CaMeRa: A Computational Model of Multiple Representations

HERMINA J.M. TABACHNECK-SCHIJF
ANTHONY M. LEONARDO
HERBERT A. SIMON
Carnegie Mellon University

This research aims to clarify, by constructing and testing a computer simulation, the use of multiple representations in problem solving, focusing on their role in visual reasoning. The model is motivated by extensive experimental evidence in the literature for the features it incorporates, but this article focuses on the system's structure. We illustrate the model's behavior by simulating the cognitive and perceptual processes of an economics expert as he teaches some well-learned economics principles while drawing a graph on a blackboard. Data in the experimental literature and concurrent verbal protocols were used to guide construction of a linked production system and parallel network, CaMeRa (Computation with Multiple Representations), that employs a "Mind's Eye" representation for pictorial information, consisting of a bitmap and associated node-link structures. Propositional list structures are used to represent verbal information and reasoning. Small individual pieces from the different representations are linked on a sequential and temporary basis to form a reasoning and inferencing chain, using visually encoded information recalled to the Mind's Eye from long-term memory and from cues recognized on an external display. CaMeRa, like the expert, uses the diagrammatic and verbal representations to complement one another, thus exploiting the unique advantages of each.

1.0 INTRODUCTION

In an Economics 101 course, the professor walks to the blackboard and begins to explain a problem. She uses speech and graphs to get her point across. Suddenly, a student raises his hand and asks a question. The professor poses a question in return and a lively discussion follows. Finally, she returns to the problem at hand, examines the board a bit, and resumes where she left off. To do this, she must first build a verbal and diagrammatic representation of the problem she is explaining, and later, after her short-term memory information has

Direct all correspondence to: Herbert A. Simon, Carnegie Mellon University, Pittsburgh, PA 15213.
Anthony M. Leonardo is now at the California Institute of Technology, in Pasadena, CA.
been lost during the interrupting discussion, she must rebuild this representation from the diagrams left on the blackboard.

An immediately noticeable quality of this kind of expert problem solving is the ease and frequency with which an expert uses multiple representations. Indeed, it would be difficult to find a current mathematics, physics or economics textbook without graphs, or a blackboard bare of diagrams after a lecture in one of these domains. The economics expert whose behavior is reported in a later section of this article found it extremely difficult to explain even simple economics principles without using multiple representations, visual elements in particular, and failed to do so on three subsequent trials. The expert does not merely employ multiple representations—he or she is often dependent on them.

Novices, however, have many difficulties using multiple representations. Protocols taken from beginning solvers contain piecemeal collages of representations which are rarely, and often inappropriately, combined (Tabachneck, 1992). Physics teachers say they must often force students to draw and use diagrams instead of crunching equations. To summarize, experts rely on multiple representations, whereas novices are often unable to use them successfully.

Experts use multiple representations because different representations of a problem are seldom equivalent computationally, even when they contain equivalent information (Larkin & Simon, 1987). It may be easier to explain the concept of "cognition" verbally, but to explain the workings of a steam engine visually. Moreover, many complex tasks have parts that are better explained in different representations. By modeling an expert's explanation of an economics problem, we will illustrate how the diagrammatic and verbal representations complement one another, and thus allow the expert to exploit the unique advantages of each. In certain tasks, the diagrammatic representation has computational advantages over the verbal representation: less search for information, easier recognition of information, and simpler inference processes (Glasgow, 1993; Larkin & Simon, 1987). However, the verbal representation has its own advantages; for example, in evoking explanations of events in terms of already-familiar processes (e.g., describing why prices fall when supply is greater than demand, provided that terms like "price," "supply," "demand," "surplus" are already understood, and information giving their meanings is stored in memory).

An expert also can use a visual picture such as a graph or diagram as a summary, or, when a picture is being constructed as part of an explanation, as a dynamic place-holder. An example of the latter is the usual proof that there are an infinite number of prime numbers.

We draw a horizontal number line extending from 0 indefinitely to the right. We suppose, contrary to the theorem, that there are only a finite number of primes, hence a largest prime, $P_{\text{max}}$. We mark the hypothetical $P_{\text{max}}$ somewhere on the number line, and mark another number, $N$, equal to 1 plus the product of all the primes including $P_{\text{max}}$, on the line somewhere to the right of $P_{\text{max}}$, for it is certainly larger than $P_{\text{max}} + 1$. Now we observe that if we divide $N$ by any of the primes, there will be a remainder of 1, hence (on the previous assumptions) it is a prime. But as it is larger than $P_{\text{max}}$, $P_{\text{max}}$ cannot be the largest prime, a contradiction. Therefore, the assumption that there are only a finite number of primes is false.
Here the proof is verbal, but the number line, with $P_{\text{max}}$ and the larger prime marked on it, keeps track of the stage of the argument and of the relative sizes of the two numbers. It serves as a supplement to short-term memory that most persons making or trying to understand the verbal argument find invaluable. This place-holder is updated throughout the explanation, in pace with the current state of the reasoning. The interaction between different knowledge modalities in our model demonstrates how the unique properties of a diagrammatic representation and the use of an external visual display can complement and be integrated with verbal representations. These multiple representations are more useful than a single representation of the same information.

We employ the term “pictorial” in order to refer unambiguously to those parts of our model, CaMeRa, that process graphs and drawings but not other visual images like words and text. Our working model provides a computational analysis of the coordinated use of multiple representations and demonstrates how they can produce expert problem solving behavior.

### 2.0 Assumptions and Definitions

It will be useful, before proceeding further, to clarify some concepts that will be central to our account: representations and their equivalence or inequivalence (informational and computational), the relation of representations to their denotations, and the meaning of the term understand.

#### 2.1 The Nature of Representations

A representation has two components: (1) a format for recording and presenting information and (2) processes for using and modifying the information. Neither a format nor a set of processes is sufficient by itself to define a representation; both must be present. Labels, axes, and lines and their definitions are part of the format of a graph. The actions needed to find the $x$ and $y$ coordinates of a point in the graph are processes that use and modify the information.

#### 2.2 Equivalence of Representations

Two representations are informationally equivalent if any information provided by the one can be provided by the other, and vice versa (see also Tabachneck & Simon, 1996). The fact that two representations are informationally equivalent does not imply that they are equally useful or efficient for the same task, that is, computationally equivalent. The effectiveness of representations for reasoning, communicating and instructing depends on how easy it is to extract the information that is implicit in them (Larkin & Simon, 1987; Palmer, 1978), which in turn depends on the representation, the tasks to which it is applied, and the user’s familiarity with both problem and representation. For example, to find the approximate equilibrium price and quantity on a supply-demand graph is relatively easy: notice the intersection point of the supply and demand lines. Finding these values algebraically is
more difficult, as it requires one to solve simultaneous equations. The diagrammatic and algebraic representations are informationally, but not computationally, equivalent.

2.3 Denotation

The denotation of a verbal or pictorial structure held in memory consists of the external information that is mapped on that structure. This external information, together with previously stored internal information that may be called upon when using the structure to perform tasks, constitutes the semantics of the internal structure. Denotation gives real-world meaning to the structure. During an act of recognition, the system applies tests to sensed objects and relations in order to determine the internal structures to which they correspond. For example, replying to "Is there a chair here?" involves "sorting" the features of perceived objects to match the ones associated with the name "chair" in verbal memory and the information associated with it. Understanding is not a single simple event, but may involve a whole sequence of processes.

2.4 Understanding

Internal representations of information are used in performing a variety of tasks. We say that a system (human or mechanical) understands the information, or that the information is meaningful, in proportion to the range and complexity of the tasks to which it can be applied successfully. For example, a person is said to understand the sentence, "Please sit down on that chair," if he or she can recognize a nearby object as a chair, can walk up to it and sit down on it.

Thus, the meaningfulness of the sentence is related closely to the ability to map back and forth between the sounds of the spoken sentence (stored in the Mind's Ear) and the scene recorded in the Mind's Eye, and to associate both of these with other information (perceptual and motor) already stored in memory. The latter information enables the person to reduce his or her distance from the chair and to perform the act of sitting down. It also enables the person to predict likely consequences of the motor acts: that certain acts will reduce the distance to the chair, that the chair will support his or her weight, and so on.

3.0 FEATURES OF THE MODEL

The CaMeRa model was designed to capture as accurately as possible current, experiment-based hypotheses about how human perception and reasoning employ the several sensory modalities, perceptual processes and memory structures to accomplish their tasks. Before describing the model formally, we will introduce in this section the principal experimentally supported features of the processes and structures that we have considered, and for each, indicate briefly how it has been incorporated in the model. The subsequent two sections will give a systematic overview of the CaMeRa model followed by a detailed description.
3.1 Different Modalities Have Different Representations

**Empirical evidence.** There is much support in the literature for separate and different modalities. Different types of short-term memories (Atwood, 1971; Baddeley, 1986; Brooks, 1967, 1968) such as a “visual-spatial scratchpad” and an “auditory loop” appear to have different capacity limits (e.g., Baddeley, 1986; Zhang & Simon, 1985). Evidence for different processes used in verbal as compared with visual representations can also be found in research on individual differences in abilities to process verbal, pictorial and spatial information (e.g., Just & Carpenter, 1985; Paivio, 1971; Riding & Calvey, 1981) as can demonstrations of the computational advantages of having such different representations (Larkin & Simon, 1987). Finally, Kosslyn offers strong evidence for both separation and difference in verbal and visual perception and cognition, separation increasingly supported by neurophysiological evidence (Kosslyn, 1980, 1994; Kosslyn & Koenig, 1992). Zhang and Norman (1994) found differences in the way their subjects worked with visual and spatial information.

**The model.** Each mode of representation in the CaMeRa model has its own data structures and operators. Thus, if a pictorial stimulus is perceived, it is encoded and stored in a pictorial data structure and interpreted with pictorial rules. Similarly, if a sentence is read, it is stored in a verbal-semantic representation. If a thought is recalled from memory, it will be processed in verbal form in verbal STM, and in pictorial form in pictorial STM (assuming the information has both a pictorial and verbal LTM representation). Pictorial operators cannot modify verbal data structures and vice versa, although each may, through associative links, retrieve information from the other. The modalities interact like two layers of memory drifting alongside each another, tied together temporarily by links in working memory, and disconnected again after the information is no longer needed (see Figure 1 for a “Gestalt” of this).

3.2 Mental Images Resemble Visual Stimuli Closely

**Empirical evidence.** In studying the representations people use, we must distinguish between external and internal information, between information in the environment and information in the brain. To understand a drawing of an “A” above a “B”, the external drawing must be transformed into an internal representation, the mental picture contained in the Mind’s Eye (Kosslyn, 1980).

Much research supports the idea that information encoded into STM by perception is represented in basically the same form as the identical information transferred from LTM to STM, so that a mental image produced by looking at an external drawing is represented in STM with the same type of pictorial structures as a mental image generated from the memory of the drawing. We call the hypothesis based on this evidence the Mind’s Eye Hypothesis.

Evidence for the Mind’s Eye Hypothesis is of three kinds: behavioral, neurophysiological, and computational (Glasgow, 1993; Tabachneck & Simon, 1996). We call attention
Referent tie between visual and verbal layer
Referent tie within the visual or verbal layer
(in LTM, these ties are dormant; they are executed only in STM)

Figure 1. The four "layers" of CoMeRo's memory. The various constructs will be explained elsewhere in the text.
especially to the research of Kosslyn and his colleagues (1980, 1992, 1994), who have demonstrated through numerous experiments and simulations in all three areas that images evoked by perception and memory produce the same configuration in the Mind's Eye, and use the same procedures. Computational accounts have also been given by Baylor (1971), and behaviorally by Finke and Shepard (1986).

The Mind's Eye hypothesis does not require that "pictures in the head" be exactly the same as pictures on a piece of paper (Anderson, 1985). Different amounts of detail may be available from each. Moreover, external pictures may remain stable while mental pictures may require frequent periodic refreshment. What the "identity" hypothesis means is that images from both sources can be operated upon by the same processes. Experiments show that certain operations are more difficult with mental imagery than with perception, such as rotating pictures mentally and then reinterpreting them (Chambers & Reisberg, 1985; Slezak, 1992). Rather than being due to a different medium or different processes, such contrasts may result from different speeds and computational costs of accessing information, different degrees of permanence, or a higher resolution in perceptual imagery. When data structures and operations are simplified, mental imagery has been shown to perform quite like perception (Finke, 1986; Pinker & Finke, 1980).

Thus, external perception has several advantages over unsupported mental imagery. However, imagery has a few advantages as well. It can combine different percepts or ideas into a new percept or idea, such as a rabbit with wings, or a new machine. (There are limits. What would it mean to image a 4-dimensional cube or a square circle?)

The model. At the working memory level in CaMeRa, the medium, the data structures and the rules that operate on them are the same both for images derived from perception and images drawn from memory. We have not yet implemented mechanisms that would cause differences in resolution between images of these two kinds, but these will appear as we implement limitations on working memory.

3.3 All Inferences Are Made on Working Memory Contents

Empirical evidence. In a sense, this is true by definition. That is to say, working memory is sometimes defined as the system that supplies the inputs for processing, functions with which to do the processing, and a storage area to receive the output of processing. The empirical content of the claim derives from consistent evidence that memories involved in current processing have different parameters for acquisition rates, access times and durability than memories used for more permanent storage. For example, the large body of evidence for the structures hypothesized in EPAM (Elementary Perceiver and Memorizer) points to short-term memory acquisition times of a few hundred milliseconds, but long-term memory times of about eight seconds per chunk; STM retention times of about two seconds (without rehearsal), but indefinite LTM retention times (Richman, Staszewski, & Simon, 1995).

The model. In CaMeRa, all the inferencing is done in working memory (the auditory and visual short-term memories). In order for the information in data to be used, it must be cop-
ied to the working memory. When learning processes are added to the present model, all changes in memory structures will also be made initially in the working memory. Whether these changes are permanent or not will depend on the type of reasoning that was done with the data. The LTM record of data used merely for computation would not be altered, but if inferencing showed that data were erroneous or incomplete, corrections would be learned and introduced into LTM.

3.4 Experts Can Use and Integrate Multiple Representations

**Empirical evidence.** An immediately noticeable quality of expert problem solving is the ease and frequency with which an expert uses multiple representations. An example of this, from data we have gathered, is the behavior of an economics expert¹ who was asked to explain to a student some principles of supply and demand, of equilibrium and of the effects of a shift in the supply schedule. While engaging in this explanation, the expert was constructing a graph on a blackboard to illustrate the various principles, interspersing his drawing with verbal explanations. A part of the protocol the expert generated is reproduced in Appendix 1, along with our analysis of how the pictorial and verbal parts of this protocol are interconnected, and how the CaMeRa model accounts for the expert’s reasoning. The corresponding reasoning using exclusively verbal means without a diagram is described formally in Appendix 2. It can be seen to be substantially more complex than the diagrammatic reasoning.

When the expert was asked to give a similar explanation without either using or referring to pictorial elements (in fact, sitting on his hands to prevent gesturing), he was unable to do so in three consecutive trials. He made visual references within 3 or 4 sentences⁵, and, when interrupted, reported he had been constructing a mental diagram and reading information from it.

**The model.** The way in which this behavior is modeled by CaMeRa will be explained in considerable detail in the next sections of this article. In general terms, the verbal and visual information of the modeled expert are linked together associatively by referents to one another. When visual data that carries verbal information (e.g., labels on variables) is accessed by recognition or otherwise, the verbal information that it might refer to is activated next. The visual and verbal “layers” of memory are closely tied together when the information is activated in working memory.

3.5 Novices Have Difficulty Using and Integrating Representations

**Empirical evidence.** In economics, equations, tables and graphs are widely used to enhance, enrich, and illustrate verbal explanations. Comments on two of our experiments, reported in detail elsewhere (Tabachneck, 1992), illustrate novices’ difficulties in interpreting verbal and pictorial representations and achieving integration of the two. Although the model does not currently emulate a novice, and cannot yet learn, it was built with the
expert/novice differences in mind. CaMeRa can easily be degraded to become a novice model by eliminating all the referents between visual and verbal material.

Experiment 1. Interpreting and Integrating Pictorial and Verbal Information

Novices in economics read a tutorial explaining in words basic principles of supply and demand, then were presented with problems (finding an equilibrium, stating the consequences of shifting a supply schedule, explaining how equilibrium is maintained). Data for the problems were presented to different groups of subjects (four subjects per group) in informationally equivalent forms (line graphs, algebraic equations, or tables), in a between-subjects design. Problems presented with the data were the same for all groups. By tracing the subjects’ thought processes with talk-aloud protocols (Ericsson & Simon, 1993), we determined that they were indeed working with the representation we had given them (Tabachneck, 1992; Tabachneck & Simon, 1992).

Answering some of the problems required only numerical computations on information displayed in graphs, equations or tables; other problems required explanations of changes in prices and quantities of goods after imposition of a tax or some other external event. To answer questions of the latter kind required linking the introductory information, problem statements and pictorial displays, all freely available from the Hypercard stack, to already stored LTM memories of terms like “sale,” “purchase,” “price,” “quantity of a commodity,” “surplus,” “shortage.”

Subjects in all three conditions provided wholly unsatisfactory answers to the explanation questions. Even though all information was readily available, these novice subjects did not integrate the verbal information in the tutorial with the information in the data sets when both sources were needed to answer the questions properly. Instead, their answers used only one representation at a time (Tabachneck, 1992). For these questions, subjects either used verbal information, or described the particular strategy they had followed on the numerical questions, without providing a rationale for the strategy in terms of the underlying economics.

On numerical questions, subjects using line graphs did better than subjects who used equations or tables. The reason for this was that the subjects often used rather superficial pictorial reasoning on these questions, relying on a simple, mechanical strategy to answer them (i.e., find and act on those parts of the display that changed from the before- to the after-situation). With the line graphs we used, where this amounted to reading off numbers at intersections of supply and demand lines, the strategy met with success. With equations and tables, where changes in the display required computation after reading the numbers off, the strategy produced incorrect answers. Subjects in all conditions had little success in explaining their answers. Often, they merely described their strategy, for instance, “Well, I looked at the equations and they (the intercepts) differed by 1 so I figured the price went up by a dollar.”

When the verbal material was used at all in answering the questions, representations were generally integrated on a very shallow basis, for example, connecting a label with a line, or a legend with a symbol—elements that were usually in the same data screen. On a slightly higher level, some subjects compared outcomes of different types of reasoning. For
instance, verbal reasoning may have led to the conclusion that a variable would increase; subsequent mathematical reasoning may have inferred that the variable decreased. Such contradictions produced repetitions of both types of reasoning, and very rarely (2 out of 132 answers), a more involved interaction between the different reasoning modes.

Lack of integration with the verbal information (perhaps caused partly by failure to understand the verbal information), led to a lack of understanding of what the elements in the visual display really meant. Inferences, if any, from the data displays were shallow, and often wrong. Simple perception could have yielded the answers, but only to someone who had learned to notice the relevant features of the graph, possessed the appropriate inference operators, and could translate back and forth between the operators and their real-world economic interpretations (Tabachneck & Simon, 1996). The novices, who lacked this knowledge, could not obtain the answers. As we shall see later, economics experts readily integrate the information in this way.

**Experiment 2. Generating Graphs from Verbal Information**

In a second experiment, we explored whether novice subjects could generate graphs from a verbal tutorial similar to that in Experiment 1. Four students saw relevant graphs along with the tutorial ("Graph" group), and four saw only the tutorial ("Text" group). First, we presented them with the task of explaining a simple supply-demand relation, like one presented in the tutorial, but without access to the tutorial text or graphs. No graphs were drawn by any of the Text subjects; only two of the four Graph subjects drew graphs. Although these were well drawn, the verbal reasoning was not properly tied to them, and these subjects showed no deeper understanding of the economics than the subjects who did not use graphs at all (Tabachneck, 1992).

Subjects were then given the surprise task of drawing the graph for the tutored problem, with full access to the text. Of the four subjects in the Text group, only one drew a reasonable graph. All Text subjects had difficulties assigning variables to axes and drawing the lines within the graphs. Three of these four subjects were unable to represent the causal connection between surplus and price change on the graph.

More surprising, in the Graph group, only two subjects (those who had used graphs in their previous replies, and thus had practiced their graph-drawing) drew the graph correctly. The other two, like the verbal group, were unable to represent the causal mechanism, to relate the verbal information to the visual information, though all had seen it represented explicitly several times (Tabachneck, 1992).

The model. In the expert model, the visual and verbal information is associated by referents; for example, an intersection of supply and demand lines is associated with “equilibrium.” When we will simulate novices, the referents embodying knowledge of subject matter, but not those embodying basic perceptual skills, will initially be lacking and will have to be learned. Novices will be as skilled as experts in associating the point of intersection of two lines on a graph with the word “intersection”; but noting that two lines intersect does not automatically evoke the verbal statement that “the equilibrium price is the price at which the quantity demanded equals the quantity supplied”; to construct each such subject-
specific referent will call for explicit processing. Initial inability to integrate the two representations explains in part the deficiencies in novice performance, as attested by the evidence of the protocols, as well as the common experience in classrooms, where students who have been provided with a verbal explanation do not immediately understand the corresponding graph when it is displayed on the blackboard to “clarify” the argument. Of course we would expect to find other differences between novices and experts besides this one.

3.6 Pictorial Representations Are Often Computationally Advantageous

**Empirical evidence.** Larkin and Simon (1987) have shown that for certain tasks, pictorial representations (1) require less search for information, (2) allow easier recognition of information, and (3) admit simpler inferencing processes than do verbal representations.

Our analysis of the expert’s protocol shows that the external pictorial display could have benefited him in various ways. First, it supported pictorial reasoning. Second, it evoked recognition processes to access information in memory. Third, the dynamic build-up of the picture served as a summary of the processing, decreasing the short-term memory load and allowing the expert to concentrate on reasoning instead of on the task of maintaining his internal representations (Tabachneck, Leonardo, & Simon, 1994). All these advantages can benefit the expert, but not a novice, as only the expert holds in LTM the semantic meanings of the symbols and patterns that were drawn on the board. Novices can benefit only to the extent of their knowledge of these meanings.

**The model.** All of these aspects are used in the model as they were by the expert. In its simulation of mental imagery, CaMeRa uses a representation of diagrams as node-link structures (cf. Larkin & Simon, 1987), consisting of straight and curved lines and their intersections, together with verbal labels.

3.7 An External Representation Serves as Supplementary Memory

**Empirical evidence.** Kotovsky, Hayes, and Simon (1985) have shown in research on Tower of Hanoi isomorphs that, if subjects are not allowed to use an external pictorial representation, problem solving difficulty increases exponentially as the representations and rules grow more complex. This is an example of how the external representation can carry some of the burden of encoding constraints and function as external memory.

Zhang and Norman’s *representational effect* is closely related to this topic (Zhang & Norman, 1994). In their words “different isomorphic representations of a common formal structure can cause dramatically different cognitive behaviors” (p. 88), a finding that has also been noticed by others, including Hayes and Simon (1977), Kotovsky, Hayes, and Simon (1985), Simon and Hayes (1976). Zhang and Norman found that as one modifies the external representation to embody implicitly more of the constraints of the problem, subjects are faster and less error prone. For instance, here are three ways to encode the relation...
A > B > C: green > red > blue, large > medium > small, and three full, smooth cylindrical coffee cups of different diameters, which only stack one way; that is, large above medium above small. The color relationship, as encoded by the retinal receptors, has no ordinal features, but the large/medium/small and the coffee cups do. Subjects were worst at encoding the color relationship, in which the external representation offered no memory aid. Next came the large/medium/small relationship. In the case of full coffee cups, the rule didn’t even have to be given: it was obvious that the only way they stacked was a larger one on a smaller one. Subjects recognized that if they would stack the full cups the other way, coffee would spill all over!

Koedinger and Anderson (1990) emphasize the important role that recognition of visual chunks plays for experienced geometry teachers. The teachers did not think in terms of separate lines and angles, but rather of combinations like “line-angle-line,” which forms the basis for one of the proofs for congruent triangles, or line-bisecting-two-parallel lines. Each such composite unit is treated as a chunk rather than as three lines. The novice geometry students, in contrast to the teachers, did not recognize these chunks but saw separate lines and angles.

Larkin and her colleagues also showed how familiar patterns in pictorial representations could serve as an access route or index, both to factual knowledge and to information about actions and strategies (Larkin, McDermott, D. P. Simon, & H. A. Simon, 1980).

**The model.** In the CaMeRa model, the graphical representation serves both to initiate and summarize reasoning. The expert recognizes familiar chunks as they are encoded in visual memory from the drawing. This recognition pushes the reasoning forward. The interplay between immediate reasoning and recognition largely removes the need for a goal stack in the economics problems that CaMeRa solves.

### 3.8 There Are Different Subrepresentations for Spatial and Object Information

**Empirical evidence.** Considerable neurobiological evidence suggests that the primary visual cortex sends projections to at least two other areas of the brain where high-level visual representations are then formed. Spatial information (e.g., location) is sent to the posterior parietal lobes and the superior colliculus, while object (e.g. shape, color) information is processed in the inferior temporal lobes (Farah, 1990; Kosslyn, 1994, pp. 70–71). These two subsystems are interconnected, as well as having numerous projections to other areas of the visual system and to the frontal cortex.

**The model.** There are two different types of pictorial data structures in CaMeRa’s short-term memory. One type is an object data structure, which holds high-level information about the shape, color, and form of an object. The second type are spatial structures, which, unlike object structures, are quite small, containing only the spatial location of the object with which they associate.
3.9 In Perception, Object Is Separated from Ground

_Empirical evidence_. In order for an object to be recognized, it has to be separated from the background—one must be able to recognize and interpret a particular graph whether it is in a book or on the blackboard.

_The model_. To accomplish the separation of object from ground, the model employs rules derived from gestalt principles of perceptual organization. The model’s early-visual system together with the gestalt principles tend to uncover an object’s “nonaccidental” properties. A property is nonaccidental if it remains more or less unaffected by scale changes, rotation and translation (see Kosslyn, 1994, Chapter 5). The model’s early-visual system, a simple parallel network that does feature detection, uncovers points that are different from other points on the bitmap; the gestalt principles look for connections between points (this is discussed further below).

3.10 Objects Can Be Represented in at Least Three Distinct Reference Frames

_Empirical evidence_. Many researchers have studied how the world around us is represented in our mind. There is evidence that the reference frame for a percept may be (1) object-oriented, (2) background-oriented or (3) self-oriented. In the first case, the reference frame moves with the object (so that the object appears stationary); in the second case, it moves with the background of the scene, perhaps in the middle distance; in the third case, the frame moves with the viewer. In walking down the street one can experience these alternative frames by (1) focusing one’s gaze on a nearby object, (2) focusing it on the middle distance beyond the object, or (3) focusing it on the lower part of one’s body.

_The model_. In a model, a representation is object-oriented when coordinates of all points are fixed relative to the object in question. It is background-oriented if the coordinates are fixed relative to a background point; it is self-oriented if they are fixed relative to the viewer. In its present state of development, CaMeRa does not make these distinctions, but associates each object on the blackboard with a fixed background location. To extend its capabilities of simulation to situations where a person is moving in relation to objects and/or environment, capabilities would have to be added for transforming from one coordinate system to another.

4.0 CaMeRa: _THE EXPERT MODEL_³

Before getting down to the nitty-gritty detail of the CaMeRa model, we will conduct an overview of its main components and run through a general account of how it extracts information from an external blackboard or from verbal propositions and converts that information into a structure in pictorial short-term memory (pSTM).
3.18

Input through low-level visual perceptual processes

Output through motor processes

(1) External Display

(2a1) Visual Buffer

(2a2) Spatial Node-Link Relations

(2a3) Object

(2a) Pictorial Short Term Memory

(The Mind's Eye)

(2b1) Verbal Auditory Buffer

(2b2) Verbal Data Structures

(2b) Verbal Short Term Memory

(The Mind's Ear)

(3a) Pictorial Long Term Memory

(3b) Verbal Long Term Memory

Sound

(3a) Pictorial Long Term Memory

Figure 2. The architecture of CaMeRa. Pictorial and verbal representations are computationally different from each other, both in form and function, and thus are highly constrained in their interactions. Refer to text for details.

4.1 Anatomy of CaMeRa

CaMeRa consists of (1) a pictorial external display (Figure 2, left-hand side), (2a) pictorial and (2b) verbal short-term memory (STM) (Figure 2, center), and (3a) pictorial and (3b) verbal long-term memory (LTM) (Figure 2, right-hand side). The pictorial external display is a bitmap which represents the contents of a sheet of paper or a blackboard (we will use the term “blackboard” throughout the article to refer to this bitmap). The pictorial STM (pSTM), also called the Mind's Eye, consists of (2a1) the visual buffer (a second bitmap, Figure 2, left-hand side of pSTM), high-level (2a2) spatial and (2a3) object information structures (Figure 2, right-hand side, top and bottom, of pSTM), and (2a4) the productions (not shown in the Figure) that operate on them. The verbal STM Mind’s Ear (vSTM), similarly, is the union of (2b1) the verbal auditory buffer, (2b2) propositional list structures in STM, and (2b3) their related productions (not shown in the Figure). At present, the auditory buffer has not been implemented.

Verbal memory structures are modeled as propositional list structures, while high-level pictorial structures, in pSTM and pLTM, consist of node-link relations. Knowledge of both kinds is organized into small chunks, which can be connected by associations within or between modalities, but modified only from STM structures of their own modality.
CaMeRa combines a parallel network, used to process the low-level pictorial aspects of the visual buffer, with rule-based processes used for higher level pictorial and verbal reasoning. The rule-based processes depend on pictorial recognition cues (i.e., cues transmitted from external stimuli, via STM to pictorial LTM) and verbal memory cues (i.e., cues transmitted from STM to verbal LTM). The model does not use a goal stack, and its productions are not limited to any particular task domain but refer to basic geometric features of patterns.

The design of CaMeRa owes an important debt to Jeff Shrager's model of gas laser physics, which is based on representations that, while specific to different modalities, associate easily with one another (Shrager, 1990). CaMeRa also has some resemblance to ISAAC (Novak, 1977), which can solve physics problems by constructing and reasoning from drawings. ISAAC integrates previously learned information from several schemas in order to draw a picture either on the computer screen or internally, in the Mind's Eye. It temporarily assembles information as needed to continue solving a problem, and draws inferences from this integrated drawing.

Kosslyn's 1980 book, *Image and Mind*, outlines a model that accounts for interactions between long-term memory and visual short-term memory. The model focuses on low-level visual processes, such as "zoom" and "scan", and does not encompass interactions with perceptions. Our model, in contrast, is more concerned with higher-level problem solving processes: How they are influenced by perception, and the interaction of pictorial and verbal percepts and memories.

While many properties of the human representational system are still poorly understood, we have seen that a considerable body of evidence from both neuroscientific and behavioral experiments supports the Mind's Eye hypothesis employed in CaMeRa, and has guided the construction of the model. Symmetrically, the Mind's Ear hypothesis ties verbal short-term memory (vSTM) and its processes to the auditory and language centers of the brain, and implies that the representations created from percepts in the verbal working memory use the same medium, data structures, and processes as representations created from verbal Long-Term Memory (vLTM).

CaMeRa's semantic understanding is currently limited to topics related to the elementary economics of supply and demand, mentioned earlier. As we are primarily interested in visual cognition, the model, as now implemented, emphasizes the visual processes, and CaMeRa's verbal representations are not developed as elaborately as its pictorial ones. For example, we have not yet incorporated natural language processing capabilities in CaMeRa, as other architectures have (e.g., the Soar architecture: Lehman, Fain, Lewis, & Newell, in press; Lewis, 1993; Nelson, Lehman, & John, 1994; Rubinoff & Lehman, 1994; and the CAPS architecture: Just & Carpenter, 1992; Just & Thibadeau, 1984). Moreover, because we are examining expert reasoning, the knowledge that is used is assumed to be well established and does not require the system to learn, and STM capacity limits are not especially critical. All of these limitations will need to be removed as we extend the model to new tasks and to simulating the behavior of novice subjects.
Overview of CaMeRa's Performance

Storing verbal information. The vSTM and vLTM representations contain quasi-linguistic structures—propositional list structures associated with semantic information about real-world meanings and implications. These representations are able to interact with those of the pictorial systems, through STM, in a highly regulated and limited manner, so that, for example, the propositional information can be used to help interpret the mental image. This is an important feature of CaMeRa, which we will discuss in more detail later in the article.

Forming a mental image. As we have just seen, CaMeRa contains representations of a pictorial external display (the "blackboard"), pictorial short-term (pSTM) and long-term (pLTM) memories and verbal short-term (vSTM) and long-term (vLTM) memories. It uses the external display just as the expert does, to draw upon, recognize from, reason from, and to refresh the contents of STM. Presented with a diagram (a bit map) showing one or more supply and demand curves for a commodity, CaMeRa will inspect the diagram and construct, in a combination of bit-map and list-structure representations, a corresponding mental image in pSTM that contains nearly the same information as the diagram. (We will show later how the expert carries out essentially the same process.)

At present, no capacity limits are placed on the pSTM information; this is a task for the future. pSTM contains referents associating visual with verbal information, while the diagram does not. Except for these associations with verbal memory, diagram and mental image then are very similar informationally but not at all equivalent computationally.

First stage of transfer to pSTM. The visual buffer is a geometrical bitmap representation of pSTM that simulates initial images generated by external percepts (light on the retina) and internal percepts (LTM and STM structures). The early filtering and feature extraction processes of perception take place in the visual buffer. This is analogous to computations performed in the primary visual cortex. We have modeled these processes in CaMeRa with a simple but highly effective parallel network.

Re-representation in pSTM. The node-link, or list-structure, components of pSTM are a higher-level representation of the information contained in the visual buffer, and consist of two substructures. One represents object form and the other object position. Both of these structures can be created from information in the visual buffer or in pLTM. pLTM information transferred to the high-level pSTM structures can also be projected onto the visual buffer, thus creating a complete mental image (i.e., one with both a bit-map and a list-structure component). This is done in order to facilitate complex visual reasoning, which requires both STM's parallel network and its node-link structures. Reasoning from the visual buffer can reveal information that was not readily apparent from the more abstract pSTM object form and position structures (another example of computational inequivalence of the representations).
An example. Let us see, in very broad terms, how the system performs a particular task which we will later compare with the behavior of the expert. The blackboard already has a supply and a demand line drawn on it, with an equilibrium point at their intersection [Figure 3a]. This information has also been stored in the expert's mental image (visual buffer). Successive stages of information on the blackboard are shown in the left panels of Figures 3c and 3d, and information in the visual buffer in the middle and right panels (object form information and object position information, respectively).

The expert wishes to explain why the price, if it were currently above the equilibrium price, would be driven down to that price. He does this by considering (in a pictorial representation) a higher price, showing that at this price there would be a surplus (greater supply than demand), and then reasons verbally that the price would decline. [The appearance of the blackboard at this point is shown in Figure 3b.] Thus, he uses both pictorial and verbal representations in his explanation.

To be more specific, before the explanation just described has taken place, CaMeRa's blackboard has the appearance of Figure 3c. In its Mind's Eye CaMeRa marks a price, p, on the y axis above the equilibrium price, p, and draws a horizontal line through it. It then transfers this information to the blackboard representation (via a "motor" command), causing it to be projected onto the visual buffer. There the parallel network detects the intersections of this new line with those already present, and transfers the information about intersections back to the Mind's Eye. The flow of information in this example is thus high-level pSTM → blackboard → visual buffer → high-level pSTM.

Next, CaMeRa searches the Mind's Eye to find the intersection of the new line with the supply line, and draws internally a vertical line to find the x-coordinate of that intersection, which it labels "quantity supplied at p'." After this, CaMeRa transfers the vertical line and label to the blackboard. These constructs are projected into the visual buffer, where the parallel network detects the intersection of the new vertical line with the x-axis. This intersection is represented in the high-level pSTM, and a command is sent to read the x-coordinate (located at this point).

CaMeRa then repeats this entire sequence of perception and reasoning for the intersection of the price line with the demand line, labeling the new coordinate "quantity demanded at p'." Now the system is able to focus, in the Mind's Eye, upon the horizontal segment joining the two points that have been analyzed, notices that the supply quantity is larger than (to the right of) the demand quantity, and, transferring this information to vSTM, reasons verbally that, as supply exceeds demand, there is a surplus. (The reasoning is based on knowledge in LTM about the meaning of "surplus.") It then continues to reason verbally that when there is a surplus the price will drop, and that this process will continue until equilibrium is reached (i.e., at the equilibrium price, supply equals demand and there is no surplus). [CaMeRa's blackboard now looks like Figure 3d.]

Features noticed and recognized on the blackboard, combined with the heuristic that new features focus attention, drive the problem solving effort without need for an explicit goal stack. Recognizing information in the external display involves a sequence of events: first matching information on the blackboard to information stored in pLTM, and then copying the pLTM information into the corresponding pSTM structure. In the economics
Figure 3. The expert's blackboard and CaMeRa's visual buffers before and after "surplus" reasoning. The left-hand component of the visual buffer displays information about objects; the right-hand component displays information about their spatial arrangement (the coordinates of points, of labels, and of the ends of lines). The inner circle in each buffer represents the fovea, the area within which details can be recognized. The outer circle in each buffer shows the boundaries of the focus of attention. Refer to the text for details.

application, the only information other than the blackboard contents that is needed to move the problem solving forward is the initial problem statement. This type of control architecture is advantageous because it does not place heavy demands on STM capacity. Adding an explicit goal stack would only consume more space in STM and slow the rate of problem solving.
4.3 Technical Specifics of Programming

The reasoning part of CaMeRa is written in OPS5, a production system language built on Common Lisp. For an extensive overview of what a production system is and does, see Newell and Simon (1972). An OPS5 production system model consists of data structures, production rules, and conflict resolution rules (Cooper & Wogrin, 1988). The production rules, defined as condition(s) → action(s) statements, may have varying degrees of complexity. External image-processing and computation routines, written in LISP, are used on the “action” side of certain productions in CaMeRa in order to augment the capabilities of OPS5. Productions carry out either verbal or pictorial actions, but not both simultaneously.

The parallel network (LISP-based) which is used to perform the feature extraction on CaMeRa’s visual buffer sends the results of its computations to the OPS5 system where they are further processed by the appropriate high-level rules. In addition, the reasoning system of CaMeRa may activate the parallel network through its external image processing routines (e.g., drawing a line on the blackboard), and cause it to scan a certain area of the visual buffer and return the corresponding results.

4.4 Memory Systems

All the memories within CaMeRa are contained within the same architecture, though each has its own representation. The relations among the different layers of memories (LTM, STM, external display), and the different modalities (pictorial, verbal) within each layer are depicted in Figure 2. Representations within the same modality have more in common than do the representations of different modalities (e.g., pLTM and pSTM resemble each other more closely than pLTM and vLTM). Different sensory modes of the model may associate (referents: dashed arrows on Figure 2) with one another within a memory layer (e.g., pSTM to vSTM). These associations maintain a link between related information. For instance, in the example given in the previous section, the verbal concept of surplus and the corresponding propositions about surplus in vSTM are associated with the horizontal line segment in pSTM in which the supply coordinate is larger than the demand coordinate. The system, having used diagrammatic reasoning to detect the surplus, transfers this information to vSTM, which continues the analysis verbally.

However, modifications (solid arrows on Figure 2) are initiated only by STM, and only to memory systems of its own modality. For example, pSTM can modify pLTM directly, and the blackboard, but not vSTM or vLTM. Modifications of the latter require a sequence of further steps. This architecture accommodates variance in the input formats of different sensory systems by restricting direct modification between memories of different modalities. These restrictions ensure that events involving data-structure/operator incompatibility do not occur (such as visual operators modifying pieces of verbal information and leaving them in a visual format unreadable by verbal operators).

4.4.1 Long-term Memory (LTM)

LTM is essentially a database of information objects. Its contents can be increased, modified and deleted only through the STM of the same modality. Knowledge obtained from
LTM is physically represented in STM, as opposed to being linked with LTM only through associations. Nor is STM the activated part of LTM (as it is in the ACT systems of Anderson, 1985)—it is a separate entity. We separate STM from LTM for several reasons. Defining STM as the activated part of LTM would cause the modification of LTM information during reasoning. An important consequence of this would be to destroy valuable knowledge stored only in LTM. It is difficult to see how a system that constantly erased the information it had struggled to learn could survive for any significant length of time.

A final reason for separating LTM from STM is the format of the stored material. In our model, knowledge in LTM is not represented in the same format as knowledge in the visual buffer (which is one part of pSTM; the other part of pSTM, consisting of node-link structures, does have a similar representation to pLTM). This difference in storage formats yields differences in how the information may be used in reasoning. For example, an image in the visual buffer yields its geometric relations more readily than the equivalent information in LTM.

Both the pictorial and verbal LTM objects have associations to other LTM structures of their own modality, and to relevant objects in other modalities. Both LTMs also indicate whether the represented object is unique or may exist in variants of its original LTM form. This property allows the system to create STM exemplars of an instantiated general object, and to know that these are multiple instances of the same kind of object, by virtue of the fact that they associate to the same “mother” structure in LTM.

**Pictorial LTM (pLTM).** pLTM representations are image-like. We assume that the pSTM representations generated from LTM are computationally equivalent to those generated from perception. Any inferencing that one could do on the image from paper, one could do on the representation of the same information from LTM. This is, of course, limited by the rate at which information can be sampled from each source, which is generally much higher for perception (a parallel process) than recall from LTM (a serial process).

The drawings CaMeRa makes on the blackboard while reasoning are graphs made up of various relations between lines and labels. In pLTM, these lines are represented, for convenience, as mathematical equations (slope, intercept, etc.). Labels are represented as vectors of bits. These pLTM representations are compressed versions of the lines and labels that would exist in the visual buffer (that is, they consume significantly less storage space). The equations used to define lines in the pLTM are scaled to an internal coordinate system as opposed to an external one like retinotopic space.

In order to understand spatial relations, CaMeRa must project the LTM structures onto the visual buffer of pSTM. This is done by first representing the pLTM data in high-level pictorial STM structures (like the one shown below in Figure 5), and then using projection functions to display this information in the visual buffer. The projection functions are processes that translate compressed images (like equations) into full images on the buffer. The middle panels of Figures 3b and 3c depict the projections onto the buffer of object forms; the right-hand panels, the projections of object positions (i.e., the coordinates of “interesting” points like intersections, ends of lines, and labels).
The mathematical relations stored in pLTM are like schemas. The variables of such equations are represented in pLTM structures with default prototype values. For example, the supply line is represented in pLTM with a generalized slope of one and an intercept of zero, so that, from it, different supply line exemplars (in this case, linear equations) may be created. The relation for linear growth would be represented in a different pLTM structure than the relation for exponential growth. Mathematical distinctions of this nature are the types of things that a novice student (or a novice CaMeRa) would have to learn over time.

Exemplars in pSTM are represented as modified copies of the original pLTM relation, with the same or different parameters. Particular supply line instances might differ only by their intercept, or some might be straight lines and others curved. The mathematical form of these structures is a computational convenience; we do not believe that the brain represents objects as mathematical equations. CaMeRa's representations are intended as a reasonable approximation of what is going on, sufficiently veridical to allow us to model all of the relevant cognitive processes. The proof of this particular pudding lies in the success or failure of the simulation attempts.

Figure 4 shows an example of the pictorial and verbal LTM structures and their connectivity. The supply line is represented in pLTM as a linear equation with a slope of 1.0 and intercept of 0.0. An association links the pLTM structure to the analogous structure in vLTM. This association is also a part of LTM—it is information which has been learned over time. The related structures of different modalities must be bound together before problem solving can proceed correctly. pLTM structures can associate themselves with other pLTM structures: the pLTM representation of the supply line in Figure 4 would also have a link to the pLTM representation of the equilibrium point (not shown).

**Verbal LTM (vLTM).** A single generic form of propositional relation sufficed to represent all of the verbal knowledge needed by CaMeRa in solving the simple economics problems we used.

conceptC is the typeR relation between term1 and term2

Figure 4 also illustrates the supply line in its vLTM representation. The vLTM structure represents the proposition, "Supply is the Proportional relation between Price and Quan-
tion, " and has a link to the corresponding pictorial representation in pLTM. The central idea behind this form of verbal representation is that knowledge is represented in the brain as associations among symbols; structure and hierarchy are produced by these associations. Different novel structures can be formed, depending on what concept one starts from. Because connections are one-way (A → B does not necessarily imply B → A), the concept one begins with will define the network that one can elicit from vLTM by following its associations. As with the pLTM representations, vLTM structures can associate themselves with other vLTM and pLTM structures. For example, the vLTM representation of the supply line shown in Figure 4 would also have a link to the vLTM representation of the equilibrium point (not shown).

4.4.2 Short-term Memory (STM)

As all problem solving, computation, and high-level cognition takes place in STM, a production must obtain the data that it needs to satisfy its firing conditions from STM. If some critical piece of information is not present, then other productions must recognize this state and act to load the relevant data into STM from LTM, the blackboard or another external source.

It is well established that both vSTM and pSTM have limited capacity. For instance, Baddeley has proposed that vSTM capacity is the number of elements that can rehearsed before they decay, and the duration of this "articulatory loop" has been estimated to be about two seconds (Baddeley, 1986; Zhang & Simon, 1985). CaMeRa's STM currently does not have such capacity limitations. Because the expert's knowledge is so well rehearsed, and the problems are so simple, STM capacity plays an insignificant role in the problems we have examined. When we extend CaMeRa to account for novice behavior, the limits of STM capacity will become critically important and will be incorporated into the model.

STM structures can contain all of the kinds of information that the LTM structures of their modality can contain. They are defined as specific exemplars of LTM structures or as unique objects. STM structures may also contain associations to other STM structures in their own and different modalities.

The Mind's Eye. The Mind's Eye represents the synthesis of the visual buffer, object structures, and spatial structures of pSTM, and the processes that work on these structures.

The visual buffer. The visual buffer, the physical location of mental percepts, is the area in which feature extraction and other low-level, highly parallel, visual processes occur. It contains both perceptual images (the internal representation of the blackboard) and mental images retrieved from pLTM. Because it is represented as a multi-layered bitmap, its topographic organization has interesting similarities with the primary visual cortex. However, CaMeRa's visual buffer and the parallel network that operates on it make no claims to be a neurophysiological model of the visual system. We discuss the visual buffer further in the Perception section, below.
Object structures. Mind's Eye object structures hold higher level information about the shape, color, and form of objects. These structures contain all of the information held in pLTM, and in addition are able to represent new material not contained in LTM (i.e., novel objects). They are defined either as exemplars of LTM knowledge, or as new non-LTM data. For example, a cube rotated at an angle would be represented as an exemplar of a prototype cube stored in LTM. A picture of something which had never before been seen or imagined would be represented as a new object not classified in LTM (e.g., a cube, seen for the very first time). CaMeRa's pSTM objects are fairly large structures, containing numerous properties that allow them to interact with the visual buffer, associate with vSTM and pLTM, and build representations of newly acquired objects.

Spatial structures. Mind's Eye spatial structures, unlike object structures, are quite small, containing only the spatial location of the object with which they associate. They provide a spatial representation of the external environment, where the coordinates refer to retinal space (in contrast to the internal coordinate system used to determine the equations of lines represented in pSTM object structures and pLTM structures). Perhaps coincidentally, the size difference between the object pSTM structures and the spatial structures in CaMeRa is also found in the brain—the inferior temporal cortex (objects) contains a significantly larger number of neurons than does the posterior parietal cortex (spatial) (Kosslyn, 1994, p. 84; c.f. Van Essen, 1985).

Processes. Both the Mind's Eye object and spatial structures associate to the vSTM, pLTM, and to each other. The connections to pLTM link the object to its LTM prototype. The association to the vSTM provides the object's verbal name and label. The association to other Mind's Eye object structures enables the formation of complex pictorial constructs like intersections, in which the main Mind's Eye object represents the intersection point, and the associated Mind's Eye objects represent the component lines that define the intersection.

Figure 5 illustrates how the system represents the shifted supply line while it is explaining issues of supply and demand. It is drawn by CaMeRa on the blackboard while explaining the effect of a new tax on the original supply schedule. pLTM and knowledge about the amount of the new tax are used to transform the generalized supply line stored in LTM into the appropriately shifted supply line shown in the Figure. This new line is represented by a schema in the high-level object structures of pSTM, and then drawn directly onto the blackboard. Its appearance on the blackboard causes an image of it to be projected onto the visual buffer, where the parallel network and gestalt rules act to send spatial information about the shifted supply line to pSTM where a spatial representation is formed and linked to the object representation of the demand and supply situation that was created earlier.

The pSTM object structure has an association to the pLTM representation of the supply line (link not shown) that was depicted in Figure 4. Because it is a specific exemplar of the supply line, its parameters do not exactly match those in the generalized pLTM structure: the slope is the same but the intercept is transformed to represent the prototype supply line shifted upwards. The Mind's Eye object structure is closely linked to a Mind's Eye spatial
structure that gives the endpoints of the line in retinal space; this structure forms an association to a specific area of the visual buffer.

**Perception.** A parallel network which performs feature extraction, and gestalt principles used to scan the visual buffer for structure and form are the main components of the perceptual system.⁷

*The parallel network.* The network functions by perceiving non-modal areas (discontinuities) in the local topology of the object. As all connections between pixels on the bitmap are presumed to be of equal, positive weight, no training is necessary. By detecting points that depart from the local modal value, the network identifies all perceptually significant areas on the visual buffer (similar to the method described in Marr, 1982). This process is depicted in Figure 6.

The first layer of the visual buffer represents information projected directly into the brain from the external world. Pixels are either on (activation = 1) or off (activation = 0). The first filtering process builds a representation of the topology of the object by summing the activations of all the units surrounding a given unit $U_i$. This summation is the activation of $U_i$ in layer 2. Points in clusters mutually increase each other's activations, while isolated points retain only their initial activation. For example, if the input was a 3 x 3 square, the activation of the center of the square in the second layer of the visual buffer would be eight (one for each of the eight units surrounding the center point).

Processes in the second layer determine which areas of the entire topology are the most significant, by calculating how similar the activation of unit $U_i$ is to the average activation of its regional topology. The function $A(n, U_i)$ gives the average activation of the local area around $U_i$ (a square array of length $n$). $U_i$ is activated in the third layer if:
Layer 1

Layer 2

Layer 3

This figure depicts the processing sequence for any pixel \( U_i \) in the first layer of the visual buffer

- **=** \( U_i \)
- **=** units which project to \( U_i \) in the next layer

\[
f(U_i) = 1 \text{ if pixel } U_i \text{ on}
0 \text{ if pixel } U_i \text{ off}
\]

\[
f(U_i) = \sum_{n} \left[ A(n, U_i) \right] - 1
\]

where \( A(n, U_i) = \) the average activation of \( n \) units surrounding \( U_i \)

\[f(U_i) = 1 \text{ if pixel } U_i \text{ on}
0 \text{ if pixel } U_i \text{ off}
\]

If the activation of point \( U_i \) exceeds a certain threshold in layer 3, then \( U_i \) is sent to other processing areas of the visual system as a perceptually significant point.

Figure 6. Topology of the parallel network for feature extraction.

\[
\left| \left\{ \frac{U_i}{A(n, U_i)} \right\} - 1 \right| > t \quad \text{where } t \text{ = the firing threshold}
\]

Consequently, if the activation of \( U_i \) in layer 2 is very close to the activation of its local area, \( U_i \)'s activation in layer 3 will be zero—it is not perceptually significant as it can not be distinguished from the points surrounding it.

The result of this process is a mapping in layer 3 of all the perceptually significant features found in the image on the visual buffer, for example, intersection points, line endpoints, labels, object outlines (see Figure 6). The coordinates of these points are sent to the higher-level object and spatial representations of the pSTM for further processing. Productions may then request the visual buffer to identify a specific feature by matching the fovea-sized area around the point to patterns stored in pLTM, or to scan and recognize a set of associated objects. In order to do this, the model turns to the gestalt principles of organization.

The gestalt principles. CaMeRa employs the gestalt principles of Good Continuation, Good Form, Proximity, and Familiarity (Goldstein, 1984). Gestalt rules related to motion, convexity, etc., were not required in our tasks. We implemented the gestalt principles serially, rather than in parallel, because the serial design captured the relevant cognitive principles in a lucid and explainable manner. To coordinate the four gestalt rules, the model first saccades to the closest perceptually significant point on the visual buffer. This is the point in layer 3 which is the shortest distance from CaMeRa's current focus of attention. It then determines whether the small area of CaMeRa's fovea surrounding this point contains a familiar pattern. If the fovea's contents match a pLTM pattern, action is taken. For exam-
ple, recognition of a line segment evokes Good Continuation and Good Form. This causes the model to focus its attention on the nearest endpoint of the line, and then scan the entire line through a series of short saccades, following the simplest connected path. We limit recognition to the fovea because this is the area of the retina which has the highest visual resolution. Visual information extracted from regions beyond the fovea is in general not detailed enough to perceive the high-frequencies which give an object its defining features.

Mozer, Zemel, Behrman, and Williams (1992) have designed a parallel network model of object segmentation (based on the phase-locking of related features), which naturally produces the behavior of some of the gestalt rules we employ. Comparison of the behavior of the two systems suggests that many of the differences between our serial design and his parallel one may be only implementational and not functional. In further support of this point, although CaMeRa was not intended to segment images, it is able to do so, by virtue of the architecture of its visual buffer and the processes that operate on it. CaMeRa is able to discriminate geometrical objects (squares, diamonds, triangles, etc.) that are overlaid on each other (so far we have tested up to four overlaid objects).

**The Mind’s Ear.** The Mind’s Ear is the combination of the auditory buffer (not yet implemented), vSTM data representations and the productions that operate on them. The propositional list structures lie at the core of vSTM. The Mind’s Ear represents propositions of the type described earlier: \texttt{concept} is the \texttt{type} relation between \texttt{term1} and \texttt{term2}. In addition, the Mind’s Ear contains associations to pSTM and vLTM. vSTM can create exemplars of objects from vLTM, and can represent novel objects not contained in LTM. The association to the object representation in pSTM provides the denotation of the object.

Although the Mind’s Ear representation is fairly complete, the task we have given the system did not require any verbal external input except for the initial problem statement, which we stored directly in vLTM. Our expert solved the economics problems in a highly pictorial fashion, and did not use verbal external storage devices (writing down propositions, or the like). Such devices would, however, be very important in a novice learning situation, where verbal feedback from the instructor and reviewing written materials would play critical roles in the learning process.

The vSTM structure shown in Figure 5 is a representation of the supply schedule, and associates itself with the appropriate pSTM object structure for the Supply Line 2. The vSTM representation also contains information that denotes it as a specific exemplar (instance 2) of the generalized schedule in vLTM to which it is linked. The vSTM structure of Figure 5 would also have a connection to a pSTM spatial object, representing the visual buffer coordinates of the label. Both the expert and CaMeRa draw the label when it is needed for referring to a certain pictorial object. A label provides a shortcut from the pictorial to the verbal representation in LTM. A pictorial object in the Mind’s Eye is labeled by associating a verbal name with that object and drawing an icon representing that name either on the blackboard, the visual buffer, or both.

The representation in Figure 5 allows CaMeRa to understand what the Supply Line 2 is, where it is located on the blackboard, and what it looks like. This information can then be used for inferencing or to drive the problem solving process onward.
**Pictorial External Display (Blackboard).** The blackboard (or any external display) has two functions. One is to provide cues to items stored in LTM; these cues take the forms both of simple features, like line segments, and emergent properties like intersections, removing, as we observed, the need for a goal stack. In addition, the blackboard also serves as an external memory system for information that must be available immediately upon demand, and as a placeholder in the reasoning process by recording steps of the process dynamically as they unfold. The blackboard may contain both recognized and unrecognized (novel) structures.

In our model, we assume that information on the blackboard is constantly being projected onto the visual buffer, and is always available to CaMeRa for examination. When something is drawn on either the visual buffer or the external display, the processes of visual perception immediately extract and analyze the new salient features. CaMeRa, like the expert, continually adds elements to the blackboard during its explanation, and uses this display as a record of how far the explanation has progressed.

Were the model to be interrupted midway in its reasoning and then later return to its task, the state of the blackboard would allow CaMeRa to resume its explanation. The contents of the blackboard would remain intact even though CaMeRa's STM had been erased during the interruption. The processes of visual perception would allow CaMeRa to extract quickly the relevant cues from the blackboard and reconstruct its position in the reasoning sequence, allowing problem solving to continue from the point where it had left off. This is much faster than starting the problem over, because the perceptual processes used in the cue extraction are much faster than the processes of visual reasoning that were used initially to generate the contents of the blackboard.

CaMeRa also looks to the blackboard for input when problem solving has stalled, and it needs geometric or additional pictorial information to move beyond an impasse. If no such information is needed, and it can rely purely on the high-level object and spatial representations currently residing in pSTM, it will not access the blackboard or visual buffer. In the future, when STM capacity limitations have been implemented, the images on the blackboard will be used to refresh the visual buffer and the high-level pSTM structures.

**Verbal External Display.** The verbal external memory, not yet implemented, would input the initial goal statement to the system. We assume that this statement would be read and represented internally by a procedure like that used in existing language processing systems. Parsing procedures contained in the language system would create propositional structures in the form that CaMeRa understands. It is after these processes have done their work that CaMeRa, at its present stage of implementation, takes over. A later stage of the model will leave the initial statement in the verbal external display, and gradually build up the vLTM from it, via the vSTM, by referring back to it—as the subjects actually do. A model of this general kind, implemented within the Soar framework, can be found in Polk (1992).

### 4.5 Reasoning and Inference

We have already provided a number of examples that show how the CaMeRa architecture is used in reasoning. All reasoning and inference involve collaboration between the Mind's
Eye, the Mind’s Ear, and LTM. If information for a computation is not available in one or more of these systems, then the blackboard is consulted. Each element of the expert’s reasoning is linked to the next by LTM, blackboard cues, and the inference rules operating on them. As we have seen, the dynamically created graph on the blackboard is thus used to continue reasoning and to summarize past reasoning, so that the system is interruptable. For instance, as soon as supply and demand lines are drawn, their intersection is noticed, thereby cueing LTM, through the Mind’s Eye, to begin to reason about the equilibrium (see Larkin, 1989 for more research on how people handle interrupted processes through use of visual cues).

The reasoning employed by CaMeRa is called immediate reasoning, as that term is used in the literature on situated action (Vera & Simon, 1993), where the environment interacts in an integral way with the reasoning process (Agre, 1988; Suchman, 1987). Another system that reasons in this way is Vera, Lewis and Lerch’s (1993) model that learns how to use an automated teller machine—the various buttons and labels on the machine cueing the behavior. As in our model, their system builds a memory consisting of small substructures and depends heavily on recognition, using the external cues to reconstruct the actions needed to operate the machine. Neither model needs a structured plan. Vera et al.’s model provides a more detailed account of linguistic processes than does the system described here, but a less detailed account of pictorial processes. Both systems provide psychologically rooted examples of symbolic systems that operate in a situated way.

4.5 Productions

In this section we will provide some final examples of CaMeRa’s productions by showing how a complete representation of a line would be constructed, and by creating an equilibrium point. The productions do no error checking. That is, although it would be possible to ensure in multiple ways that the memory elements used to satisfy the conditions of the production contain the appropriate information, the model makes only the minimal requisite checks. This appears to be consistent with human processes—people assume that knowledge is associated correctly with other knowledge, and only search actively for errors when something appears inconsistent to them (Tabachneck, 1992).

Using the example of Figures 3 and 5, consider CaMeRa’s task of drawing the new supply line while it is explaining a shift in the supply schedule. Previously, the problem statement has been sent to the Mind’s Ear, processed, and transferred to the Mind’s Eye. The model has used this description to produce an initial graph containing a demand line and supply line, and it has explained various simple economics principles, all primarily on the basis of the information stored in pLTM and vLTM. The pLTM information includes the equations of generalized supply and demand lines, the axes of the graph, and representations of equilibrium point, surplus, and shortage. The vLTM knowledge includes the meaning of the supply and demand schedules, and the concepts of equilibrium, shortage and surplus. Using this information and its drawing abilities, CaMeRa can produce an initial graph and deduce various concepts such as “at price X₁ there is a state of surplus”, and “at price X₂ there is a state of equilibrium,” etc.
The first step in drawing the shifted supply line is for CaMeRa to represent the line as an object in its pSTM using the pSTM draw_line production (see upper-right-hand side of pSTM in Figure 5). This causes a node-link object structure for the Supply Line 2 to be created, representing the equation (with shifted intercept), and a link to the pLTM. Second, using the vSTM production, label_line, the new pSTM representation is labeled by creating a vSTM structure for the Supply Line 2 (right-hand side of Figure 5), and an association formed between this structure and the node-link pSTM object created moments before. At this point, the Supply Line 2 is not yet depicted on the visual buffer or the blackboard. It has only been "drawn" in a high-level pSTM object structure.

CaMeRa has now associated the verbal semantics of the supply schedule with the supply line, an internal label, but has not yet drawn a physical line or label on the visual buffer or on the blackboard. Next, CaMeRa draws the Supply Line 2 on the blackboard using a motor routine (motor_draw_line). This causes the line to be projected onto the visual buffer (left-hand side of pSTM in Figure 5), and subsequently causes the creation of a pSTM node-link structure (through parallel network, gestalt, and high-level processes) for the spatial location of the supply line 2 on the visual buffer (lower-right-hand side of pSTM in Figure 5). A similar processing sequence produces a spatial representation of the label and its relevant associations. This spatial structure is immediately linked (vertical arrow) to the pSTM object structure for the Supply Line 2, mentioned earlier.

![Diagram](image)

**Figure 7.** A more complete representation of the Supply Line 2 in CaMeRa. Both LTM and STM components are shown. Solid lines indicate that modifications may take place between the two associated structures, while dashed lines only allow for information to be accessed.
The end result of this reasoning sequence is the representation shown in Figure 7, containing the STM and LTM structures and the relevant connections between them (the spatial pSTM structure for the label and its link to the vSTM structure that represents the semantics of the line is not shown). As mentioned in the previous paragraph, drawing the shifted supply line on the blackboard activates the parallel network and its low-level feature extraction processes. A number of new intersections will then be noticed and their coordinates sent to the productions of pSTM in order to determine if they require further analysis. This will drive the reasoning process onward and eventually allow it to locate the new equilibrium point and identify the surplus and shortage conditions.

A more complex representation is produced during the explanation of why the intersection is an equilibrium point. Using the processes described above, the basic graph, consisting of axes, supply line, demand line and labels, is constructed. The topology network observes a new intersection point emerge as the demand line is drawn and intersects the supply line. This point is passed along to the high-level pSTM representations, where it is further processed and identified as the equilibrium. A chain of inferencing procedures determines the coordinates of the equilibrium point and the states of the graph that occur above and below it. The result is the highly elaborated equilibrium representation, shown in Figure 8. Certain symmetries are immediately apparent. On the left, the pictorial and

Figure 8. The complete representation of the equilibrium point in CaMeRa. All structures are shown as circles; the shading of each circle denotes the structure's modality and memory type. Adjacent circles which are physically touching each other have an association between them. Solid lines also indicate associations. Note that this diagram does not specify whether or not the associations are strong (and allow for modifications) or weak (and only allow for information access)—the purpose of this figure is to make clear the symmetries that appear in the complex representations constructed by CaMeRa. Refer to the text for details.
verbal structures that represent the supply line clearly parallel and complement one another. A similar relation can be seen on the right for the demand line. In the middle lie the component structures of the equilibrium concept itself, both pictorial and verbal.

The representation becomes even more complex after the expert reasons about why the economic system, if displaced from equilibrium, returns to it. Appendix 1 provides a comparison of the model's reasoning about this stability with the expert's protocol, using corresponding visual and verbal segments from the trace of the model and the protocol of the expert, along with a brief explanation of the matching. CaMeRa's processes during the performance of this task are quite similar to those described above for the task of inferring the consequences of a shift in the supply line, and the reader can check in detail which of these processes are reflected directly in the protocol, and which are not. Not surprisingly, internal perceptual processes—for example, recoding information from the internal bitmap to the Mind's Eye, and vice versa—are not verbalized because they are presumably not within conscious awareness.

Appendix 2 shows how the same reasoning in both tasks could, in principle, be accomplished verbally, without using the diagram. We leave to the reader the construction of a third, algebraic, representation of the arguments, which can be derived by a rather straightforward translation of the verbal axioms into the language of algebra, then solving certain pairs of simultaneous equations and noticing inequalities between quantities.

5.0 CONCLUSION

Our economics expert who is simulated by CaMeRa had, after long training and experience, distilled a model of supply and demand relations that exploits unique aspects of both graphical and verbal representations. Employing this model, he uses pictorial memory to draw relevant graph parts, to recognize significant graphical features and thereby drive forward the reasoning, and to reconstruct the record or summary of reasoning already accomplished that is embedded in the traces left on the blackboard. Verbal memory is used to supply causal reasoning, and verbal labels give semantic meaning to the various parts of the graph.

As the representation is being constructed, it provides cues for the next steps in reasoning. These cues jog recognition, supplying the grist for immediate reasoning and allowing the expert to work forward without an overall plan. For example, visual recognition of a shortage or surplus is cued by drawing a line parallel to the x-axis at the presumed selling price, which generates two intersections, one with the supply line and one with the demand line, and a horizontal line segment connecting them. The line segment cues the visual LTM, which, in turn, activates the appropriate verbal term ("shortage" or "surplus") for the new segment, as well as causal verbal reasoning to continue the explanation. Thus, the expert works forward, pushed in turn by verbal and visual memories, actively constructing the whole account from small, basic elements.

CaMeRa, the model we have built to account for the expert's reasoning, consists of an integrated production system and parallel network. It contains representations of a pictorial external display, a pictorial short-term and long-term memory, and a verbal short-term and
long-term memory. CaMeRa uses the external display as the expert does, for drawing, reasoning, recognition, and input to STM. CaMeRa shows that the use of different representations for different modalities still permits communication between the diverse memories. Even though a memory can only be altered directly by the processes that belong to its own representation, associative links can be forged between different representations.

The next steps in developing further the capabilities of CaMeRa are to stimulate STM limitations, extend the verbal component of the model, increase the perceptual sophistication so that it can read and segment more complex diagrams, and incorporate learning capabilities. We expect that CaMeRa's successor will enhance the already considerable reasoning abilities of the current model, and will be able to solve problems in several domains as well as learn basic principles as a novice would in new domains.

CaMeRa demonstrates that reasoning need not be exclusively a verbal process, for inferences are also drawn by manipulating visual images. In the economics situation we have been examining, observing that a supply curve and a demand curve intersect with \( x \) and \( y \) coordinates, \( q \) and \( p \) respectively, cues the conclusion that supply and demand will be in equilibrium when the price is \( p \) and the quantity bought and sold is \( q \). Similarly, observing that the horizontal line corresponding to a price, \( p' \), intersects with the demand curve to the left of its intersection with the supply curve cues the conclusion that at price \( p' \), there will be a surplus of the commodity. (Verbal reasoning will then infer that the price will be driven downwards.) Scientists (e.g., Einstein) have often insisted that they do not "think in words." CaMeRa implements an important kind of thinking, using images, that is not verbal.

A second major lesson from the experimental data is that the use of a pictorial display to understand or explain a situation calls upon much more than basic, task-independent, perceptual processes. Presumably, the eyes of novices are as able as those of experts to detect lines, their intersections and their coordinates. What distinguishes the experts from the novices is not the sharpness of their vision but their possession of task-related knowledge that enables them to recognize important features of the display, and having recognized them, to interpret them meaningfully in the context of the problem domain. The model described here allows us to see how this interaction between perception and previous knowledge, as well as the interaction between different perceptual and memory modalities, can take place in a system possessing functional properties and organization consistent with what we know today of the human brain.

Finally, CaMeRa is a new member of the small family of computer architectures that illustrate by actual implementation the kinds of symbolic processes that are capable of extracting information from different sorts of representations, transforming one representation into another, and thereby exploiting the special computational powers that each mode of representation contributes to enhancing human reasoning.

Acknowledgment

This article is a joint product, all of the authors being involved in all aspects of it. The project originated in the experimental and theoretical work of the first author's doctoral dissertation. The second author played a principal role in both the development and the
implementation of the computational model. All three authors share equally the responsibility for the final product.

Preliminary versions of CaMeRa were discussed at the 1994 and 1995 Cognitive Science Conferences (Tabachneck, Leonardo, & Simon, 1994; Leonardo, Tabachneck, & Simon, 1995).

This research was supported in part by the James S. McDonnell Foundation, program of Cognition in Education, Grant No. 92-05.

NOTES

1. Relying on the precedent of Ebbinghaus, who used himself as a subject, the third author of this article served as the expert in economics in these experiments. On both occasions of giving a protocol, he was not warned that one would be elicited, hence could make no advance preparation. The expert protocols were obtained after the basic structure of CaMeRa had been fixed, whereas the protocols of novice behavior were gathered before modeling work began. All the protocols, like the other evidence cited in this article, should be interpreted as guides to the construction of the model and tests of its sufficiency to produce the basic phenomena that had been observed, rather than as conclusive tests of its accuracy or generality.

2. After the difficulty he experienced in the experiment, the expert, with considerable additional work later in the day, was able to put down on paper an explanation completely void of visual references. He found this not to be an easy task, and he reports that he found himself unable to avoid using his own mental imagery to guide his thought while carrying it out.

3. CaMeRa is available on the World Wide Web. You can run the model yourself and set some of its parameters. The address is: http://www.cs.cmu.edu/afs/andrew.cmu.edu/user/13/al28/camem.html

4. This bitmap is currently set to 150 x 150 pixels. However, we make no claims as to the relation between pixels and neurons.

5. It is also difficult to see how exemplars of even simple objects could be built or modified in LTM in real time (Simon, 1979, estimated it takes about 1-10 seconds to build an LTM element). However, see the discussion of retrieval structures, which permit rapid processing in LTM under strictly defined circumstances, in Richman, Staszewski, and Simon (1995).

6. We intend no implication that the actual memory structures correspond to equations; we simply have not undertaken a veridical representation at this grain level, as almost nothing is known about it physiologically. See further discussion of this point herewith.


REFERENCES


APPENDIX 1:
COMPARISON OF PROTOCOL WITH SIMULATION DURING EXPLANATION OF EQUILIBRIUM

The first column gives segments 43 to 55 of the expert's protocol while he was explaining why the intersection of the supply and demand lines was an equilibrium. The second column provides an analysis of the protocol in terms of processes, with interpolations during periods when the expert was silent. The third column provides a trace of the processes executed by CaMeRa. The goodness of fit of CaMeRa to the protocol can be assessed by comparing the processes inferred in the second column with the trace in the third column. The fourth column provides some explanatory comments. "ME" stands for "Mind's Eye."

Notice that on 18 occasions the verbal protocol, gestures, and drawing of the expert provide explicit evidence for the corresponding steps in the formal analysis. The remaining steps in the analysis were inferred as minimal requirements for information transfer from one internal modality to another.

During the intervals between protocol segments 47 and 48, and again between segments 48 and 49, CaMeRa carries out a whole series of (internal) noticing and labelling processes that identify the intersections and other features introduced by drawing the new price line, features that are used to derive the notion of surplus (e.g., that the intersection with the supply line is to the right of the intersection with the demand line). As knowledge about these features is already stored in the expert’s LTM, we can postulate that identifying (recognizing) them requires only a few hundred milliseconds each, so that each of the “long” intervals would be of the order of several seconds in duration, comparable to the intervals in which a single externalized action is carried out. The internal processes are not verbalized (something we would infer from the generally accepted theory of verbal protocols), and in this case, usually not even signalled by pointing gestures. Some other expositors of the same material might gesture more freely.

The verbal reasoning that evoked the idea of using a proof by contradiction (protocol segment 44) was assumed by CaMeRa to be already available in the expert’s memory rather than being constructed step by step on this occasion. The same is true of the reasoning from the surplus to the lower equilibrium price. Appendix 2 shows how both pieces of reasoning can be carried out verbally at the cost of considerable complexity that the diagrams eliminate.

An example will illustrate the method used to interpolate between the explicit protocol statements, drawing acts and gestures:

The first step in constructing the analysis column is to interpret the explicit protocol statements, including acts of drawing and gesturing. For example, statement 43, *Why do we think that’s where the market would stay?* becomes: *explain why equilibrium is stable;* and 44, *Because, suppose you have a higher price* becomes: *posit higher sales price than equilibrium price.* As protocol statement 44 is a contrary-to-fact conditional, we infer that it is motivated by the expert’s planning to make his argument by contradiction. That leads us to postulate two intermediate processes: The idea of proof by contradiction is evoked, which leads (1) to formulating, then setting, the subgoal of taking the system out of equilibrium,
which leads (2) to the idea of positing the higher price. Thus, the analysis postulates that
certain information in long-term memory must be evoked in order to lead the expert from
the action expressed in one explicit statement of the action expressed in the next statement.
These intermediate steps all correspond to internal processes. From our knowledge of
human recognition times, omitting an external response, we would expect each of these
steps to consume a few hundred milliseconds.

The protocol analysis uses seven “actions”:

(A1) V: Set Goal
(A2)–(A5): Retrieve further information on this topic, using existing connection from
one structure to another
(A2) V → V: verbal information (verbal structure to verbal structure)
(A3) P → V: verbal information (pictorial structure to verbal structure)
(A4) V → P: visual information (verbal structure to pictorial structure)
(A5) P → P: visual information (pictorial structure to pictorial structure)
(A6) P-recognize-emerging-features (features emerge as items are drawn on top of or
connected to each other, such as intersections; the expert recognizes them as relevant to
the reasoning)
(A7) P-recognize-lack-of-feature (such as absence of a referent)
<table>
<thead>
<tr>
<th>Protocol</th>
<th>Analysis</th>
<th>Model Run: CaMeRa</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(observed actions in italics)</strong></td>
<td><strong>(V-Verbal; P-Pictorial)</strong></td>
<td><strong>(V-Verbal; O-Pictorial Object; S-Pictorial Spatial)</strong></td>
<td><strong>Note that CaMeRa’s (0,0) coordinate point is in the left upper corner.</strong></td>
</tr>
<tr>
<td><strong>(direct evidence — bold)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43 Why do we think that's (A1) V-Set-Goal: explain why equilibrium where the market would stay?</td>
<td></td>
<td>(P1) HYPOTHESIS_GENERATOR Why is the equilibrium point the equilibrium point? Consider the relation between the quantity of the product supplied and demanded when the price of the product is larger or smaller than that of the equilibrium price.</td>
<td>The hypothesis generator is a set of verbal reasoning modules which posit setting a higher sales price in order to force the system out of equilibrium. These modules are already in the expert’s repository.</td>
</tr>
<tr>
<td></td>
<td>(A2) V→V: List the further verbal information on this topic: proof by contradiction: set goal to take system out of equilibrium</td>
<td></td>
<td>The model chooses to draw the higher price line first, and a lower price line later, just like the expert.</td>
</tr>
<tr>
<td></td>
<td>(A1) V-Set-Goal: take system out of equilibrium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44 Because, suppose you have a higher price</td>
<td>(A2) V→V: List the further verbal information on this topic: posit higher sales price than equilibrium price</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(A1) V-Set-Goal: set sales price higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(draws horizontal line above the equilibrium point)</td>
<td>(A4) V→P: List the further visual information on this topic: draw higher-price line above and parallel to the equilibrium-price line</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62. O:DRAW_LINE</td>
<td>Drawing the higher price line.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>63. V→O:LINK</td>
<td>Linking the visual object &quot;higher price line&quot; to the verbal object PRICE2, and labeling the line in the ME.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>64. S:DRAW_LINE</td>
<td>Drawing the higher price line on the blackboard.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(A7) P-Recognize-lock-of-feature: need referent for equilibrium-price line</td>
<td></td>
<td>NOTE: the model had found and drawn the label for the equilibrium-price line earlier on, when intersection was first drawn.</td>
</tr>
<tr>
<td>45 well, this is equilibrium price P</td>
<td>(A3) P→V: List the further verbal information on this topic: label P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(labels P)</td>
<td>(A5) P→P: List the further visual information on this topic: draw label P on blackboard</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Protocol | Analysis | Model Run: CaMeRa | Explanations
---|---|---|---
(A7) P-Recognize-lack-of-feature: need refer-ent for higher-price line | 65. S:LABEL OBJECT | Labeling the higher price line on the blackboard

46 call the higher price \( P \) prime (labels \( P' \))

47 then, at \( P \) prime (gestures to emphasize intersection \( P' \) and supply-line)

(A6) P-recognize-emerging-features: intersection of line \( P' \) and Supply-line (part of 64)

| Fovea (85, 65); Intersection: (87.0 63.0) |
| Fovea (75, 65); Intersection: (75.0 63.0) |
| Fovea (65, 65); Intersection: (63.0 63.0) |
| Fovea (17, 63); Intersection: (15.0 63.0) |

66. O→S:DETERMINE_INTERSECTION_LINES

| These internal steps are not reflected in the protocol: |
| Convert information from pixel array to ME object format |
| Determining what lines are part of an intersection. Point (15.63) is the intersection of line (0.0 30.0) (Price2) with the line ([INFINITE, 0.0] (Y-axis)). |

67. O:LINE_LINK

| Converting ME intersection component lines data for intersection [0.0 30.0 , [INFINITE 0.0] into pointers to relevant ME structures for those lines |

68. O:READ_Y_COORDINATE

| Reading [ME] Y-axis value as 30.0 for intersection of ME lines [ (0.0, 30.0) ([INFINITE, 0.0] ) |

Since we are ON the Y-axis, we can read off the Y-coordinate for this intersection, it's 30.
<table>
<thead>
<tr>
<th>Protocol</th>
<th>Analysis</th>
<th>Model Run: CaMeRa</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>86. O→S: DETERMINE_INTERSECTION_LINES</td>
<td>Isolating component line equations of pixel intersection point [87.0, 63.0] created by pixel line [0.0, 63.0]. Converting equations into ME object format... Component lines = ( [(0.0 30.0) (1.0 0.0)] )</td>
<td>Now seeing what lines belong to the intersection of P2 and (-1.0 50.0) (the demand line). The equations for those lines are converted to a format that can be used by the ME.</td>
<td><strong>NOTE:</strong> the model first elaborates on the intersection with the Demand line, while the expert first goes to the intersection with the Supply line. For the sake of brevity, we are eliminating the model's Demand line reasoning; it is exactly analogous to the Supply line reasoning.</td>
</tr>
<tr>
<td>87. O: LINE_LINK</td>
<td>Converting ME intersection component lines data for intersection [ 0.0 30.0 , 1.0 0.0 ] into pointers to relevant ME structures for those lines</td>
<td>Linking the ME structures of the demand line and P2 line to the ME intersection structure...and...finding the ME Y-coordinate of the pixel value of the intersection point of P2 and Demand line by interpolation - it's 30. Since we know the intersection of P2 and the Y-axis, we can infer the Y-coordinate of the intersection of the P2 line and the demand line.</td>
<td></td>
</tr>
<tr>
<td>88. O: INTERPOLATE_YCOORDINATE</td>
<td>Estimating (ME) y coordinate of pixel intersection point [ 87.0 , 63.0 ] to be 30.0 for intersection of lines [ 0.0 30.0 , 1.0 0.0 ]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Protocol Analysis Model Run: CaMeRa Explanations

(A3) P→V: List the further verbal information on this topic: need X-axis coordinate value

(A1) V→Set-Goal: find X-axis coordinate value

(A4) V→P: List the further visual information on this topic: to find X-axis coordinate value, project perpendicular line (Q3) from intersection to x-axis

89. O SEEK_X_COORDINATE

Initiating procedures to find the X-coordinate of pixel point (87.0, 63.0) for the intersection of ME lines [(0.0 30.0) (1.0 0.0)]. A VERTICAL coordinate line must be drawn through the point in order to read the axis value.

90. O DRAW_LINE

drawing VERTICAL_LINE with ME slope INFINITE and intercept 30.0 into mind's eye representation.

91. V→O: LINK

Linking ME coordinate line with Verbal STM structure QUANTITY 3 and labeling it with 43.

92. S DRAW_LINE

drawing a line with pixel slope INFINITE and intercept 87.0 onto external memory.

Fovea (89, 63) Intersection: (87.0, 63.0) (with supply line)
Fovea (87, 73) Intersection: (87.0, 75.0) (with P line)
Fovea (85, 85) Intersection: (87.0, 87.0) (with demand line)
Fovea (87, 133) Intersection: (87.0, 135.0) (with X-axis)

93. S LABEL_OBJECT

labelling ME spatial VLINE as QUANTITY 3

48 as you can see on the graph (gestures to graph)

(A6) P-recognize-emerging-features: intersection of line Q3 with X-axis

94. O→S DETERMINE_INTERSECTION_LINES

Isolating component line equations of pixel intersection point [87.0, 135.0] created by pixel line [INFINITE, 87.0] Converting equations into ME object format... Component lines = ((INFINITE 30.0) (0.0 0.0))

What intersection lines belong to the [87.0, 135.0] intersection?

It's "INFINITE 30.0" (Q3) and "0.0 0.0" (the X-axis)
<table>
<thead>
<tr>
<th>Protocol</th>
<th>Analysis</th>
<th>Model Run: CaMeRa</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>95. O:LİNE_LINK</td>
<td>Converting ME intersection component lines data for intersection [INFINITE 30.0, 0.0 0.0] into pointers to relevant ME structures for those lines</td>
<td>Linking these lines to the [87.0, 135.0] intersection point structure, and converting them to a format that can be projected on the ME. Reading the X-coordinate of that line off the ME, so in ME measurements. The coordinate is 30.</td>
<td></td>
</tr>
<tr>
<td>96. O:READ_X_COORDINATE</td>
<td>Reading [ME] X-axis value as 30.0 for intersection of ME lines [ (INFINITE, 30.0) (0.0, 0.0) ]</td>
<td>Thus, we are estimating that the X-coordinate of the point (87, 63), the intersection of P2 and the Supply line, is 30 as well.</td>
<td></td>
</tr>
</tbody>
</table>

(A3) P→V: List the further verbal information on this topic: This is the X-axis value of intersection of line P' and Supply-line

97. O:INTERPOLATE_X_COORDINATE | Estimating [ME] x coordinate of pixel intersection point [87.0, 63.0] to be 30.0 for intersection of lines [0.0 30.0, 1.0 0.0] |

Gestures to graph.

(A6) P-recognize-emerging-features: intersection of line P' and Demand-line

(A3) P→V: List the further verbal information on this topic: need X-axis value

As noted before, the section on the intersection of the Demand line with the P' line, and finding it's X-coordinate has been omitted for the sake of brevity. [Notice that both the expert and CaMeRa could have abbreviated this process by assuming, rather than showing, that "to the right of" on the diagram is equivalent to "larger quantity," and using the corresponding action to make the inference that supply exceeds demand at P'.]
<table>
<thead>
<tr>
<th>Protocol</th>
<th>Analysis</th>
<th>Model Run: CaMeRa</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A6) P-recognize-emerging-features: segment of line P' between intersections with Demand and Supply lines</td>
<td>98. O→S:SEE_X_SEGMENT Recognizing segment at ME points: 20.0 30.0 for intersection of line set :[-1.0 50.0 1.0 0.0 ]. Representing new segment in ME and EM</td>
<td>Now that we are aware of the intersections of P2 with the demand and supply lines, and know their X-coordinates, we can recognize the line segment between those intersections as significant. In the same production, we are representing the segment both in the ME and EM.</td>
<td></td>
</tr>
<tr>
<td>49 the supply at that price (A3) P→V: List the further verbal information on this topic: surplus</td>
<td>99. V→O:ANALYZE_SEGMENT Studying segment [ 20.0 30.0 ] The line HLINE intersects both DEMAND and SUPPLY at y = 30.0. The line VLINE intersects DEMAND at x = 20.0. The line VLINE intersects SUPPLY at x = 30.0. DEMAND is LESS [than] SUPPLY</td>
<td>Now we analyze the exact meaning of the segment in the verbal LTM. We find that the segment means that at this price (P2), demand is less than supply …</td>
<td></td>
</tr>
<tr>
<td>50 the demand at that price is much smaller (points at the intersection between P' and the demand line)</td>
<td>(A7) P-Recognize-lack-of-feature: need reference for surplus-segment (A3) P→V: List the further verbal information on this topic: label surplus (A5) P→P: List the further visual information on this topic: draw label surplus on blackboard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51 (labels difference with curly bracket and &quot;surplus&quot;) and so there would be a surplus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol</td>
<td>Analysis</td>
<td>Model Run: CaMeRa</td>
<td>Explanations</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>-------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>And we know what happens when there's a surplus</td>
<td>(A2) V→V: List the further verbal information on this topic:</td>
<td>100. V→O: RECOGNIZE_SEGMENT</td>
<td>... and that this state is called surplus which then leads to a string of verbal reasoning.</td>
</tr>
<tr>
<td>the sellers get anxious</td>
<td>a surplus makes sellers anxious, too many goods on hand</td>
<td>Recognize SEGMENT as SURPLUS based on previous analysis. When there is a surplus, the sellers want to lower the price in order to get rid of the excess of goods. This causes buyers to purchase more of the product and sellers to make fewer of the product. The process continues until the surplus is gone and the equilibrium has been reached.</td>
<td></td>
</tr>
<tr>
<td>and the buyers get fussy,</td>
<td>a surplus makes buyers choosy, product flooding market</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and so the price goes down.</td>
<td>(A2) V→V: List the further verbal information on this topic: sellers must get rid of goods, so they lower the price</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Effects of price changes.

1a. If price rises, then amount demanded decreases.
1b. If price falls, then amount demanded increases.
2a. If price rises, then amount supplied increases.
2b. If price falls, then amount supplied decreases.
3. If price is at equilibrium, then supply price equals demand price, and supply quantity equals demand quantity.

Surplus or shortage

4a. If supply quantity exceeds demand quantity (surplus), then price falls.
4b. If demand quantity exceeds supply quantity (shortage), then price rises.

Definitions of supply and demand schedule

5a. If supply schedule shifts downward then, for any P, amount supplied will be less. for any Q, the price will be greater.
5b. If supply schedule shifts upward then, for any P, amount supplied will be greater. for any Q, the price will be less.
6a. If demand schedule shifts downward then, for any P, amount demanded will be less. for any Q, the price will be greater.
6b. If demand schedule shifts upward then, for any P, amount demanded will be greater. for any Q, the price will be less.

Notice that the properties expressed by axioms 1 and 2, and 5 and 6 can be seen directly on the diagram as soon as positively sloping supply curves and negatively sloping demand curves are drawn. Axioms 3 and 4 are not directly shown on the diagram, but must be inferred separately. This can be seen from the expert’s protocol, segments 51–55.

APPENDIX 2

Verbal Reasoning from Supply Shift

We can now use these axioms to reason verbally that if a tax is imposed, and the supply schedule correspondingly shifts downward, the new equilibrium price will be higher than the old, but by less than the amount of the tax. The advantage of being able to “see” the relations between the old and new intersections of demand and supply schedules directly in the diagram is illustrated dramatically by this complex verbal argument.

Suppose equilibrium is at P1, Q1 and supply schedule shifts downward without a change in price then amount supplied will now be less than Q1 (5a) but demand will be the same as before
and there will be a shortage at the original price
the price will begin to rise (4b)
amount supplied will begin to increase again (1a)
amount demanded will begin to decrease (2a), both decreasing shortage
There will be a new equilibrium at a price higher than P1 (3)
(At that new price, demand (hence supply) will be less than before (1a)
therefore the price rise will be less
than the (price) shift of the supply schedule).

In summary,
if the supply schedule shifts downward, less will be supplied, and because of shortage, price begins to rise, causing demand to begin to fall.
With continued rise in price, consequent rise in supply and fall in demand,
a new equilibrium is reached.
At the new equilibrium, the price is higher
and both supply and demand are smaller than before.
As supply is less than before,
the change in price must not be as great
as the downward price shift in the supply schedule.

The expert’s argument to explain why the point of intersection of supply and demand schedule is the equilibrium point can be carried out verbally in a similar fashion.

Suppose that the price is above the intersection of the demand with the supply curve.
then the supply will be greater (2a)
and the demand will be smaller (1a)
so that supply will exceed demand (algebra),
giving a surplus, and a consequent decrease in price (4a).

But by the same argument,
the price cannot be lower than the intersection price (4b).

In the actual expert protocol, the first steps in the argument above are carried out diagrammatically, the final argument from the surplus, verbally, essentially using 4a and 4b.