predicting the effects of changes in an object's location or orientation, they are not yet sufficient to account for the entire range of inferences that underwrite planned behavior in humans. What we should like to find is that the solution to the prediction problem outlined above can be scaled up in order to handle not only spatial inferences but causal inferences as well. Moreover, it will need to be demonstrated that the relevant computational representations are not susceptible to either the prediction problem or the qualification problem. Thus, in order to truly solve the frame problem, what seems to be required is nothing short of a computational implementation of the full-blown scale model metaphor for thought.

4.3 Virtual Forethought

This might seem a tall order. Given that books continue to be written about the frame problem, it seems fair to say that computational systems of this kind have not yet emerged onto the scene of cognitive science. Be this as it may, they do exist—and they can be purchased from upwards of $350 retail.[4] They are found in a sector of computer science that is, as yet, somewhat far removed from cognitive science. Specifically, researchers interested in Computer Generated Images and Virtual Reality (VR) Models have unwittingly made great strides toward supplying what might be viewed as a computational model of human forethought. This model preserves the precise virtues that make the scale model metaphor for thought so attractive.

4.3.1 Productive Polymesh

Much like computational matrix representations, VR models generally involve coordinate specifications (in this case, in an (x,y,z) co-ordinate 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. Co-ordinate 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—forming what is known as 'polymesh' (Watt 1993, p. 24). As an illustration, here we see some polymesh spheres that were constructed out of 32, 48, and 192 polygons, respectively:

Fig. 6. Representing a sphere with 32, 49, and 192 polygons.

Also like computational matrix representations, the productivity of the polymesh medium varies with the number of modeling elements—in this case, with the number of polygons. The more polygons you have, the more different things you can represent. Representational productivity is, of course, a property of thought that any

---

[4] I am happy to report that academic discounts are available through JourneyEd.com

viable model will have to explain, and, as anyone who has seen 'The Phantom Menace' is aware, the polymesh medium is *highly* productive.

4.3.2 Ray Dream Studio 5.02: A computational model of forethought.
The forte of VR modeling is actually the interaction of surface features and lighting. In addition to modeling the effects of lighting, however, computational modeling media have also been created that support the representation of causal interactions. As I will demonstrate presently, the pay-off is a determinate computational solution to the frame problem. To show this, I have created a set of models using an off-the-shelf program called Ray Dream Studio 5.02.

If you will recall from the beginning of my talk, I showed you a problem that most intact humans would have little trouble solving. The goal was to pick the implement that, when pulled, would bring the banana within reach. If the scale model metaphor for thought is correct, humans construct the cognitive equivalent of a scale model of the problem and use their model to predict the consequences of pulling on each of the implements. That is, according to this account, our mental representations are amenable to the same kind of truth-preserving manipulation that scale models are. To show that VR models also exhibit these basic powers of truth preservation, I created a model of the problem in Fig. 1.

Having created the model, first one and then the other rake was moved from the back of the enclosure to the front. If the system has the truth-preserving powers that we should like to see in a model of human forethought, then the outcomes of this manipulation should be that (a) the motion of the toothless rake does not alter the location of the leftmost banana, and (b) the motion of the T-bar results in the rightmost banana being moved closer to the opening. That is, we should like to find that the difference between moving the toothless rake and moving the T-bar is that in the latter case the banana comes along for the ride. As you can see, this is exactly what took place (see 'waskan-wayne1a.avi'). The outcome of this alteration to the representation mirrored what would happen in light of the corresponding alteration to the represented system—and without requiring any high-level rules or operators.

If you will recall from the discussion of the frame problem, the qualification problem raised some serious worries about the viability of logic-metaphoric accounts of forethought. It was shown that our knowledge of the consequences of alterations to even simple physical systems could not be formalized in the manner that many had hoped it could. For instance, with regard to the banana problem, my pulling on the T-bar will cause the banana to slide within reach, provided that, among other things, there is not a hole in the table that the banana will fall through. This, and countless other qualifications, would need to be added to the antecedent of the corresponding rule.

Scale models do not suffer from this limitation because they admit of the same qualifications as the domains they represent. 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. If the models constructed with Ray Dream Studio 5.02 work in a similar manner, one of the great virtues of the scale model metaphor for thought will have been carried over into the computational realm.

In order to examine this possibility, the model that I just showed you was altered in one simple respect—that is, a hole was put in the table between the T-bar and the opening to the enclosure. Once again, the results (shown in 'waskan-wayne1b.avi') were highly promising. Instead of the banana being carried along to the edge of the container, it fell through the hole. Hence, another advantage of the scale model metaphor for thought has been shown to carry over into the computational realm. Like our own predictions, and the predictions generated through the use of scale models, predictions generated on the basis of virtual reality models are implicitly qualified in an open-ended number of ways.

The other major problem that I discussed was the prediction problem. I illustrated the prediction problem with the help of the door, bucket, and ball setup. If you will recall, we would need countless distinct rules or operators if we wanted to formalize

what we know about the consequences of various alterations to this system. Like the qualification problem, the prediction problem results from the fact that our knowledge of the consequence of alterations to simple physical systems cannot be formalized in the manner that many had hoped it could.

Scale models, on the other hand, do not suffer from the prediction problem. We saw that the same is true, at least within the limited domain of updating spatial relationships, of computational matrix representations. VR models, however, avoid the prediction problem with regard to both 3D-spatial and causal relationships. To show this, I constructed a model of the door, bucket, ball setup and carried out a series of alterations.

The starting condition for the first alteration has the bucket resting atop the door and the ball over the bucket (see 'waskan-wayne2a.avi'). The only direct manipulation to the ensuing chain of events is that the door is opened rather abruptly. The best result that could be hoped for in this case would be that the bucket and the ball would to fall to the floor. This, as you can see, is exactly what occurred. Once again, the only direct manipulation to the chain of events was the opening of the door. The side effects followed automatically, and without the need for a rule framed with respect to the properties of doors, buckets, and balls.

In a completely new scenario, the bucket is turned upside-down and placed over the ball. The bucket is then moved through the doorway, and it is subsequently raised. Were this alteration carried out with respect to either the actual door, bucket, ball system or a scale model of this system, we should expect to find the ball (or the scale model of the ball) underneath the bucket. This, as you can see, is also what happens in the virtual reality model (see 'waskan-wayne2b.avi'). We could continue using the same model in this way to generate any number of predictions.

Finally, you might recall Janlert's suggestion that a system not beset by the frame problem should admit of incremental additions. For systems that rely upon rules framed with respect to particular objects and properties thereof, simple additions to the represented domain give rise to a profusion of new relations and consequences that need to be captured in terms of further rules. With scale models, on the other hand, simple additions to a represented domain pose no such problems. For instance, if a board were added to the door, bucket, ball setup, one could take this into account simply by adding a board to one's scale model of that setup.

As we saw, computational matrix representations also deal with additions to the represented domain quite gracefully. The same is true of VR models. To show this, I simply added a board to the door, bucket, ball model. This time, 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 through the doorway. One again, what we would expect to happen in the world (and in a scale model thereof) took place in the VR model (see 'waskan-wayne2c.avi'). That is, the ball bounced off of the board instead of rolling through the door.

4.3.3 Problem Solved

The remarkable fact is that these VR models, which were constructed with an off-the-shelf animation program, do not suffer from the frame problem. Considering that Ray Dream Studio 5.02 was, I presume, never meant to be considered a computational model of mental representation and predictive inference, its successes in this regard are extraordinary. Like scale models, and for what look to be the very same reasons, VR models do not suffer from the frame problem. That is, because the relevant constraints characterize the representational medium itself, anything constructed from the materials constituting that medium will obey those same constraints.

4.3.4 Psychological Plausibility of Virtual Forethought

There are, admittedly, some disanalogies between scale models and VR models. For starters, because the constraints governing scale models are inherent; 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. With

VR models, on the other hand, the constraints characterizing the medium are primitive but not inherent. For this reason, one can, in effect, change what an object is made of without having to construct that object anew. I leave it to my readers to decide whether this is, from a psychological-modeling standpoint, a strength or a weakness.

Another disanalogy between scale models and VR models is that only the former are constrained by the actual laws of physics. The primitive constraints characterizing VR modeling media need not, and often do not, accurately reflect those principles operative in the world. In Ray Dream 5.02, for example, the outcomes of collisions are not determined by such factors as mass, momentum, or the storage and release of energy due to compression. The simple bouncing behavior of the ball in the second model was the result of a primitive rebound variable that governs how bouncy objects are. Although this may seem like a shortcoming, there is, somewhat surprisingly, a case to be made that this is not unlike how physics-naïve individuals represent the outcomes of collisions.[5]

In a seminal study conducted by Chi, Glaser, and Rees' (1982), for instance, novices and experts were asked to categorize a set of physics problems. Chi et al. found that novices categorized problems on the basis of their surface features, while experts categorized problems on the basis of the underlying physical principles they exemplified. Andrea DiSessa (1983) later examined the manner in which physics-naïve individuals understand the nature of bouncing behavior in particular, and the results were similar. DiSessa discovered that one physics-naïve individual, called '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 the world.

This 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 (and not unlike physics-naïve individuals), these models 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 and bananas, buckets, doors and balls, and so on.[6]

A look at the existing literature reveals that there are, in fact, other interesting parallels between the inference powers of physics-naïve individuals and those exhibited by VR models (see Chi, Glaser, and Rees 1982; De Kleer and Brown 1983; DiSessa 1983; Larkin 1983; Norman 1983). Many naïve-physics researchers have even proposed that individuals create mental models of the world and 'run' their models in order to predict the behavior of the physical contingencies they encounter.

5. VR Models and Autonomous Humanoids

The implications of these finding for the possibility of constructing an autonomous humanoid are straightforward. The techniques of VR modeling provide a solution to a problem that has long been a stumbling-block for AI research. Unlike operators (or frame axioms, meaning postulates, or what have you) VR models can be used to support predictions regarding the consequences of an open-ended number of alterations to systems they represent—and without the need for endless qualifications. In other words, all indications are that VR models do not suffer from the frame problem.

---

[5] Though this might seem less surprising if members of the Ray Dream programming team happen to be physics-naïve. More likely, however, they are physics-savvy and recognize a computation-sparing shortcut when they see one.

[6] Although Hayes (1995) has famously suggested that AI researchers incorporate the principles of naïve-physics in their models, the present approach seems at rather at odds with his prescription of an expert-systems style axiomatization.

A truth-preserving representational medium is, of course, only one of many systems that figure into full-blown, goal-directed reasoning. Also implicated are mechanisms responsible for weighing options and implementing those actions that bring one closer to ones goals. It seems worth considering whether or not such goal-directed reasoning can be modeled using a hybrid of GOFAI and VR techniques. It may be, in fact, that developers of the latest versions of SOAR (i.e., SOAR 7.3 and higher) have already laid substantial groundwork for this endeavor by devising ways of interfacing SOAR with 'external' inference mechanisms.[7]

Of course, the problems related to goal-directed reasoning themselves only comprise a subset of the ones that will need to be dealt with before we can realize the dream of creating an autonomous humanoid. At the very least, problems related to memory, attention, pattern recognition, and motor control will all have to be dealt with. I do, however, think that by solving the frame problem we have moved one sizable step closer to our ultimate goal.

References

Block, N. 1990. Mental Pictures and Cognitive Science. In W. G. Lycan (ed.), *Mind and Cognition* (Cambridge, MA: Blackwell), pp. 577-606.

Chi, M.T.H., R. Glaser, & E. Rees. 1982. Expertise in Problem Solving. In R. J. Sternberg (ed.), *Advances in the Psychology of Human Intelligence* (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 7-76.

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: Cambridge University Press.

De Kleer, J. & J.S. Brown. 1983. Assumptions and Ambiguities in Mechanistic Mental Models. In D. Gentner and A. L. Stevens (eds.), *Mental Models* (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 155-190.

Devitt, M. & K. Sterelny. 1987. *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge: The MIT Press.

DiSessa, A. 1983. Phenomenology and the Evolution of Intuition. In D. Gentner and A. L. Stevens (eds.), *Mental Models* (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 14-33.

Fodor, J. 1975. *The Language of Thought*. New York: Thomas Y. Crowell.

Fodor, J.A., J.D. Fodor, & M.F. Garrett. 1975. The Psychological Unreality of Semantic Representations. *Linguistic Inquiry*, 6, 515-531.

Haselager, W.F.G. Unpublished manuscript. Connectionism, the Frame Problem and Systematicity.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. Cambridge: The MIT Press.

Hayes, P.J. 1995. The Second Naive Physics Manifesto. In G. F. Luger (ed.), *Computation and Intelligence* (Menlo Park, CA: AAAI Press), pp. 567-585.

Janlert, L. 1996. The Frame Problem: Freedom or Stability? With Pictures we can have both. In K. M. Ford and Z. W. Pylyshyn (eds.), *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence* (Norwood, NJ: Ablex Publishing), pp. 35-48.

Johnson-Laird, P.N. & R.M.J. Byrne. 1991. *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Köhler, W. 1938. *The Place of Value in a World of Fact*. New York: Liveright Publishing Corporation.

Kosslyn, S.M. 1980. *Image and Mind*. Cambridge: Harvard University Press.

Larkin, J.H. 1983. The Role of Problem Representation in Physics. In D. Gentner and A. L. Stevens (eds.), *Mental Models* (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 75-98.

Lindsay, R.K. 1988. Images and Inference. *Cognition*, 29, 229-250.

McCarthy, J. 1986. Applications of Circumscription to Formalizing Common-Sense Knowledge. *Artificial Intelligence*, 28, 86-116.

McCarthy, J. & P.J. Hayes. 1969. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer and D. Michie (eds.), *Machine Intelligence* (Edinburgh: Edinburgh University Press), pp. 463-502.

Mishkin, M., L.G. Ungerleider, & K.A. Macko. 1983. Object Vision and Spatial Vision: Two Cortical Pathways. *Trends in Neurosciences*, 6, 414-417.

---

[7] Indeed, though I am not privy to the details, outward appearances suggest that something like computational matrix representations are being used instead of operators in order to keep track of changing two-dimensional spatial relationships.

Norman, D.A. 1983. Some Observations on Mental Models. In D. Gentner and A. L. Stevens (eds.), *Mental Models* (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 7-14.

Palmer, S. 1978. Fundamental Aspects of Cognitive Representation. In E. Rosch and B. Lloyd (eds.), *Cognition and Categorization* (Hillsdale, NJ: Lawrence Erlbaum Assoc.).

Povinelli, D. 1999. Toward a New Theory of the Evolution of Human Social Intelligence: The Reinterpretation Hypothesis. Paper read at Twenty-Fifth Meeting of the Society for Philosophy and Psychology, at Stanford University.

Pylyshyn, Z.W. 1984. *Computation and Cognition: Toward a Foundation for Cogntive Science.* Cambridge: The MIT Press.

Rips, L.J. 1990. Paralogical Reasoning: Evans, Johnson-Laird, and Byrne on Liar and Truth-teller Puzzles. *Cognition*, 36, 291-314.

Sterelny, K. 1990. The Imagery Debate. In W. G. Lycan (ed.), *Mind and Cognition* (Cambridge: Blackwell), pp. 607-626.

Waskan, J.A. 1999. The Medium of Thought: A Model of Representational and Inferential Productivity, Philosophy-Neuroscience-Psychology Program, Washington University, Saint Louis.

Watt, A. 1993. *3D Computer Graphics*. 2nd ed. Harlow, England: Addison-Wesley Longman Limited.