This appeared in edited form as "Digital Humans Wait in the Wings" in the Nov. 2000 *Scientific American*, 72–76. The unedited form here is the one I prefer.

The Cameraless Movie

Alvy Ray Smith
7 Jun 2000

Recent film *Being John Malkovich* explores the fascinating implications of an actor's body—that of John Malkovich—occupied by a second person's mind. A female occupier, for example, manages to have sex with another woman heterosexually. Taking the idea one step farther, a male occupier, John Cusack, playing an accomplished puppeteer, manipulates yet a third person in the form of a lifesize marionette in a ballet at Lincoln Center—through the hands of the occupied Malkovich. This film, although not cameraless, highlights the heart of the Cameraless Movie concept—separation of artist and representation.

The Completely Digital Movie has all visuals generated by computer—actors, sets, locations, visual effects—and all audio too—dialog, music, location sound, sound effects. All editing and mixing of the components proceeds digitally. Digital distribution and digital projection complete the idea. The Cameraless Movie comprises a subset of this large idea, as does its audio equivalent, the Recorderless Movie. For the Cameraless Movie, let us assume the partially digital audio generating and recording technology of today—dialog, in particular, being human generated. In other words, the Cameraless Movie is strictly a visual concept. The equally interesting Recorderless Movie deserves separate treatment. This issue of *Scientific American* treats digital distribution and display elsewhere [pointer here] so I shall concentrate on completely digital production of visual content.

Notice that the Cameraless Movie as defined above already exists. *Toy Story* first claimed that distinction in 1994, and three others have followed: *Antz*, *A Bug's Life*, and *Toy Story 2*. True, all these used cameras for final recording to film, a minor matter because the digital source material for each can be inserted directly into a digital distribution and projection system. These four movies are 3D animations, created by Pixar for Disney except *Antz* created by Pacific Data Images for Dreamworks. [Figure 1 shows a typical modeling and rendering sequence of steps for 3D film generation.] More importantly, all are cartoons. The Cameraless Movie vision encompasses all films, particularly those with human actors, not cartoons, so I focus on the challenge of humans.

Some might argue that *The Rescuers Down Under* (1990) deserves the title of first Cameraless Movie, and that its successors, *Beauty and the Beast*, *Aladdin*, and *The Lion King*, predate *Toy Story*. All of these 2D "cel" animations (because artists painted the frames on celluloid in the original technology) relied completely on the computer except for the original drawings—all done by hand—and final film. I exclude them from consideration, although I had a hand in originating their technology, because I believe the technology will remain in the realm of cartoons.

The larger idea, the Completely Digital Movie, combines the Cameraless Movie and the Recorderless Movie with digital creation of the art—the story, the dialog, the music, the acting, and so forth—which I have carefully excluded from the Cameraless Movie discussion because it leads into the controversial territory dominated by the three great modern religions of Artificial Intelligence, Virtual Reality, and Artifical Life—or in short, to the digital simulation or "realization" of everything. In particular, digital simulation of actors forces

me to venture here at all. As I will show, in order to avoid this murky, faith-driven area, we must distinguish between the art of acting and the representation of actors. The Cameraless Movie focuses on the latter.

When movie star Richard Dreyfuss presented a technical Academy Award to colleagues and me a few years ago, he commented, looking our way, ". . . we're both indispensable to each other. Don't forget that—the people who made *Toy Story*. We're all going into the 21st century *together*—I hope!" to much laughter, some of it nervous. His barely disguised fear fed on the oft-expressed claim, from members of the computer science profession, that simulated actors will someday take the place of real ones. Zealots would have Dreyfuss replaced with a simulacrum that walks, talks, gestures, thinks, emotes, and otherwise acts just like the man himself, without hiring him, presumably, with his hours at his prices. I and my Pixar colleagues are constantly asked, "Will actors be replaced?" I attempt an answer below for the visual aspects, but leave it to the reader to contemplate the analogous question for the Recorderless Movie, "Will actors' voices be replaced?"

The completely simulated Richard Dreyfuss requires solution to imposing problems: consciousness, emotion (if indeed the two are separable [pointer to Antonio Damasio's *Scientific American* article]), physical nuance, and acting, just for starters.[1] I am unwilling to deny that we might someday make an artificial Dreyfuss—or Marilyn Monroe, the usually favored example. My personal belief says that he and she are machines, as I am, and therefore ultimately explainable, at least to a much finer degree than we have today, but that's the highly inspirational faith many of us scientists hold, not known fact. I cannot predict (nor can anyone) when or how this might happen, nor can I (or anyone) determine whether the virtual actor would ever be cheaper or more versatile than the real item. And we might yet run into some fundamental logical roadblock as did the mathematicians, with Godel's Incompleteness Theorem, when they set out to derive all of mathematics. So rather than wildly fantasizing across eons, invoking a magical Second Coming called "emergence," or preaching world domination by machines, as have some of my professional colleagues, I offer instead what we might reasonably attain in a current lifetime. In fact, I claim that we have already done so to a small degree. This solution, without hocus-pocus, integrates the best of both worlds, human and machine, artistic and technical, with mutual benefit.

A man-machine melding that can really work, achievable without resorting to faith and without putting artists out of work, is this: *Real actors driving realistic representations of human beings*, perhaps even themselves. I am so sure of this because we have already done it. It's called animation. The movies listed above represent some of the first steps. When John Lasseter, Pixar's principal artistic director, searches for new animators, he looks neither for animation or drawing ability, nor for knowledge of computer modeling programs, and certainly not for programming experience; he asks for acting skill. By this conception, an animator is just that special kind of actor who can make us believe that a pencil drawing or a collection of colored geometric polygons has heart, gets angry, and outfoxes the coyote. Alternatively, an actor is an animator of his own body; he or she makes us believe that the body we see on the screen or stage is that of an entirely different human being. So we already have actors, called animators, driving not very realistic representations of human be-

[1] Not to mention the philosophical question of qualia, or how we know "blueness," for example. [All footnotes are provided for editorial reference and verification, or for my own information. They are to be omitted from the final article.]

ings—or animals or objects. All that's missing then from the full Cameraless Movie vision is realism, the appearance of reality, and that will come, as we see below after reviewing another widespread example of the separation of humans from their representations.

Anyone who has played DOOM, Quake, Ultima Online, or many other such games has represented himself or herself by a graphical character or object, sometimes called an avatar. Screen fantasies in films such as *The Matrix* take the concept to unbelievable heights. Provocative, if not overblown, cyberworld developments are the electronic Halloween venues where people freely represent themselves, with little or no reality check. More than a negligible number choose an avatar of the opposite sex. Some choose animals or objects. But animators have done this for years. The point is that humans already commonly drive self-representations. As with movies, the quality of the representations still suffers, especially on the bandwidth-starved web of today. For the Cameraless Movie, think of an actor driving a first-rate avatar—a realistically rendered one, whether of himself, another human, or an animal or object. Think of (real) John Cusack driving a realistic representation of John Malkovich's body. So what is the prognosis for reality?

A famous "law," Moore's Law, captures the dynamic of the ongoing digital revolution. Let me express it in an uncommon but potent way: *Everything good about computers gets ten times better every five years*, or *"10× in 5."* The Law is usually, and arcanely, expressed as the density of transistors on an integrated circuit chip doubling every year and a half. Mathematically this is equivalent to a factor of ten every five years.[2] And "everything good" about computers directly tracks the density of transistors: increased memory, faster speed, lower price. A factor of ten is more meaningful than a factor of two because it is an increase by "an order of magnitude," an expression we use purposely to connote a conceptual leap, not just a larger number (or smaller in the case of price). Currently, chip designers believe Moore's Law still has about ten more years' usefulness—another factor of 100×—before a physical wall bars further decrease in transistor size. Quantum mechanical weirdness will undoubtedly reign when the widths reduce to about ten atoms. Some currently primitive technology such as quantum or organic computing *might* then leap forward to fuel a continued revolution [pointer to *Computing with Molecules* article in Jun 00 issue].

Moore's Law applied to pictures seems to tell us that 3–17 million polygons[3] per frame in 1994—*Toy Story* complexity—should have become 30–170 million polygons per frame by the time *Toy Story 2* reached completion in 1999.[4] Interestingly, my longtime colleagues Loren Carpenter, Ed Catmull, and Rob Cook first came up with the measure of 80 million polygons as the reality threshold when we were at Lucasfilm, and I have quoted this estimate for many years, saying "Reality begins at 80 million polygons." However, *Toy Story 2* only doubled *Toy Story* complexity to 4–39 million polygons. [Figure 2 shows the actual visual comparison.] What *did* increase by a factor of 10 in that five years was total rendering time per frame. Pixar's Don Schreiter rerendered frames from both movies for this article, but on the same modern hardware for a direct comparison. Frames from the newer film

[2] The exact factor is $2^{5/1.5}$, or about 10.079×; for 10 years, the factor is 101.59×. N.B., from experience, I can predict that there will be readers who mistakenly compute this as 2*(5/1.5) = 6.67×, or 13.33× for 10 years.

[3] I have stats for two frames: 1,789,791 and 17,437,119 polygons. The former one does not include the background, for which I add another 1 million polygons. New stats: 7,260489, 10,963,795, and 14,004,859.

[4] I have stats for two frames here too: 2,992,732 and 39,438,264 polygons. I add 2 million to the former number, for the same reason but doubled to take the general increased complexity into approximate account. New stats: 18,782,826 and 32,347,209.

took 6–13 times as long to compute—roughly an order of magnitude. However, averages provided by Pixar's Bill Reeves, taken over all frames of both movies, indicate a ratio toward the lower end, at about 5, in general.

Caution! Marketers frequently misinterpret the 80 million polygon figure: Hardware manufacturers measure the speed of their graphic devices in terms of polygons generated *per second*, but for visual meaning, complexity per picture is what matters, regardless of the speed of its generation. Thus 80 million polygons *per frame* is the claimed threshold of reality. In hardware rate terms, using film's 24 frames per second, the reality threshold is 1.9 *billion* polygons per second.[5,6] Reality in realtime would be marvelous, of course, but the movie business would happily just reach reality. To measure how distant realtime still is, consider this: *Toy Story* took an average of seven hours computation for each frame. *Toy Story 2* required several hours each too, some of its most complex frames taking over 50 hours.[7]

Another caution: We don't really claim to have replicated reality at 80 million polygons. True reality might be fractal in complexity: The closer you look the more it looks the same and just as complex. We mean that the visual complexity achieved at that level sufficiently interests most of us to be accepted for reality, especially when not the focus of attention. In fact, *all* media approximate reality. For example, current television (eg, a newscast) samples reality vertically with only 486 lines—and we contentedly believe that we thus see reality.

And a third caution: To think that true realism is even our goal is a mistake. Movies are never real. Dialog proceeds pauselessly, sets feature false fronts, lighting is artificial, makeup heightens facial characteristics, and editing and pacing play with time. What "realism" means here is a convincing, believable representation of reality. Usage of the representations might be totally surreal. It is in this context that I have often said, "Reality is just a convenient measure of complexity." It establishes a pleasing level of complexity, not necessarily a target or a limit. To many, however, the search for realism is the Holy Grail, so I address these remarks to them.

We calculate the complexity measure for a canonical one million pixels then prorate easily to a particular resolution. For reference, *Toy Story* and *Toy Story 2* have frames of 1.4 million pixels, *A Bug's Life* 1.8 million pixels, and *Monsters, Inc* (Pixar's next) will have 2 million pixels:[8]

As you look out into the real world, imagine it projected onto a rectangular, but continuous, screen. We seek to synthesize this viewport into reality digitally. Picture the screen being divided into a rectangular array of a million little squares, one per final pixel. Con-

---

[5] The video rate of 30 frames per second implies 2.4 billion polygons per second.

[6] The new Sony PlayStation 2 supposedly generates 25–50 million polygons per second. That's 1–2 million polygons per frame at best. Not shabby for realtime—in fact, marvelous—but far short of even the threshold of reality: That's 50 million full-color polygons, 25 million textured polygons per second. (The company uses 75 million polygons per second to illustrate raw speed but these are unusable polygons for real scenes.) The lower number is the one that uses polygons most like the polygons Pixar uses to render scenes (ignoring motion blur and other sophisticated techniques beyond the range of today's hardware to execute in realtime).

[7] Actual time during our experiments was 33 hours. During production, however, the machines were not optimized as they now are. They were swapping heavily, and took 50–60 hours.

[8] *Toy Story* and *Toy Story 2* have frames of 1536×922 = 1,416,192 pixels, *A Bug's Life* 2048×871 = 1,783,808 pixels, and *Monsters, Inc* will have 1920×1038 = 1,992,960 pixels.

sider a ray sent out from your eye into the world through this screen. The bundle of all the rays passing through one square averages into one color in the corresponding pixel. Thus a pixel represents, with a single color, the average of everything in the scene intersected by its bundle of rays. Entire galaxies might reside in a single bundle! Actually, the bundles may not have a rectangular cross-section and might overlap, but the simplified "little square" model captures the essence. [A Figure 3 could be used here. See my rough sketch below.]

By the argument, every bundle "sees" about four levels of surfaces before the intersected surfaces completely obscure the ones behind, including the galaxies (not visible to the naked eye anyway). Each level of surface comprises contributions from about eight polygons.[9] Thus 32 polygons will contribute on average to each bundle's pixel, hence 32 million polygons for each million pixels of frame resolution. At the time we considered frames with 2.5 million pixels,[10] hence 80 million polygons per frame.

The numbers from this argument have held up empirically—and remarkably, considering its arbitrariness: Why not eight levels of surfaces and 16 polygons per layer? Moreover, it disquiets that the complexity of reality, a fixed notion, should vary depending on the final resolution of a representation of that reality. We really want an argument that gives the complexity of reality over a rectangular pyramid of space—that occupied by the final screen's view—regardless of the resolution of the sampling of that screen. Instead, with encouragement from Pixar colleagues Bill Reeves, Ed Catmull, Oren Jacob, and Galyn Susman, I argue that polygon counting, or any geometric measure, misses the point. In other words, current geometric complexity nearly suffices, so what else does realism require? If geometry consumes only 10 percent of the problem, as Pixar reports, what constitutes the rest?

Shape and shade combine for believable representations. Shape, a 4D concept, covers the geometry but also the motion of objects, such as actors. Consider the nuance of facial expression. Movement requires the same faithfulness to reality as does static shape[11]. Shading colors the objects and includes the illumination of them as well as their material and textural makeup. Lighting has arisen as a particularly challenging task, direct computation of the physics being too hard—Mother Nature does it in parallel and realtime, but computers serially and laboriously. Consider the difficulty of getting the lights just right with turnaround times per frame on the order of hours. We can predict that the shading portion of this other 90% of the effort should ease remarkably when the Moore's Law 100× factor drops hours to seconds of rendering time.

Humans add the requirement of heightened accuracy of representations, including movement. We do not forgive errors in pictures of ourselves, having a word for an almost-but-not-quite-human: monster. The computer graphics community currently opts for simplifying humans—making them obvious cartoons—rather than risk our perceiving them as monstrous. Or it exploits the monstrosity, as in *The Mummy*. Or it opts for animal representation, as with the mouse in *Stuart Little*. My friend Barbara Robertson, longtime commentator on and reporter of the computer graphics scene and fellow supporter of the point

[9] Polygons are assumed to range in size from one-half to one-fourth the width of a little square—ie, from 4 to 16 polygons per layer per pixel. Carpenter et al used 8 as an average.

[10] 2048×1226 = 2,510,848 pixels

[11] If tree-lined street image is used from *Toy Story 2*, we can make the point that subtle movement of the leaves is more important to the reality of the scene than the geometric accuracy of the still frame.

of view expressed here, suggests that the eponymous mouse holds the title for best artificial actor in a live-action film as of this writing, factoring in complexity of representation as well as acting ability. Perhaps the Academy of Motion Pictures Art and Sciences should designate an award for this new category—for an avatar *and* its actor/animator.

The problem becomes designing tools to model intricately detailed representations, without the tools themselves becoming hopelessly complex. As suggested here the efficient solution will have humans driving the representations. Of course, physical modeling will also help as physics comes within the reach of ordinary computers by the turning of the Moore's Law crank. And techniques such as motion capture and image-based rendering will allow us to measure and use reality itself more accurately. Nevertheless, as the number of controls for a character increases from the hundreds (Woody in *Toy Story*) through the thousands (Al in *Toy Story 2*) to, say, hundreds of thousands, or even millions of controls for believable humans, something must be done to deliver them in intuitive form to their human drivers. We do this for control of our own bodies, after all. When I see a friend, production of a giant smile happens automatically, seeming to proceed from the single command "dear friend."

So Richard Dreyfuss should relax and contemplate a screen body that doesn't necessarily age. The Cameraless Movie needs his talent. He should start talking with his intellectual property attorney about protecting his representation. Does the real he or his avatar command the higher salary? The essential problem becomes that of interfacing the actor to the realistic model, and his adapting to that way of things. Animator/actors currently explore this area, but, as I have argued, one should not therefore conclude that cartoons constitute the limit of the idea. It took us 20 years from initial conception to reach the first completely computer-generated movie. Perhaps the next 20 will bring us the full Cameraless Movie—from artists and technicians creating together.

Reading and Viewing

Damasio, Antonio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Harcourt Brace & Company, NY, 1999.

Robertson, Barbara, Building a better mouse, *Computer Graphics World*, Vol 22, No 12, Dec 1999.

A Bug's Life, Disney, 1998; *Aladdin*, Disney, 1992; *Antz*, DreamWorks, 1998; *Being John Malkovich*, USA Films, 1999; *Beauty and the Beast*, Disney, 1991; *Stuart Little*, Sony, 1999; *The Lion King*, Disney, 1994; *The Matrix*, Warner, 1999; *The Rescuers Down Under*, Disney, 1990; *The Mummy*, Universal, 1999; *Toy Story*, Disney, 1994, *Toy Story 2*, Disney, 1999.

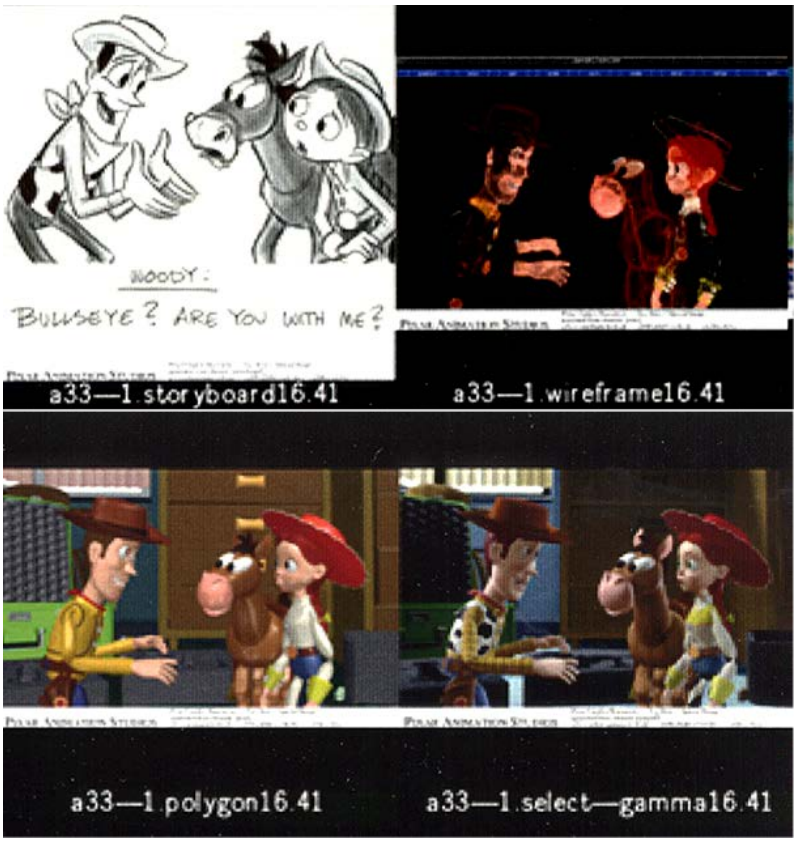


Figure 1. *Toy Story 2* development sequence. (top left) The original storyboard pencil sketch. (top right) Crude 3D (3-dimensional) computer graphics realization of characters for placement, mostly as "wireframes" or outlines. (bottom left) Simple rendering check with colored polygons—typically small triangles and quadrilaterals—with background. (bottom right) Final rendering at movie screen resolution with full lighting of textured and patterned material. [Labels underneath are Pixar's control numbers for editorial reference.]

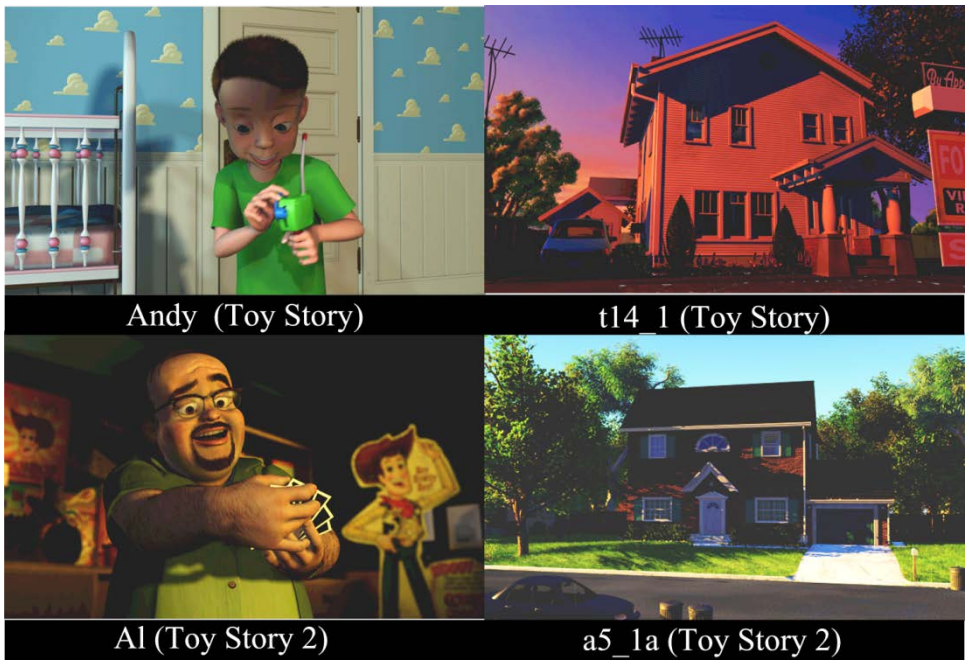
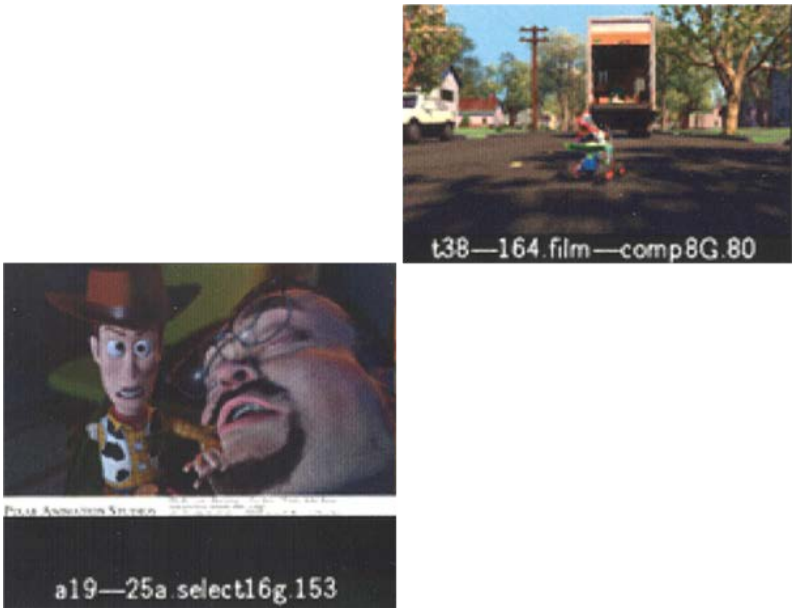


Figure 2. Comparison of *Toy Story* (top row) complexity with that of *Toy Story 2* (bottom row), five years later, with human characters on the left and complex outdoor scenes on the right. The geometric complexity increased by a factor of two between the two movies, but the computation time by a factor of ten, an order of magnitude. The additional computation is spent on better lighting, texturing, and surface representations. Not evident from these stills is the increased subtlety of motion as well—for example, the gentle fluttering of the leaves in the outdoor scenes of *Toy Story 2*, greatly increasing the sense of reality. [The labels in the right column are Pixar's control numbers. I don't have the corresponding numbers for the frames on the left. Pixar's Don Schreiter knows which frames these are. NB, these are the frames actually used for the measurements I reference in the article.]

[Altenative Figure 2 components. Should there be some problem with the frames I have chosen for Figure 2 on the preceding page, the frames shown above, in corresponding order and with their Pixar control labels, are second choices. Notice that the lower right picture is the same as before, however. The "control number" below it is the label that Don Schreiter used in our communications. I don't have an alternative for it.]



[Alternatives to the alternatives for Figure 2. These are my third choices, shown in corresponding positions. Pixar control numbers underneath.]

[Cover suggestion. I extracted this square image from the remarkable frame (see preceding page) a19_25a.select16g.153. Seems like it should be used for something! Assuming Pixar would agree to the cropping.]

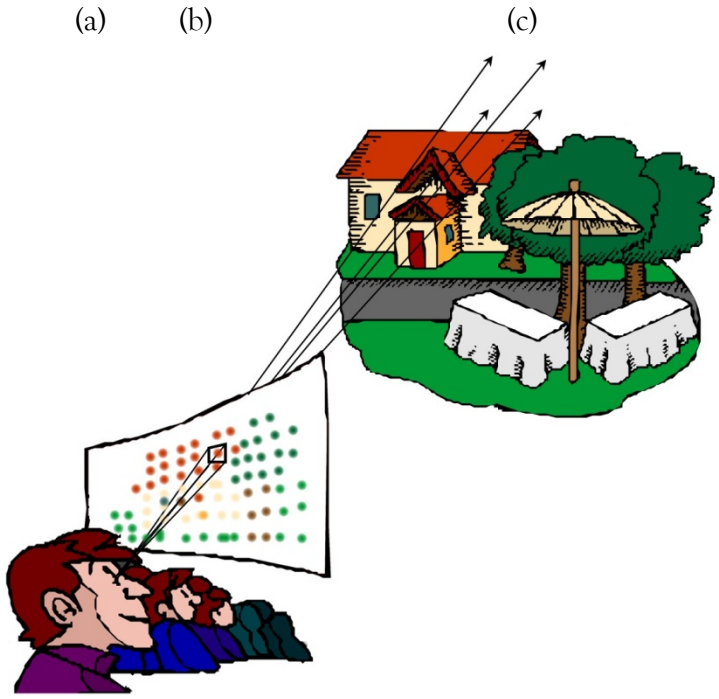


Figure 3 [rough draft]. A solid bundle of rays from a viewer (a) passes through one square on the screen (b) acting as a window into the reality (c) beyond, which the screen is to represent. One dot of color, a pixel, will be placed on the screen inside the little square to represent *all* of reality intersected by the bundle of rays passing through it. This is repeated for about two million dots on the screen. The projector spreads each dot slightly, and the eye melds them into a continuum. The computer graphics problem is: How is (c) modeled and how is the bundle average computed so that the relatively small number of dots on the screen believably represents the complexity at (c)? [Note to illustrator: Do not put colored squares on the screen. It is a common misconception that pixels are little squares of color. They are not. They are points of color. The little square represents the cross-section of the solid volume of reality averaged into the one point of color located in that square. An idea is to have a thought balloon on the viewer showing the continuous picture his brain forms from the dots. The bundle of rays is not just the four rays shown but a solid cone of rays intersecting the little square in all possible places. I have just drawn the four corner rays of the bundle above. Also, the intersection of the bundle with the "reality" beyond needs to be indicated somehow. In the rough draft above, the bundle presumably intersects just the roof, the average color of which is simply the roof color in this cartoon. To really show the problem, the bundle would intersect several surfaces of different colors.]