Indexes Are "Out," Models Are "In"

Glorianna Davenport
MIT Media Lab

Is there any sign of intelligent life beyond keyword searching? My word processor's helpful spell-checker constantly urges me to substitute "composting" for the word "compositing." My Web search for the proverbial "Genie in a bottle" instead yields "Gen-X in a battle" with a 12-percent likelihood of correctness. Should I consider this brilliant, witty, and insightful commentary from my user-friendly laptop computer? Or is it just plain stupid? Oh, brave new world, that has such software in it!

What would it take to build a search engine possessing the knowledge, intelligence, and resourcefulness of my favorite research librarian? Her ability to interpret my inquiries, knowledgeably expand them, and then extrapolate them into a rich model for search and retrieval makes her an invaluable and pleasurable resource. She takes pride in knowing her library thoroughly, both spatially (where to physically find a book) and temporally (how her book inventory has changed and evolved). As she formulates her plan of attack, her sophisticated understanding of language, culture, experience, and other knowledge-domain models simultaneously converge and are re-mapped onto the reality of her library. The task of information retrieval becomes an interactive adventure of human dimensions, full of the satisfaction of continuous discovery.

Keyword matching is a crude and unsatisfying method for sampling the information content of complex sources such as the World Wide Web. Cooks seldom go to the Farmer's Market simply to find something "green," for example. These truncated representations fail to model the larger meaning embedded within the source (or the inquiry)—they leave behind all vital contextual information, and they strip away any basis for appraising the quality and veracity of the source. The worthless appears as the equal of the worthy.

In the electronic future, when all books are digitized and available on line, I pray that we have a more skilled and interesting guide than Yahoo or WebCrawler staffing the circulation desk.

Behind every theory lurks a model

Last fall, in MIT's "Report of the President,"[1] Charles Vest encouraged us to ask the "unanswerable" questions. As an example, he focused on the fascinating and maddeningly complex problems of weather prediction. We are currently devoting vast resources—satellite networks, sensor arrays, multiple radars, supercomputers—to an imperfect but practical system that only sometimes works. Will we ever develop a top-level model that allows us to predict climatic change precisely and accurately? Are we collecting the right type of data in sufficient quantity to discover its underlying mechanisms? Are there other ways of thinking or different observational perspectives that would allow us to build and operate better weather models?

Since the earliest of times, philosophers and then scientists have tackled the "unanswerable" questions about our world by formulating models. Why do we experience day and night, winter and summer? Why do the stars move in the sky? In ancient Eastern philosophy, it was obvious that these things happened because the world is a huge, flat plate carried on the back of a giant turtle. Of course, a flat plate has its limits, and proponents of this model predicted that if you went too close to the edge of the world, you would fall off. Needless to say, the turtle-and-plate model has undergone considerable refinement over the centuries. Slowly, under observational and theoretical pressures, it shifted to the Copernican model: the earth is obviously a round ball flying through the void, and we orbit the sun because we are attracted to it by the Universal Gravitational Constant—a more scientifically measurable sort of giant turtle.

The Copernican model is a good match with empirical observation and has the added benefits of being usefully predictive, accurately quantifiable, and precisely computational. However, much of the comforting and beneficial information provided by our knowledge about the giant turtle—including its name, its origin and history,

1070-986X/96/$5.00 © 1996 IEEE

and its true spiritual motivation for doing what it does—becomes unavailable to us through the Copernican model.

A good model, like a good story, is a redescription of the world (or at least a selected portion of the world, isolated and mounted on brackets). It is a scaled down, codified representation of objects, processes, and their interrelationships—a universe in a teacup. Whether a model is simple or complex, representation is the key to its power and usefulness. Superficial descriptions are not enough—we must have some meaningful glimpse of internal mechanisms and connectivity as well. When *x* happens, what does *y* do? The true power of models lies in our ability to play "what if?" games with them, to speculate by twiddling with the inputs and seeing the consequences.

The world is full of models, in all shapes and sizes: flight simulators, toy trains, spreadsheet programs, DNA, weather maps, the law. Of course, some models are better than others. What models ignore is often as important to understand as what they keep. Some models of real-world phenomena are the result of our ability to observe objects and processes and emulate them in a system—a "bottom-up" approach. Others choose to capture a theory, run real data through it, and compare its predictions to actual events—a "top-down" approach. Then there is the "black box" approach: Most engineers know that a handful of statistical data and a clever bit of logarithmic curve-fitting will allow them to fake the transfer functions between input and output of just about any physical system they encounter. The results are convenient but particularly dangerous—any real understanding of the underlying phenomena is removed from the model itself. Hopefully, in the future we will have sufficient knowledge to replace the "black boxes" with genuine mechanisms.

One likely approach to answering the unanswerable will undoubtedly involve building a vast library of proven micro-models and linking them together in various ways to form powerful, wide-ranging "universal models." Perhaps, as is the case with many complex systems, the effort may prove more revealing and productive than the results.

Brains in a box

For many years Marvin Minsky, one of the fathers of artificial intelligence, has contemplated the human brain. How do people think? How do we learn? How do we know? How do we maintain a continuity of ideas?

To derive a general model, Minsky thought about thinking. How could a stimulus order up the right amalgam of responses in the brain? Seymour Papert's corollary—"You can't think about thinking without thinking about something"—soon proved its mettle.

We appear hell-bent to drag simple models into the realm of storytelling and make them the foundation of automated storytelling systems.

In his 1986 book *Society of Mind,*[2] Minsky collected hundreds of examples that revealed small modules of what would grow into a general theory. The book itself is presented as a collection of self-contained, one-page modules; larger threads of meaning run across pages, across chapters, and throughout the book. Some disappear for a while, only to reappear later. The book makes sense even if you don't read the pages in order—by walking through a collection of pieces, a view of the whole begins to emerge.

One problem with most machine-based reasoning, Minsky argued, is that the machine lacks "common sense." When we think about storytelling, this seems apparent. *Snow Crash*[3] and other cyber-imaginings notwithstanding, the process by which we interpret story, let alone generate it, is complex and possibly perverse. Ask any author why they love creating stories; few will admit that part of the delight of their activity lies in creating secrets and building mystery into their world. James Joyce once argued that he wrote Ulysses to keep university professors on their toes; he also commented that Finnegan's Wake should take us as long to read as it took him to write (about seventeen years). Ferreting out the secrets of an author gives immense pleasure to the audience and allows for multi-layered readings of the story that can be shared with others, or not. This task of interpretation requires more "common sense" than we can currently define and catalog, let alone program into machines. There are also looming issues of "uncommon sense" reasoning required to understand complex works.

Nonetheless, we appear hell-bent to drag sim-

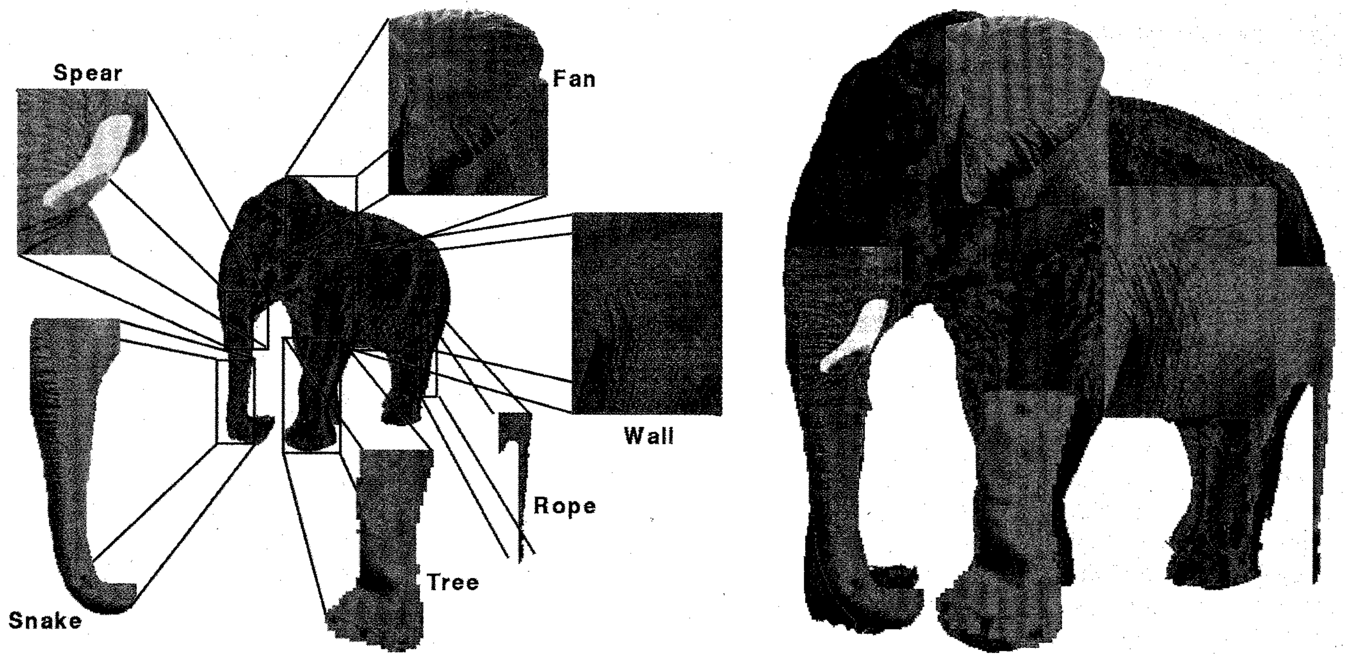


Figure 1. Ramesh Jain's illustration of "An elephant and six blind men." Video currently offers only a local perspective of what the camera was seeing at a point in time. What we want is a global perspective, an environment where the system understands the full semantics of a scene or world view.

ple models into the realm of storytelling and make them the foundation of automated storytelling systems. Virtual worlds with "chat rooms" peopled by avatars; teleconferencing to faces texture-mapped onto coarse, roughly modeled heads; object-oriented transmissions of various kinds; 3D sound environments—all rely on simple, limited-aspect models of the story space. A few systems also require some modeling of user preferences and expectations, but the extraction of this information is either cumbersome and intrusive or is gleaned only after long hours of use. None of these experimental models pays meaningful attention to the cognitive and sensory satisfactions of storytelling.

Doug Lenat's "Cyc" project was the first attempt to build "a single intelligent agent whose knowledge base contains tens of millions of facts, beliefs, categories, relations, problem solving methods, etc., that define our 'consensus reality'."[4] This lofty effort has proven to be a Herculean task, resulting in an enormous collection of data and code.

So far, experience suggests that general-purpose common-sense engines must by necessity be huge and cumbersome things. But the activities of thinking, creating, expressing, and communicating all draw on our ability to build abstract representations out of specific perceptions. A major obstacle to effective common-sense reasoning is often language itself. It is possible, however, that through prolonged use and refinement, common-sense engines will eventually boil down to a small strand of "DNA," a compact generative grammar linked to stories of experience.

In the future, we may flip common-sense reasoning on its ear precisely because we discover narrative intelligence and story is the true driver for technological change, rather than a poor cousin who makes unreasonable demands on the representational infrastructure. The constraints of storytelling (inherent in its sequential, spatial, and temporal nature) will bring order to the combinatorial chaos that current common-sense reasoning embodies.

Little pieces in search of a big model

For many years, V. Michael Bove has worked to extract 3D objects from videotape. He soon concluded that to "fix" the broken medium of television, we need to build the image up from metadata. Bove recently organized the "Objects of Communication" symposium held at the Media Laboratory in May 1996.[5] The symposium brought together an impressive group of practitioners: Chris Reader, Ramesh Jain, Reinhard Koch, Tamas Sandar, Andrew Scott, Dave Marvit, Frank Dutro, Steven Ackroyd, Julie Dorsey, Barry Vercoe, Joe Chung, and Bove himself. All use 3D spatial models as the basis for their work. Their specialties covered a wide range, including architectural design, scene design, entertainment titles, audio design, and medicine.

While this convocation highlighted the use of descriptive models to construct spaces, it essentially ignored the power of story, character, or action models. Only Chung paid passing homage to adaptive models. However, the sweep of examples provided a front-row seat to the pantheon of cultural and professional practice in an object-oriented age.

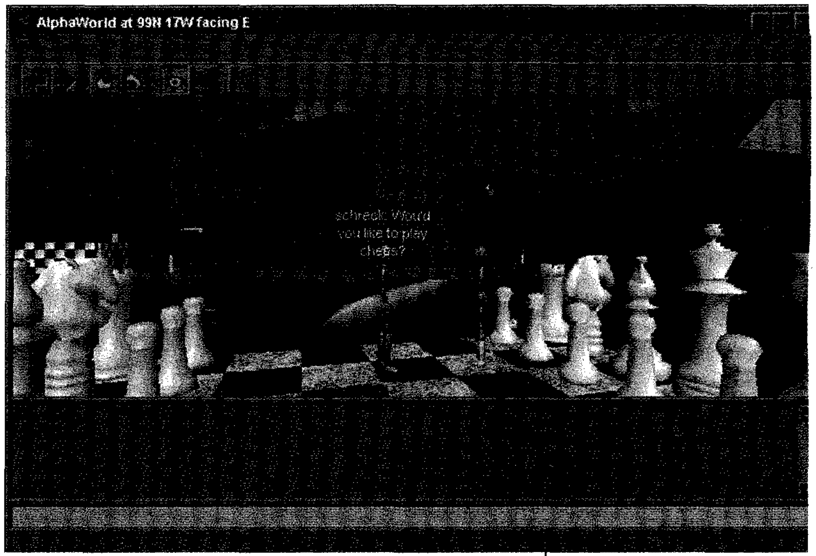
In overview, the symposium's work fell into three categories: digital 3D models created as approximations of real physical space (while these models are primarily descriptive, their use is often predictive in nature); 3D models of virtual spaces made for the containment of objects; and Hypertext models containing a functional adaptive dynamic.

The speakers who promoted the virtual modeling of real spaces gave us the clearest insight into the recent transformations of real-world practice. Frank Dutro, currently of Silicon Graphics Studios in Los Angeles, described the use of 3D modeling as a previsualization tool that can shave time and dollars off movie production. Dutro used set models from the films *Addams Family Values* and *The Firm* as examples to extol the benefits of simple models; when used in the field, speed is usually more important than precise accuracy.

The beauty and fully-realized detail of Julie Dorsey's work for the Metropolitan Opera Company stood in stark contrast to the quick, functional simplicity of Dutro's models. In the world of stage performance, lighting design is an essential and time-consuming task. Lights and their operators are an expensive commodity; failure to convey the desired mood and message of a scene is even more expensive. In the future, Dorsey hopes to adjust her simulations to more fully encompass human perceptual phenomena.

According to Ramesh Jain, a truly interactive video system must be able to reason semantically about its information content. In a University of California at San Diego demo, 20 cameras were used simultaneously to capture one sporting event. A primitive form of semantic reasoning enables the user to select who or what they are most interested in following, and the system dynamically selects the most illustrative camera view (see Figure 1).

Virtual worlds for socializing and "chat" are popular applications that currently sit at the low end of the model-based spectrum. In these applications, the objects in any given view may be driven by multiple players. Dave Marvit showed us how "World's Chat" and "Alpha World" (shown in Figure 2) allow participants to navigate synthetic 3D rooms under the guise of an avatar of their choice. However, it was in Reinhard Koch's work on "Realistic 3D Modeling of Persons and their Environment" that my jaded taste was finally inspired by the future of fake people. Would you rather be embodied as a goddess come down to earth or as a stiff, mannequin-like replicant topped off with a photographic image of your own head?



schreck: Would you like to play chess?

Figure 2. A screen shot from Alpha World. One visitor invites another to play chess.

Metadata compression and flexible, on-the-fly compositing of images was another hot topic addressed at the symposium. Chris Reader made a presentation on MPEG-4 that rated high on the technology scorecard. This likely next round of digital imaging is based on the idea that every identifiable item in a still or motion-picture image will, in the future, be composited on-the-fly into a model environment as called for by a script or a GUI. While it is unclear whether MPEG-4 is ready to become another standard—we still have insufficient knowledge to build a comprehensive and durable standard—there is no doubt that the underlying concept is a highly desirable future development. This provides a workaround of the limited-bandwidth issue. More importantly, it will allow for limitless flexibility and personalization in the delivery of video imagery: if you don't like Walter Brennan in the lead role, you can always drop in Humphrey Bogart instead; if you don't like Art Deco furniture, you can always substitute Duncan Phyfe.

Bove addressed his own research and the Media Lab's experience in the area of structured video, focusing on a short experimental piece called "Two Viewpoints." This interactive

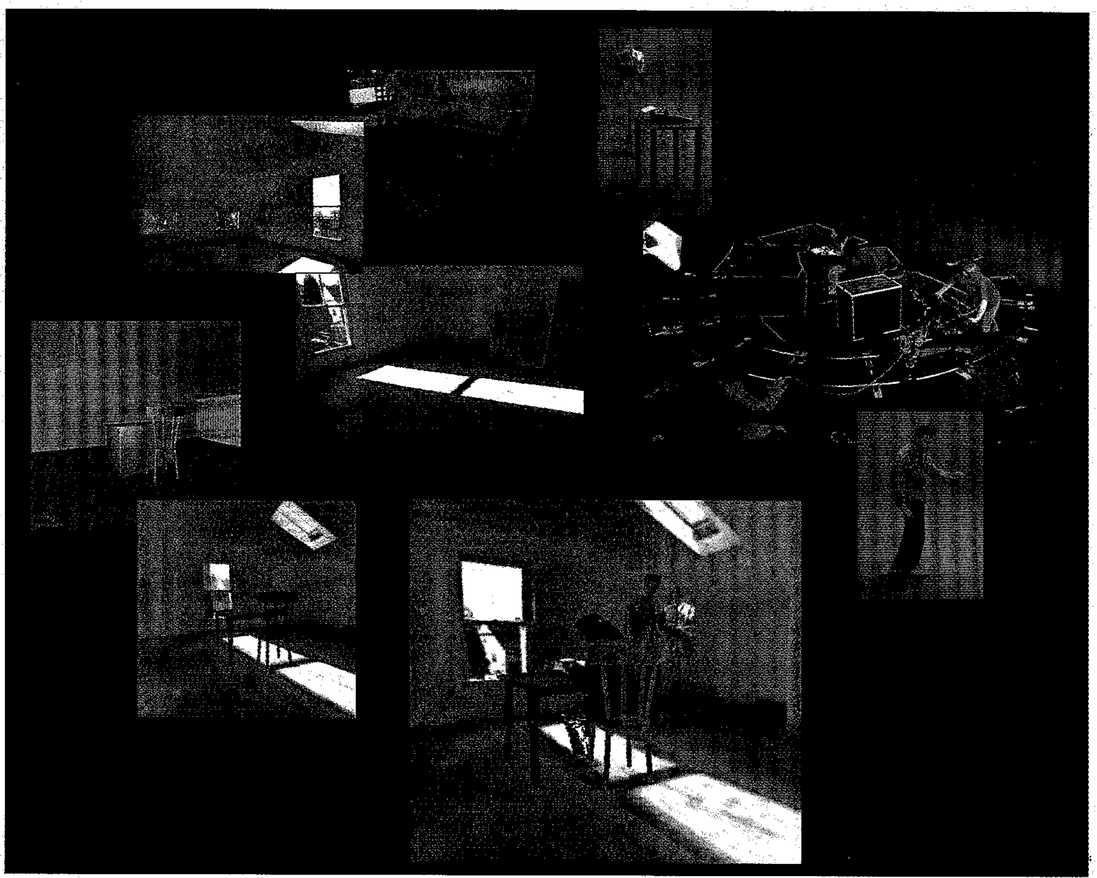


Figure 3. Shawn Becker's program takes still images as input, untangles perspectives and oclusions, and models the place in three dimensions. This could significantly alter the purpose and significantly increase the value of location discovery. http://www.media.mit.edu/~sbeck

sequence was shot on video and assembled using special hardware and a new, prototype scripting language called Isis.[6] It demonstrated that story production for object-oriented display is possible today, but only just.

If anyone feels discomfort about how we will construct entertainment using such a layered system, their fears will be allayed (or reinforced) by "The Wallpaper," a.k.a. "Two Viewpoints." This short interactive piece is the Media Lab's homage to "Yellow Wallpaper," a short story by 19th-century author Charlotte Perkins Gilman that has become a classic of feminist literature. Having advised and observed the production of this project, I promote this new world of object-oriented video production with trepidation. Nonetheless, the 3D setting, an attic room, is a technical and aesthetic tour de force. Shawn Becker used a series of still photographs to construct a complex, natural-looking space with realistic lighting, illustrated in Figure 3. As in other model-based video work, multiple (five) cameras captured the action performed on a blue-screen stage. The actors were then extracted from the blue-screen frame and saved as 2D objects. The system makes use of the wealth of views from the cameras, compositing, switching between cameras, and synthesizing new views on-the-fly (under direct user control) to obtain the "proper" perspective of the character.

This level of abstraction gives rise to new creative problems. The system of twirling around in 3D space, as if you are a camera on a mission, does not intrinsically generate a coherent and compelling story view. Cinematography has never been about simple pointing and clicking. Repositioning the camera is a reasoned activity, with or without Frank Dutro's models. The five stationary cameras flattened the action. I quickly realized we had effectively postponed any creative decision-making about point-of-view to postproduction, the assembly or display-scripting stage. One of my graduate students, David Tamés, put considerable effort into developing a tool for annotating subjective and objective camera positions to contour the story.[7] This work showed us how difficult it is to define cinematographic strategy computationally.

The "Two Viewpoints" project clearly revealed that, although we had succeeded in creating a model-based channel for presentation, we had failed to put any of that model on the front end of the system. This inspired me to take a second look at "Dramatica."[8] Both a theory and a software product, Dramatica was developed by Melanie Anne Phillips and Chris Huntley and distributed by Screenplay Systems. It is a good first attempt to build a general story model into software. Dramatica presents a model of the story mind that evolved from their theory that stories which fit our traditional sense of narrative (read "Hollywood") are modeled on the human mind as it deals with an "inequity." This theory eventually led the authors to create a sort of Rubik's cube, a cosmology detailing the range of character, action, and point-of-view elements an author must work with in order to create the progression of character growth and change essential to a satisfactory story.

Dramatica provides a rudimentary model for writing character-based, plot-driven narratives. According to its theory, the author must fashion an objective story with an antagonist and a protagonist, and a subjective story with a main character and an obstacle character or guide. The software provides a 3D mapping structure to guide the author in "story forming," "plot encoding," and "story weaving." Unfortunately, the current version is unable to model either interactive or cinematic strategies. Once you lock a character into a role type, the model is static and lacks the range of dynamic reversals that might be necessary in a world of model-based video, where we can dynamically modify the story stream, calling up transitions in setting and character action on-the-fly. On the other hand, structured video without a story-modeling engine will always seem hollow, lackluster, and incomplete. Changing point-of-view is not a simple matter of switching cameras. It requires a transformation of the story mind.

I considered taking the script and filmed images of "The Yellow Wallpaper" and evaluating it through Dramatica, but I quickly realized that this was insufficient due to the limitations of a single scene "demo." While using a "Dramatica-like" model, we could conceivably rework the scene to more effectively hold a shifting viewpoint. We need a parallel method for defining cinematographic strategy, such as, "We're going to shoot the objective story space from this angle, subjective from this," and so forth. The addresses, values, and arguments of a scripting language are too obtuse for the task of authoring compelling drama. If we are ever to build effective theories and models of the cinematographer's task, we need to understand the abstractions that the cinematographer talks about, and we need to build an experience base that captures her decisions as they are made in practice.

The universe in a tea cup

In *The Quark and the Jaguar,* Murray Gell-Mann reminds us that "no gluing together of partial studies of a complex nonlinear system can give a good idea of the behavior of the whole."[9] For the last two centuries, the arts and humanities have been segmented from the world of industry, largely due to the lack of a good model bridging them. A model provides a reusable formula—but formula is sometimes anathema to creative human venture. So far, we have loosely associated theories of thinking and models for visual production and presentation. Any generative story meme must also include a model of the audience. Recent research has emphasized the need for adaptive user models, where the adaptive forces are derived from both the user's own activity and from the activities of others. The care and feeding of users, which is one of the most important nodes in any communications model, will be more thoroughly explored in a future Visions and Views column. **MM**

References

1. C. Vest, "Massachusetts Institute of Technology: Report of the President for the Academic Year 1994-1995," *Technology Review,* Massachusetts Institute of Technology, Cambridge, Mass., Jan. 1996, pp. 1-8.
2. M. Minsky, *The Society of Mind,* Simon and Schuster, New York, 1988.
3. N. Stephenson, *Snow Crash,* Bantam Books, New York, 1992.
4. D. Lenat and R.V. Guha, "The World According to Cyc According to D. Lenat and R.V. Guha," *MCC Tech. Report No. ACA-AI-300-88,* Austin, Texas, 1988.
5. N. Negroponte, "Object-Oriented Television," *Wired,* July 1996, p. 188.
6. S. Agamanolis, "High-Level Scripting Environments for Interactive Multimedia Systems," MS thesis, Mass. Inst. of Technology, Cambridge, Mass., 1996.
7. D. Tames, "Some Assembly Required," MS thesis, Mass. Inst. of Technology, Cambridge, Mass., 1995.
8. M.A. Phillips and C. Huntley, *Dramatica: A New Theory of Story,* Screenplay Systems, Burbank Calif., 1996.
9. M. Gell-Mann, *The Quark and the Jaquar: Adventures in the Simple and the Complex,* W.H. Freeman, New York, 1996.

Contact department editor Davenport by e-mail at gid@media.mit.edu.