

## *Video Epistemology In- and Outside the Box: Traversing Attentional Spaces*

**Jay Lemke**

*University of Michigan*

### EXPERIENCING VIDEO MEDIA

In its earliest form, we encountered video images inside the magic box of early broadcast television. To some observers it might well have seemed that inside that box there was an animated dollhouse inhabited by miniaturized people, caught in some eldritch sorcery. On the tiny stage, classic theatrical dramas were being performed. The bulky boxes held a magical space, a space that in turn magically held our attention. A space that was rescaled relative to that of our living rooms, but that also seemed a spatial extension of them into another dimension. Some of these boxes had wooden doors, to be opened like any other cabinet or tiny closet. As children, we pressed our noses against the glass, wondering if we could somehow get inside this ante-room that opened up to elsewhere.

Early television programs played with the fascination of the boundary between living rooms and the living space of the video world. Children wrote on transparent screens (or naively on the glass itself) and the video characters would complete their pictures or give them the magic codes of their games; for example, *Winky Dink and You* (Prichett & Wyckoff, 1953–1957) or *Captain Video* (Brock, 1949–1955). This was the first interactive video. The recent film *Pleasantville* (Ross, 1998) revived the fantasy of television as a portal into a 1950s black and white world, as Woody Allen's 1985 *Purple Rose of Cairo* similarly imagined a 1930s movie screen as an interface between the worlds of a movie and of the theatre where it was playing. We never perceive cinematic and televisual worlds as two-dimensional surfaces, but always as three-dimen-

sional spaces that are worlds in their own right and, at the same, time spatial extensions of our own local places.

I want to consider here issues of epistemology and verisimilitude for video both as a medium we experience and as a medium through which we record data for research in the learning sciences. These are linked because we make use of video data by viewing the video record, and we cannot understand the epistemology of video as representation unless we also understand the processes by which we make meaning with video when we experience it. I propose that we consider the semiotic uses of video in terms of the ways we meaningfully (and feeling fully) move across and through immediate and mediated attentional spaces. I will develop the notion of multiple “attentional spaces” to stand for both the telemediated worlds to which we pay attention (real or fictional) as well as those more immediately navigable by our bodies.

I am starting from an experiential or phenomenological perspective because, regardless of the underlying technology, whether film or videotape or digital display, there is a quality of experience that remains constant. I will certainly also be considering this experience as a semiotic construction of meaning and as a social practice characteristic of a community and its culture. But I do not want to lose sight of materiality, or of the bodily and felt qualities of the experience. Feeling as well as meaning matters, and indeed, feeling matters to meaning, and vice versa. I am not limiting materiality here to the proprioceptive or visceral body. Contrary to myth, video is not a passive medium; we *do* act and move in using the technology and in response to its images and sounds. Embedded in interactive multimedia, as it increasingly is, video becomes an even more active medium. We make meaning with video and its extensions (video games, interactive video, interactive multimedia and hypermedia) in real space and in real time. Video genres have chronotopes (Bakhtin, 1981); typical patterns of movement among sites (*topoi*) and of pacing action in time (*chronos*), and there are chronotopes also for the kinds of activities in which we use video, including those for research.

I want to examine the diversity of video genres—commercial, ethnographic, and research records and the commonality and variety of ways in which we respond to them—as virtual realities, as representations, and as data. Before doing so, I need to sort out first some issues of the nature of the medium and then some of the ways in which we experience it as part of larger activities in which we traverse across different sites and other media. Afterwards, I want to sketch one possible future direction for the use and study of this medium in our work in the learning sciences.

## VIDEO AND KINDRED MEDIA

First, a few preliminaries. My primary object of concern here is the experience of video as a space of visible and audible dynamic activity: We see and hear something happening in a virtual space and time that has some indeterminate relationship to the immediate space and time that our bodies inhabit. In this sense, experienced video may not be significantly different for research purposes from experienced film or cinema, although there are certainly differences in “feel” arising from the conventionally greater resolution of film images (a temporary difference, already disappearing) and the conventionally larger spatial scale of the cinema screen. By “video medium,” I will

mean primarily the sensory-attentional features of the medium to which we respond in making meanings and in experiencing feelings, rather than the material technologies that record and present these features to our senses. We can call “video technology” the assemblage of operable artifacts, and the conventional ways of using them, that enable us to record and play back video data streams. There are also, of course, “video texts” that we experience as having some coherence or cohesion of meaning and feeling over their run time of play back, and “video genres” that are typical and recognizable cultural forms with reference to which we make sense of what is happening in most video texts. These “product” genres are the result of “process” or activity genres: The conventional ways in which video texts of various genres are made or produced (including scripted, unscripted, and/or directed action, camera-work, editing, etc.).

With the definitions given, both video and film experience share substantially the same audio-visual semiotic; the same interpretative conventions for their salient sensory features. They have more or less different technologies, but very similar or identical genres, and there are many texts that exist in both media. We all know that video technologies are used increasingly in cinema production, along with digital technologies, and there is an increasing convergence among these technologies and also among the media. Watching a film on a big screen in a theatre is still a somewhat different social and phenomenological experience from watching substantially the same text (perhaps with different aspect ratio, another temporary difference) at home on even a large television monitor, much less on a smaller flat screen digital video monitor. The extreme scale of the IMAX technology (a screen that fills peripheral vision, tens of meters high and wide, or in older similar technologies wrapping around a curved screen) does produce unique phenomenological effects of immersion and telepresence, similar to those of Virtual Reality (VR) technologies (screen-based caves or goggle-based optics). We get strong visceral and vestibular (proprioceptive) sensations of motion, and so a more vivid sense of a three-dimensional spatial environment.

In between video and VR is the increasingly popular medium of video games. Whether presented on televisions or computer monitor screens (or in arcades on dedicated systems), video games animate a dynamic, interactive, apparently three-dimensional world, but one in which the realism is generally reduced compared to most film and video texts by the more cartoonlike actors, scenery, and in-world artifacts. Video games however, do include motion-capture sequences made with live actors, cut-scene video, and other elements that overlap with those of the recorded video medium. In the case of co-produced films and video games such as *The Matrix Reloaded* and *Enter the Matrix* (Shiny Entertainment, 2003; Wachowski & Wachowski, 2003) or *The Lord of the Rings: The Return of the King* (Electronic Arts, 2003; Jackson, 2003) not only is there great overlap in features, meanings, feelings, and so forth, but the media themselves are mixed and hybridized. The film, video, and videogame texts are also marketed in relation to one another. The greatest difference is that the videogame medium is highly interactive in ways that the video medium itself is not: We more frequently engage in physical contact and motor movement interactions with the technological artifacts that provide our interface to the medium (keyboards and game controllers vs. television or VCR remotes and buttons), and we more regularly see and hear changes in the videogame features that we interpret as direct responses to our ac-

tions, just as we do in the world of immediate experience. I emphasize this new member of the family of videolike media because of its prospective importance for research in the learning sciences, to which I will return later.

The video medium of course has long included genres of reduced realism, beginning with animated cartoon films, and has also experimented with hybrids of visually realistic genres (the feature film) and cartoon animations, such as in *Who Framed Roger Rabbit* (Zemeckis, 1988). On this analytical dimension as well, there is a quasi-continuum of media and genres linking film/video and video games.

In the cases of video games and IMAX movies we are most vividly presented with the paradox of virtual spaces whose spatial and temporal relationships to our immediate surroundings are indeterminate. (For immersive VR media, there may be no visual or auditory input from our unmediated surroundings, although there is still a doubled tactile-kinesthetic world.) We may focus on the interface and try to imagine the screen as a portal between two worlds, or two places in the same world (as is the case with video-conferencing experiences), or we may yield to the illusion of virtual presence in the video world, whether as observer or as participant. We can imagine that we could, or already have, passed through the portal, or that the interface technology exists in some sense in both worlds, and that in video games, by touching the controls in this world, we are performing actions in the other world. Video-system (TV, VCR) remote controls can already be viewed as precursors of this interface metaphor, and in fact, are readily integrated as artifacts with videogame controller units.

#### ATTENTIONAL WORLDS AND TRAVERSALS

In the cases just described, we may either feel experientially that we are living in two worlds at once, or that we are effectively immersed “in” the virtual world. I want to speak of both the immediate and the mediated worlds in the same terms semiotically, affectively, and insofar as possible, bodily and materially (including spatiotemporally). It seems useful to construct a notion of “attentional worlds” to indicate that experientially, we attend to the happenings, sights, sounds, meanings, and feelings of a particular space or place, even if it is a virtual one (i.e., one with which we do not have all the same material interactional affordances as we do with what we take to be the world mediated by our unaided senses and physical contacts). We are capable of attending simultaneously to, or at least cycling rapidly among multiple attentional worlds. We have the experience of moving from one attentional world to another in many aspects of our more and more technologically mediated lives. We attend to the meeting we are in, to the space inhabited by the person we are talking to on our cell phone (or with whom we co-habit), to the relatively flat (i.e., feature-poor) spaces of text windows on our laptop, and perhaps also to the video presentation on a room screen or on our laptop screen.

Even without mediating attentional technologies, we are accustomed to shifting the primary focus of our attention rapidly and on many time scales, from the differing second-to-second salience of features in the sensory environment, to the slower shifts between different conversations and activities in a complex setting (a crowded room at a party, shopping at a mall with a loose-knit group of friends). Over longer time scales

we also make meaning along the *traversals* (Lemke, 2002a, 2002b, 2003) of our channel surfing and web surfing, or the multiple scenes of our daily life. In the complexity of daily life, and increasingly with the affordances of new technologies, we are offered choices as to how we shall take up these affordances and what habits of engaging with multiple attentional worlds we each will prefer.

What kinds of meanings do we make, and what kinds of feelings do we experience, as we attend to simultaneous and sequential attentional worlds, with a constructed sense of spatiality, interactivity, and dynamic temporality within and across them? How do we integrate or reconcile the different paces and apparent flows of time, the interruptions and resumptions of events or activities, the disjointed topological spaces of our multiple attentional worlds? It seems to me that these questions are fundamental to any reflexive and critical use of the video medium in our research in the learning sciences.

### THE GENRES AND USES OF VIDEO IN RESEARCH

The kind of research with which most of us are concerned is the study of human learning and behavior in natural settings. We may also wish to study human meaning-making practices, affective responses, and the production, use, and interpretation of artifacts, whether more toollike or more textlike. I am going to propose that we also need to study human learning and behavior within, between, and across virtual attentional worlds. In the tradition of this kind of research, the use of video has multiple ancestry. It descends from the use of photography, of audio recording, and of ethnographic film making. I will be concerned less with the lineage of the medium and more with the lineage of the practices of using these materials in research.

Two early uses of photography in the human sciences were time lapse and stroboscopic photography in the study of human movement (e.g., Edgerton & Killian, 1939) and ethnographic photography (Collier & Collier, 1986) and film making (e.g., Mead & Bateson, 1952) in cultural anthropology. The first made use of the temporal dimensions of the medium, a forerunner of slow-motion and fast-forward video, and of cinematic animation. It enabled us to study the detailed articulations and transitions from one microtimed moment of a movement to the next. It points us to the importance of temporality in our uses of video. The second made us realize how little of what can be seen in any scene we actually attend to: Re-viewing a photographic record with a different “set,” a different purpose or interest, the original photographers, and others, could see far more than was remembered or noted on the occasion preserved in the photograph. Photographs were re-usable and almost endlessly rich as a data type. This points us to the role of the viewer in our use of research video.

One can also note that photographs can be juxtaposed and grouped and linked in a large number of possible ways, each of which can potentially inspire a particular story or document a particular insight, provided only that some interpretable basis of commonality or contrast is present (e.g., that all were taken in the same village in the same year, or in two different villages being compared, etc.). That the same is true of video episodes from a unified or coherent corpus (say, many short videos made by different students in a class each exploring some aspect of the same town) was recognized

by Goldman-Segall (1998), who created *Constellations* (and more recently *Orion*) as hyper textual environments for grouping, linking, and annotating video clips, precisely to allow us to aggregate multiple “points of viewing” in our research in the learning sciences.

Audio recording was employed by dialectologists eager to capture authentic samples of local speech and folklorists seeking to preserve local story traditions. Even more than time-lapse and stroboscopic photographs, audio recordings allowed us to play with time (e.g., slowing or looping speech to better transcribe it), and they too were susceptible to re-analysis, such as with acoustic oscilloscopes, or when Hymes (1981) advanced a theory of the role of prosodies in traditional story telling.

Ethnographic video descends most directly from ethnographic film and the pioneering work of Gregory Bateson in setting up a camera in his own home to record daily life. Now the issue of temporality comes very much to the fore in the process of production (i.e., recording) as well as in the use of the recorded product. If we leave the camera running all day and night, we miss nothing, but we then need 24 hours to view the recording. With fast forward, we can inspect the recording more quickly, but very imperfectly (especially for speech). If we want to do a detailed analysis on days or months worth of tapes, it could take years, perhaps more years than mere mortals have. This is still a primary problem of video research.

Ethnographic video is an even richer source of data for re-analysis than are photographs or audiotape. Not only does it include all the information that could be obtained from these two sources separately, but it provides information on the temporal relationships of speech and sounds to visually depicted actions and events. One of the first distinctive results from such video research was the discovery of interactional synchrony, both between speech and unconscious movements for a single individual, and between both of these for two different individuals in conversation or extended social interaction (Condon & Ogsten, 1967; Kendon, 1973). Moreover, it provides us with data on pacing and rates of speech and action, and on their potentially significant modulation and variation (accelerations and decelerations, contrasts of rapid and hesitant, interruptions and resumptions, breaking off, self-correction, pauses, etc.) These are perhaps even more evident in abstract animated diagrams, but the temporal dimension of video also provides the perceptual basis for many of our inferences of cause-effect relationships.

As researchers, our uses of the video medium have primarily been to record behavior and social interaction in their ordinary contexts and settings, and to make these recordings available for later and more leisurely inspection. Once again, the temporal affordances of the medium have been critical for us. Human social activity goes by too quickly for us to remember its details. With video, we can either slow it down or replay it again and again. We can then transcribe it and link the transcription (of speech, action, and events) to the relevant segments of the video, available for replay and further analysis (as well as for linking to other segments). We have been led by this tradition of using video to a very *microscopic* approach to the study of human learning and behavior. We have learned a great deal from it, but the approach does fundamentally distort and limit our view of human learning and social interaction. How so?

First, we typically examine behavior mainly on very brief time scales to the neglect of longer time scales. There are thousands of excellent analyses of 5 min episodes in school classrooms, very few of whole 40 min lessons, and almost none of either whole school days for individual pupils or teachers or a whole week (much less a whole year) in the life of one class. Similarly so for learning under artificial laboratory conditions. We do not even have a means of describing in brief the salient patterns in learning or social activity over longer time scales. Time-lapse video might give us one kind of start toward that, but by itself, it will not necessarily reveal salient patterns unless we already know what we are looking for.

Second, by putting brief episodes under the analytic microscope, we magnify small details and minor events out of all proportion to the flow of activity on a longer time scale. A hesitation here, a mis-speaking there, a self-correction or a mis-statement that may loom large in an utterance-by-utterance analysis may in fact be missed by most of the participants and have no consequences for the rest of a lesson. If a teacher uses a term 30 times in a lesson and mispronounces it once, will students care or pay much attention to the error? Some such examples are obvious, but more generally, how can we judge the relative significance of short-term events for longer term developments if we do not even have a descriptive apparatus for the longer time scales?

Meaning is not just made moment by moment. It is also made across longer time scales. So also with learning and development, with the performance of identity, with invention and discovery, with the emergence of social relationships, with the conduct of projects and tasks, the solution of problems, the creation of artifacts and texts. We need to evolve meso and macroscale uses of video to balance our inherited tradition of microscale uses.

Finally, video texts as research data are “unedited” in the sense that no cuts and joins have been made, no time periods left out, and generally there is a view from only a single camera, not intercut with other spatial–perspectival viewpoints. We also have a tradition of edited ethnographic film and video as interpretative presentations, in which there *are* juxtapositions of temporally separated scenes, recorded on different occasions, and with commentary added. Epistemological alternatives and their supporting technologies such as Goldman-Segall’s *Constellations* or *Orion* allow us in effect to make diverse interpretative complexes of video segments from a shared video database, to view other people’s, and to link and annotate these new wholes as well as the original segments. We also have available to us now picture-in-picture and picture-with-picture video that allows us to see (almost) simultaneously the synchronized video streams from different cameras viewing the same scene. A new technology (Pea, 2003) allows 360-degree video recording and a play back in which we can zoom in on the part of the scene on which we wish to focus, thus reducing the limitations of re-usability of data that come from the viewpoint selections of the original directional camera. Although audio is less directional, it is still true that there are multiple speech and sound sources in many scenes and we will want to be able to attend differentially to these without being limited by the placement of a single recording microphone. Selection among many video and audio data streams in an integrated, synchronized software environment is already possible in modest-cost research systems (Hay, 2003).

## RESEARCHERS INTERACTING WITH VIDEO MEDIA

Video is a seductive medium, especially when its visual realism and audio fidelity are high and the quality of sensory experience comes to resemble (or even exceed) that of ordinary perception of the immediate environment. It makes us believe that we are simply seeing what is there, rather than interacting with and interpreting in very specific ways a very partial (in both senses) record of an activity. This is particularly true when we view the recording at a rate of playback that seems natural for human action and speech. If we slow it down, we gain some analytic and interpretative distance and reflexivity, and even more so when we denaturalize it by speeding it up. Once we have viewed and reviewed the same scene many, many times it does begin to seem to us more like an artifact and less like a moment of life. Something of the same effect can occur if we zoom in on a detail or zoom out, or change our angle of view. The more we interact with and become conscious of deliberately affecting the visual experience presented by the medium, the less likely it is to seem merely a natural unfolding of recorded events.

But this critical perspective on our use of video is itself limited to a very specific time scale in the process of analysis, namely the minute-to-minute time scale of inspecting or replaying a part of the tape. What about the longer time scales of our uses of video; the whole longer activity of which the minute-to-minute re-viewings of segments are directly constituent parts? The activity in which we select which tape to view and what we are going to do with it, or look for in it? The activity of creating the record itself? The larger project of which the use of the video is one part? How can we get a critical and reflexive perspective, and some prophylactic practices to avoid unwarranted interpretation and claims?

This same problem has already been analyzed in the work of Bruno Latour and other researchers in the social studies of science and technology (Latour, 1987, 1999; Lynch & Woolgar, 1990), who look at scientists and what they are actually doing in moving from data collection and analysis to published results and claims. For our purposes here, a key analytic conceptualization from that body of work is the notion of “chains of translations of inscriptions.” In this view, what is happening over longer time scales is that one inscription (for us, say, a social phenomenon perceived in real time) is being “translated” by technical practices and technological mediations into another inscription (in a different medium, here, say, a video recording the phenomenon), and then into another (a transcription), and yet another (a running description or summary, a commentary, an article for publication). In many cases, the later inscriptions are embedded in artifacts (charts, tables, texts, “constellations”) in which they are integrated with inscriptions having a different prior chain (i.e., starting from different phenomena or events). Both these integrations and the procedures for translating an inscription to a new medium or form must insure faithful or defensible connections; connections that preserve the chain of logic and consequence that supports the eventual argument of the researcher.

We have a tendency as researchers to overclaim a reversibility for these chains of translations of inscriptions that we have constructed. Our positivist traditions in the learning sciences want to make claims about the reality we assume to lie at the head of



these chains. We have followed Cartesian dualist logic too far in dichotomizing realities and representations. All the links in a chain of translations are equally real; indeed their reality is attested precisely by our ability to include them in a chain, or more generally in networks of interlocking chains. All meaningful interactions with realities are also equally mediated by culture-specific interpretative codes, and thus share the feature that seems to distinguish “representations” as such, but does not.

In this view, the video, the researcher, the camera, the play-back apparatus, the transcript, the drafts of the article, and so forth, are all interdependent parts of a network, tied together by their roles in producing a sustainable chain. This is saying more than simply that the researcher plays an inescapable role in interpreting the video. It is saying that it is only in relation to their roles in the chain (and in other chains) that any of these entities or “actants” (including ourselves) has meaning (or even a determinable existence). Latour’s ontology does not allow a view from outside, even for the purposes of taking a picture of our own activity in order to be reflexive about it. All the more reason for us to complement it, nonetheless, with a more phenomenological perspective.

What are we doing when we come to view a video? In one sense we are making a traversal, and experiencing along with it some meanings and feelings, across at least two distinct attentional spaces; that of our office or lab and its activities, and that of the events in the virtual world in the video. We are, in Latour’s sense, trying to make a translation between these two worlds, to forge an enduring connection between them that will ultimately support the claims or interpretations we will want to make about what the video “shows.” We are doing this in real time and in real space, in our own material bodies, with meanings and feelings, sensations and actions. But unlike what we might be doing in the absence of video or other virtual attentional environments, we will necessarily find ourselves trying to forge connections between worlds where time may be flowing at different rates, where space can have different relative scales, where we can move backward and forward in virtual time, where we can interrupt action and resume it without consequence, where we can move our viewpoint at arbitrary virtual speeds or not at all, and where we can view multiple attentional spaces almost simultaneously, side by side, in temporal synchrony or not.

We are not just making meaning “within” the virtual world of the video as if we were an interpreting observer present at the original event, we are also and crucially making meaning *across* the world of our research work and of the world of video events. Experientially and phenomenologically this has to feel strange to us, and in many cases even disorienting. What kinds of manipulations of the video and cyclings back and forth between the two attentional worlds, in and out of the video “illusion,” are we comfortable with or deeply disturbed by? How does this bias or limit the kinds of uses of video we make in our research?

Imagine viewing a presentation of four video windows in each of which virtual time is passing at a significantly different rate (and so showing us what changing features are salient on each different time scale), and all different from the rate of time passing in our nonvideo environment. Imagine that each window zooms in rapid sequence to fill most of your visual field, with a pause between zooms to see all the windows at the same size (but with the videos still running a pace in each). Imagine now

that you can intervene in the automatic cycle and hold a window at maximum zoom, and then jump from window to window, while also attending to a graphic display that shows where the action in each window is on a timeline in relation to what is being shown in all the other windows? Maybe you would never want to do this (or maybe we will all be doing it in 10 years), but imagine how it might *feel*. How does it feel to do the kinds of work with video that we do now? What role do such feelings play in how we do our video research? What can reflection on these feelings tell us about how we are using the video medium and about the directions we are more or less likely to take in developing the medium and our uses of it in the future?

There are obviously limits to our sensory and cognitive capacities to process multiple video streams, to live with our attention rapidly cycling among radically different attentional worlds. But there are relatively few limits to our ability to create technological aids to enable us to ratchet down the cognitive and sensory demands of such uses to a viable level. Our uses of video media are themselves becoming, and likely in the future to become even more, mediated. The audiotape transcriber machine is the forerunner of a vast technology to enable us to traverse multiple video worlds on multiple space and especially time scales.

### **One More Future for Video Research**

The future of video research is only partly about getting more complete and multiple reusable kinds of video data. And it is only partly about devising new technologies to allow us to manipulate and connect multiple video texts and multiple views of the same video data. It is most importantly, I think, about the evolution of the medium itself and about the kinds of research projects in which we will want to enroll video data.

I believe that video games point the way to one likely future for research video. A generation of social researchers that has grown up playing video games is not going to rest until we have a technology in which we can visit recorded real-world scenes with all the flexibility of movement in a three-dimensional world that video games now provide for designed simulation worlds. We are not likely for a while yet to recreate the “holodeck” of Star Trek imagination (Murray, 1998) or to have full-scale immersive virtual reality, at least for underfunded social research. But we can certainly have something equivalent in its affordances to a video game or an animated VRML scene. What will be most significant, in terms of the points I have been making here about such a development, will be the increased interactivity for the researcher user. I think we can extrapolate from experience in other advanced media (full presence VR and to a lesser degree, high-end video games) to predict that the sense of immersion and telepresence for the researcher will increase with interactivity.

I need to be a little more precise here about what I mean by interactivity. I mean a sense of agency in the virtual world, so that it is not simply an observed attentional world, but a world in which there is the full feedback cycle of efference and afference (action and perceived response), which we humans interpret as causal efficacy. On the opposite side, I believe that other, more “external” kinds of manipulation (starting and stopping the scene, replaying it, rewinding it, fast forwarding, etc.) will reduce the sense of presence and immersion. Why does this matter? Partly because of the delicate

balance researchers must maintain between the immersion that leads to insight and the distance that allows critical reflection. But more generally because video is already a powerfully affective medium. Watching a video, even on a small screen, even in black and white, even with bad audio fidelity, can make you cry. It can make you feel fear or at least apprehension and dread. It can be sexually arousing. It can make you feel good, even elated. That power increases with all the qualities that promote a sense of attentional engagement and immersive presence in the virtual world. Researchers are going to have to learn how to deal with our own affective responses to video data.

As we should already be doing. For all our commitment to reflexivity in research, where is the research on how researchers engage with video as a medium? What do we know about the complex processes of making meaning and experiencing feelings not just when attentionally engaged in the video world, but also when moving in and out of it? What do we know, systematically, about the affective elements of our engagement with video and how they play a role in the larger agendas of our research? And in our choices among research agendas and procedures in the learning sciences?

I would like to close with a modest proposal. If the arguments I have given so far make sense, then there is a larger research enterprise in which we may wish to engage. We may want to learn more generally about how people make meaning and experience feelings in engaging with attentional media and in moving across the interface between the immediate and mediated attentional worlds. And we may want to find out how people make meaning and experience feeling when they move within and among multiple virtual attentional spaces, as they do in video games, or in jumping among windows on a conventional computer display.

There is increasing interest today in ethnographic research on spatiality and temporality and on the mediation of meaning, feeling, and action by artifacts and structures in the material environment, as well as by social interaction, and how all these interact with one another. One promising conceptual approach uses Bakhtin's (1981) notion of the "chronotope": a typical pattern of timings and paces of action interconnected with movements in and among particular settings. Bakhtin saw chronotopes as important organizing characteristics of literary narrative genres. Very likely they are also organizing patterns in life activity and in the design of video games. We could use this concept to help us combine research on engagement with attentional media and research on the role of space, time, and artifacts in the mediation of culturally meaningful activity by examining these latter phenomena in the context of video games.

Video games foreground affect. They create simplified spatial environments and genre-typical chronotopes of action. They make it easy to observe artifact and social mediation of action. They afford a range of time scales and variation in pacing of action. Some are even beginning to permit manipulation of time within the game world, for example *Blinx: The TimeSweeper* (Artoon, 2002) and *Prince of Persia: The Sands of Time* (Ubisoft, 2003). In playing most video games, there are frequent opportunities to pause the action of the game to consider strategy or resources, and so there is often a cycling between the attentional world of the game play and various supplementary "screens" where other kinds of actions are performed, relevant to the game world but experientially outside it and distinct from it as attentional spaces (while still being within the overall game environment as designed). Players also move between their

immediate nongame environment and the attentional space of the game play, and sometimes these worlds interact with one another in various ways. We might find a group of players on a local-area network (e.g., students in a school computer lab during lunch hour) who are talking to each other in the conventional world, while communicating “in character” to one another in the game world, and acting collaboratively or competitively in both worlds (but not necessarily in the same combinations or alliances). We might also find players on wide-area networks, where they have both in-game interaction with other players-as-characters and CHAT or IM (Instant Messaging) interactions with the same players in real time, outside the game and the game system windows, but still within the attentional space of the computer monitor display.

A deep ethnography of computer gaming could have much to tell us not only about this new video medium, and basic issues of meaning making, affect, spatiality, temporality, and artifact mediation, but also about the dimensions of our own interactions as researchers with the video media of our present and our future. We need to frame our video epistemologies both in and outside the box, and most especially along our own traversals into, through, and out of their multiple attentional spaces.

## REFERENCES

- Artoon. (2002). *Blinx the Timesweeper* [Digital Game]. USA: Microsoft Game Studios.
- Bakhtin, M. M. (1981). Forms of time and of the chronotope in the novel. In M. Holquist (Ed.), *The dialogic imagination* (pp. — —). Austin, TX: University of Texas Press.
- Brock, M. C. (1949–1955). *Captain Video and his Video Rangers* [Television Program]. In J. L. Caddigan (Producer). USA: DuMont Television Network.
- Collier, J., & Collier, M. (1986). *Visual anthropology: Photography as a research method*. Albuquerque, NM: University of New Mexico Press.
- Condon, W. S., & Ogsten, W. D. (1967). A segmentation of behavior. *Journal of Psychiatric Research*, 5, 221–235.
- Edgerton, H. E., & Killian, J. R. (1939). *Flash! Seeing the unseen by ultra high speed photography*. Boston, MA: Hale, Cushman & Flint.
- Electronic Arts. (2003). *The Lord of the Rings: The Return of the King* [Digital Game]. USA: Electronic Arts.
- Goldman-Segall, R. (1998). *Points of viewing children's thinking: A digital ethnographer's journey*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hay, K. (2003). *Integrated temporal multimedia data (ITMD) research system* Retrieved from <http://lpsl.coe.uga.edu/live/itmd/>
- Hymes, D. (1981). *In vain I tried to tell you*. Philadelphia: University of Pennsylvania Press.
- Jackson, P. (2003). *The lord of the rings: The return of the king* [Motion picture]. USA: New Line Cinema.
- Kendon, A. (1973). The role of visible behavior in the organization of social interaction. In M. von Cranach & I. Vine (Eds.), *Social communication and movement* (pp. 29–74). New York: Academic Press.
- Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.
- Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA: Harvard University Press.
- Lemke, J. L. (2002a). Discursive technologies and the social organization of meaning. *Folia Linguistica*, 35(1–2), 79–96.
- Lemke, J. L. (2002b). Travels in hypermodality. *Visual Communication*, 1(3), 299–325.
- Lemke, J. L. (2003). The role of texts in the technologies of social organization. In R. Wodak & G. Weiss (Eds.), *Theory and interdisciplinarity in critical discourse analysis* (pp. 130–149). London: Macmillan/Palgrave.

- Lynch, M., & Woolgar, S. (1990). *Representations in scientific practice*. Cambridge, MA: MIT Press.
- Mead, M., & Bateson, G. (1952). Trance and dance in Bali [Film]. *Character Formation in Different Cultures*. USA.
- Murray, J. H. (1998). *Hamlet on the holodeck*. Cambridge, MA: MIT Press.
- Pea, R. (2003). *DIVER: Point-of-view authoring of panoramic video tours for learning, education and other purposes*. Santa Clara, CA: EOE Foundation.
- Prichett, H. W., & Wyckoff, E. B. (1953–1957). *Winky Dink and you* [Television broadcast]. In Barry/Enright/Friendly Production (Producer). USA: CBS-TV.
- Ross, G. (1998). *Pleasantville* [Film]. USA: New Line Cinema.
- Shiny Entertainment. (2003). *Enter the matrix* [Digital Game]. USA: Atari.
- Ubisoft. (2003). *Prince of Persia: The sands of time* [Digital Game]. USA: Ubisoft.
- Wachowski, A., & Wachowski, L. (2003). *The matrix reloaded*. USA: Warner Bros.
- Zemeckis, R. (1988). *Who framed Roger Rabbit?* [Film/Animation]. USA: Buena Vista Pictures.



