

# **TO LIE AND TO ACT: POTEMKIN'S VILLAGES, CINEMA AND TELEPRESENCE**

## **Notes around Checkpoint '95 project**

**Lev Manovich**

PUBLISHED IN:

Ars Electronica 1995 catalog, Linz, Austria, 1995.

In Checkpoint '95, three immobile cars are stationed in three different cities: Linz, New York, and Moscow. A little toy vehicle, equipped with a television camera, is running over the Nibelungen Bridge. An image picked up by this camera is simultaneously transmitted to Linz, New York, and Moscow. By looking at this image, a driver sitting in a stationary car remotely operates the toy vehicle.

Throughout human history, representational technologies have served two main functions: to deceive the viewer and to enable action, i.e. to allow the viewer to manipulate reality through representations. Fashion and make up, paintings, dioramas, decoys and virtual reality fall into the first category. Maps, architectural drawings, x-rays, and telepresence fall into the second. To deceive the viewer or to enable action: these are the two axes which structure the history of visual representations.

In Checkpoint '95 these axes come together. The image seen by the driver is a deception, corresponding not to the immediate space

outside of the car but to a remote location. Yet the same image enables the driver to affect physical reality at this remote location. The project also combines technologies of telecommunication, television and telepresence with references to older representational technologies: fake architecture and cinema. What are the new possibilities for deception and action offered by these recently developed technologies? Using the elements as Checkpoint '95 as a starting point, this essay will reflect on this question.

## I. To Lie

### 1.

Consider the driver in Moscow. The driver is sitting inside a stationary car. Instead of viewing the space behind the windshield, she or sees an image transmitted from a remote location. The image deceives the driver, substituting its virtual space for the real space outside the car. In short, the image is a lie. (I will return to this shortly.)

What are the elements involved in this deception? A vehicle; a window showing fake reality (in other words, a window functioning as a screen); Russia. My first association provoked by these elements is with Potemkin's Villages. According to the historical myth, at the end of the eighteenth century, Russian ruler Catherine the Great decided to travel around Russia in order to observe first-hand how the peasants lived. The first minister and Catherine's lover, Potemkin, had ordered the construction of special fake villages along her projected route. Each village consisted of a row of pretty facades. The facades faced the road; at the same time, to conceal their artifice, they were positioned at a considerable distance. Since Catherine the Great never left her carriage, she returned from her journey convinced that all peasants lived in happiness and prosperity.

This extraordinary arrangement can be seen as a metaphor for

life in the Soviet Union. There, the experience of all citizens was split between the ugly reality of their lives and the official shining facades of ideological pretense. However, The split took place not only on a metaphorical but also on a literal level, particularly in Moscow -- the showcase Communist city. When prestigious foreign guests visited Moscow, they, like Catherine the Great, were taken around in limousines which always followed few special routes. Along these routes, every building was freshly painted, the shop windows displayed consumer goods, the drunks were removed, having been picked up by the militia early in the morning. The monochrome, rusty, half-broken, amorphous Soviet reality was carefully hidden from the view of the passengers.

In turning selected streets into fake facades, Soviet rulers adopted eighteenth century technique of creating fake reality. But, of course, the twentieth century brought with it a much more effective technology: cinema. By substituting a window of a carriage or a car with a screen showing projected images, cinema opened up new possibilities for deception.

Fictional cinema, as we know it, is based upon lying to a viewer. A perfect example is the construction of a cinematic space. Traditional fiction film transports us into a space: a room, a house, a city. Usually, none of these exist in reality. What exists are the few fragments carefully constructed in a studio. Out of these disjointed fragments, a film synthesizers the illusion of a coherent space.

The development of the techniques to accomplish this synthesis coincides with the shift in American cinema between approximately 1907 and 1917 from a so-called "primitive" to a "classical" film style. Before the classical period, the space of film theater and the screen space were clearly separated much like in theater or vaudeville. The viewers were free to interact, come and go, and maintain a psychological distance from the cinematic diegesis. Correspondingly,

the early cinema's system of representation was presentational: actors played to the audience, and the style was strictly frontal.[1] The composition of the shots also emphasized frontality.

In contrast, classical Hollywood film positions each viewer inside the diegetic space. The viewer is asked to identify with the characters and to experience the story from their points of view. Accordingly, the space no longer acts as a theatrical backdrop. Instead, through new compositional principles, staging, set design, deep focus cinematography, lighting and camera movement, the viewer is situated at the optimum viewpoint of each shot. The viewer is "present" inside a space which does not really exist. A fake space.

(In general, Hollywood cinema always carefully hides the artificial nature of its space, but there is one exception: rear screen projection shots. A typical shot shows actors sitting inside a stationary vehicle, as in *Checkpoint '95*; a film of a moving landscape is projected on the screen behind car's windows. The artificiality of rear screen projection shots stands in striking contrast against the smooth fabric of Hollywood cinematic style in general.)

The synthesis of a coherent space out of distinct fragments is only one example of how fictional cinema deceives a viewer. A film in general is comprised from separate image sequences. These sequences can come from different physical locations. Two consecutive shots of what looks like one room may correspond to two places inside one studio. They can also correspond to the locations in Moscow and Linz, or Linz and New York. The viewer will never know.

This is the key advantage of cinema over older fake reality technologies, be it eighteenth century Potemkin's Villages or nineteenth century Panoramas and Dioramas. Before cinema, the deception was limited to the construction of a fake space inside a real space visible to the viewer. Examples include theater decorations and military decoys. In the nineteenth century, Panorama offered a small

improvement: by enclosing a viewer within a 360-degree view, the area of fake space was expanded. Louis-Jacques Daguerre introduced another innovation by having viewers move from one set to another in his London Diorama. As described by Paul Johnson, its "amphitheater, seating 200, pivoted through a 73-degree arc, from one 'picture' to another. Each picture was seen through a 2,800-square-foot-window."<sup>[2]</sup> But, already in the eighteenth century, Potemkin had pushed this technique to its limit: he created a giant facade -- a Diorama stretching for hundred of miles -- along which the viewer (Catherine the Great) passed. In cinema a viewer remains stationary: what is moving is the film itself.

Therefore, if the older technologies were limited by the materiality of a viewer's body, existing in a particular point in space and time, film overcomes these spatial and temporal limitation. It achieves this by substituting recorded images for unmediated human sight and by editing these images together. Through editing, images that could have been shot in different geographic locations or in different times create an illusion of a contiguous space and time.

Editing, or montage, is the key twentieth technology for creating fake realities. Theoreticians of cinema have distinguished between many kinds of montage but, for the purposes of sketching the archeology of the technologies of deception, I will distinguish between two basic techniques. The first is so montage within a shot: separate realities form contingent parts of a single image. (One example of this is a rear screen projection shots.) The second technique is the opposite of the first: separate realities form consecutive moments in time. This second technique of temporal montage is much more common; this is what we usually mean by montage in film.

In a fiction film temporal montage serves a number of functions. As already pointed out, it creates a sense of presence in a virtual space. It is also utilized to change the meanings of individual

shots (recall Kuleshov's effect), or, rather, to construct a meaning from separate pieces of pro-filmic reality.

However, the use of temporal montage extends beyond the construction of an artistic fiction. Montage also becomes a key technology for ideological manipulation, through its employment in propaganda films, documentaries, news, commercials and so on.

The pioneer of this ideological montage is Dziga Vertov. In 1923 Vertov analyzed how he put together episodes of his news program "Kino-Pravda" (Cinema-Truth) out of shots filmed at different locations and in different times. This is one example of his montage: "the bodies of people's heroes are being lowered into the graves (filmed in Astrakhan' in 1918); the graves are being covered with earth (Kronshtad, 1921); gun salute (Petrograd, 1920); eternal memory, people take down their hats (Moscow, 1922)." Here is another example: "montage of the greetings by the crowd and montage of the greetings by the machines to the comrade Lenin, filmed at different times." [3] As theorized by Vertov, through montage, film can overcome its indexical nature, presenting a viewer with objects which never existed in reality.

2.

Outside of cinema, montage within a shot becomes a standard technique of modern photography and design (photomontages of Alexander Rodchenko, El Lissitzky, Hannah Hoch, John Heartfield and countless other lesser-known twentieth century designers). However, in the realm of a moving image, temporal montage dominates. Temporal montage is cinema's main means of creating fake realities.

After the World War II a gradual shift takes place from film-based to electronic image recording. This shift brings with it a new technique: keying. One of the most basic techniques used today in any video and television production, keying refers to combining two different image sources together. Any area of uniform color in one

video image can be cut out and substituted with another source. Significantly, this new source can be a live video camera positioned somewhere, a pre-recorded tape, or computer generated graphics. The possibilities for creating fake realities are multiplied once again.

With electronic keying becoming a part of a standard television practice in the 1970s, not just still but also time-based images finally begin to routinely rely on montage within a shot. In fact, rear projection and other special effects shots, which had occupied marginal presence in a classical film, became the norm: weather man in front of a weather map, an announcer in front of footage of a news event, a singer in front of an animation in a music video.

An image created through keying presents a hybrid reality, composed of two different spaces. Television normally relates these spaces thematically, but not visually. To take a typical example, we may be shown an image of an announcer sitting in a studio; behind her, in a cutout, we see news footage of a city street. If classical cinematic montage creates an illusion of a coherent space and hides its own work, electronic montage openly presents the viewer with an apparent clash of different spaces.

What will happen if the two spaces seamlessly merge? This operation forms the basis of a remarkable video "Steps" directed by Zbigniew Rybczynski in 1987. "Steps" is shot on video tape and uses keying; it also utilizes film footage and makes an inadvertent reference to virtual reality. In this way, Rybczynski connects three generations of fake reality technologies: analog, electronic and digital. He also reminds us that it was the 1920s Soviet filmmakers who first fully realized the possibilities of montage which continue to be explored and expanded by electronic and digital media.

In the video, a group of American tourists is invited into a sophisticated video studio to participate in a kind of virtual reality / time machine experiment. The group is positioned in front of a blue

screen. Next, the tourists find themselves literally inside the famous Odessa steps sequence from Eisenstein's "Potemkin." Rybczynski skillfully keys the shots of the people in the studio into the shots from "Potemkin" creating a single coherent space. At the same time, he emphasizes the artificiality of this space by contrasting the color video images of the tourists with the original grainy black and white Eisenstein's footage. The tourists walk up and down the steps, snap pictures at the attacking soldiers, play with a baby in a crib. Gradually, the two realities begin to interact and mix together: some Americans fall down the steps after being shot by the soldiers from Eisenstein's sequence; a tourist drops an apple which is picked up by a soldier.

The Odessa steps sequence, already a famous example of cinematic montage, becomes just one element in a new ironic re-mix by Rybczynski. The original shots which were already edited by Eisenstein are now edited again with video images of the tourists, using both temporal montage and montage within a shot, the latter done through video keying. A "film look" is juxtaposed with "video look," color is juxtaposed with black and white, the "presentness" of video is juxtaposed with the "always already" of film.

In "Steps" Eisenstein's sequence becomes a generator for numerous kinds of juxtapositions, super-impositions, mixes and re-mixes. But Rybczynski treats this sequence not only as a single element of his own montage but also as a singular, physically existing space. In other words, the Odessa steps sequence is read as a single shot corresponding to a real space, a space which could be visited like any other tourist attraction.

### 3.

The next generation in fake reality technologies is digital media. Digital media does not bring any conceptually new techniques. It simply expands the possibilities of joining together different image sources



within one shot. Rather than keying together images from two video sources, we can now composite an unlimited number of image layers. A shot may consist of dozens or even hundreds of layers, all having different origins: film shot on location, computer-generated sets or actors, digital matte paintings, archival footage and so on. Most current Hollywood films, not only "Jurassic Park" or "Terminator 2," contain such shots.

Historically, a digitally composited image, like an electronically keyed image, can be seen as a continuation of montage within a shot. But while electronic keying creates disjointed spaces reminding us of the avant-garde collages of Rodchenko or Moholy-Nagy from the 1920s, digital compositing brings back the nineteenth century techniques of creating smooth "combination prints" like those of Henry Peach Robinson and Oscar G. Reijlander. However, what in the nineteenth century was only a still image now can become a moving one. A moving nineteenth century "combination print": this is the current state of the art in the technologies of visual deception.

## II. To Act

### 1.

So far, I have considered the historical connections between some of the technologies of deception, evoked in Checkpoint '95: fake architectural spaces, montage, video keying. I will now consider the second axis which structures the history of visual representations.: action.

I will begin by returning to Checkpoint '95. A little toy vehicle, equipped with a television camera, is running over the Nibelungen Bridge. An image picked up by this camera is simultaneously transmitted to Linz, New York, and Moscow. The driver sitting in a

stationary car located in one of these cities is wearing a head-mounted display. The display allows the driver to simultaneously see her hands on the car wheel as well as the image transmitted from the toy vehicle, thus making possible to remotely operate the toy vehicle. In short, the driver becomes "telepresent."

If we look at the word itself, the meaning of the term "telepresence" is presence over distance. But presence where? Brenda Laurel defines telepresence as "a medium that allows you to take your body with you into some other environment... you get to take some subset of your senses with you into another environment. And that environment may be a computer-generated environment, it may be a camera-originated environment, or it may be a combination of the two." [4] In this definition, telepresence encompasses two different situations: being "present" in a synthetic computer-generated environment (what is commonly referred as virtual reality ) and being "present" in a real remote physical location via a live video image. Scott Fisher, one of the developers of NASA Ames Virtual Environment Workstation, similarly does not distinguish between being "present" in a computer-generated or a real remote physical location. Describing Ames system, he writes: "Virtual environments at the Ames system are synthesized with 3-D computer-generated imagery, OR are remotely sensed by user-controlled, stereoscopic video camera configurations." [5] Fisher uses "virtual environments" as all-encompassing term, reserving "telepresence" for the second situation: "presence" in a remote physical location. [6] I will follow his usage here.

Both popular media and the critics have downplayed the concept of telepresence in favor of virtual reality. The photographs of the Ames system, for instance, have been often featured to illustrate the idea of an escape from any physical space into a computer-generated world. The fact that a head-mounted display can also show a televised image of a remote physical location was hardly ever mentioned.

And yet, from the point of view of the history of the technologies of deception and action, telepresence is a much more radical technology than virtual reality, or computer simulations in general. Let us consider the difference between the two.

Like fake reality technologies which preceded it, virtual reality provides the subject with the illusion of being present in a simulated world. Virtual reality goes beyond this tradition by allowing the subject to actively change this world. In other words, the subject is given control over a fake reality. For instance, an architect can modify an architectural model, a chemist can try different molecule configuration, a tank driver can shoot at a model of a tank, and so on. But, what is modified in each case is nothing but data stored in a computer's memory! The user of any computer simulation has power over the virtual world which only exists inside a computer.

Telepresence allows the subject to control not just the simulation but reality itself. Telepresence provides the ability to remotely manipulate physical reality in real time through its image. The body of a teleoperator is transmitted, in real time, to another location where it can act on subject's behalf: repairing a space station, doing underwater excavation or driving a toy vehicle over the Nibelungen Bridge.

Thus, the essence of telepresence is that it is anti-presence. I don't have to be physically present in a location to affect reality at this location. A better term would be teleaction. Acting over distance. In real time.

Catherine the Great was fooled into mistaking painted facades for real villages. Today, from thousands of miles away (as it was demonstrated during the Gulf War) we can send missile equipped with a television camera close enough to tell the difference between a target and a decoy. We can direct the flight of the missile using the image transmitted back by its camera, we can carefully fly towards the target.

And, using the same image, we blow the target away. All that is needed is to position the computer cursor over the right place in image and to press a button.

2.

How new is this use of images? Does it originate with telepresence?

Since we are accustomed to consider the history of visual representations in the West in terms of illusion, it may seem that to use images to enable action is a completely new phenomenon.

However, French philosopher and sociologist Bruno Latour proposes that certain kinds of images have always functioned as instruments of control and power, power being defined as the ability to mobilize and manipulate resources across space and time.[7]

One example of such image-instruments analyzed by Latour are perspectival images. Perspective establishes the precise and reciprocal relationship between objects and their signs. We can go from objects to signs (two-dimensional representations); but we can also go from such signs to three-dimensional objects. This reciprocal relationship allows us not only to represent reality but also to control it. For instance, we cannot measure the sun in space directly, but we only need a small ruler to measure it on a photograph (the perspectival image par excellence).[8] And even if we could fly around the sun, we would still be better off studying the sun through its representations which we can bring back from the trip -- because now we have unlimited time to measure, analyze, and catalog them. We can move objects from one place to another by simply moving their representations: "You can see a church in Rome, and carry it with you in London in such a way as to reconstruct it in London, or you can go back to Rome and amend the picture." Finally, we can also represent absent things and plan our movement through space by working on representations: "One cannot smell or hear or touch Sakhalin Island, but you can look at the map

and determine at which bearing you will see the land when you send the next fleet." [9] All in all, perspective is more than just a sign system, reflecting reality -- it makes possible the manipulation of reality through the manipulation of its signs.

Perspective is only one example of image-instruments. Any representation which systematically captures some features of reality can be used as an instrument. In fact, most types of representations which do not fit into the history of illusionism -- diagrams and charts, maps and x-rays, infrared and radar images -- belong to the second history: that of representations as instruments for action.

Given that images have always been used to affect reality, does telepresence bring anything new? A map, for instance, already allows for a kind of teleaction: it can be used to predict the future and therefore to change it. To quote Latour again, "one cannot smell or hear or touch Sakhalin Island, but you can look at the map and determine at which bearing you will see the land when you send the next fleet."

In my view, there are two fundamental differences. Because telepresence involves electronic transmission of video images, the constructions of representations takes place instantaneously. Making a perspectival drawing or a chart, taking a photograph or shooting film takes time. Now I can use a remote video camera which capture images in real-time, sending these images back to me without any delay. This allows me to monitor any visible changes in a remote location (weather conditions, movements of troops, and so on), adjusting my actions accordingly.

The second difference is directly related to the first. The ability to receive visual information about a remote place in real time allows us to manipulate physical reality in this place, also in real-time. If power, according to Latour, includes the ability to manipulate resources at a distance, then teleaction provides a new and unique kind of power:

real-time remote control. I can drive a toy vehicle, repair a space station, do underwater excavation, operate on a patient or kill -- all from a distance.

What technology is responsible for this new power? Since teleoperator acts with the help of a live video image, we may think at first that it is the technology of video, or, more precisely, of television, if we recall the original nineteenth century meaning of television: vision over distance. Only after the 1920s, when television was equated with broadcasting, does this meaning fade away. However, during the preceding half a century (television research begins in the 1870s), television engineers were mostly concerned with the problem of how to transmit consecutive images of a remote location to enable "remote seeing."

If images are transmitted at regular intervals and these intervals are short enough, the viewer will have enough reliable information about the remote location for teleaction. Modern television images are based on scanning reality at the resolution of a few hundred lines sixty times a second (the early television systems used slow mechanical scanning and the resolution as low as thirty lines). Radar images are based on scanning reality once every few seconds reducing the visible to a single point. A radar image does not contain any indications about shapes, textures or colors present in a television image -- it only records the position of an object. Yet this information is quite sufficient for the most basic teleaction: to destroy an object.

So the technology which makes teleaction possible turns out to be electronic transmission of signals, in other words, electronic telecommunication. Electricity and electromagnetism, these discoveries of the nineteenth century, are what allows for the new and unprecedented relationship between objects and their signs in teleaction. Electronic telecommunication makes instantaneous not only the process by which objects are turned into signs but also the

reverse process -- manipulation of objects through these signs.

3.

Umberto Eco once defined a sign as something which can be used to tell a lie. This definition correctly describes one function of visual representations -- to deceive. But in the age of electronic telecommunication we need a new definition: a sign is something which can be used to teleact.

## NOTES

1. On presentational system of early cinema, see Charles Musser, *THE EMERGENCE OF CINEMA: THE AMERICAN SCREEN TO 1907* (Berkeley: University of California Press, 1990), 3.

2. Paul Johnson, *THE BIRTH OF THE MODERN: WORLD SOCIETY 1815-1830* (London: Orion House, 1992), 156.

3. Dziga Vertov, "Kinoki. Perevorot" (Kinoki. A revolution), *LEF* 3 (1923): 140.

4. Brenda Laurel, quoted in Rebecca Coyle, "The Genesis of Virtual Reality," in *FUTURE VISIONS: NEW TECHNOLOGIES OF THE SCREEN*, edited by Philip Hayward and Tana Wollen (London: British Film Institute, 1993), 162.

5. Scott Fisher, "Visual Interface Environments," in *THE ART OF HUMAN-COMPUTER INTERFACE DESIGN*, edited by Brenda Laurel (Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1990), 430. Emphasis mine -- LM.

6. Fisher defines telepresence as "a technology which would allow remotely situated operators to receive enough sensory feedback to feel like they are really at a remote location and are able to do different kinds of tasks." Fisher, 427.

7. Bruno Latour, "Visualization and Cognition: Thinking with Eyes and Hands," KNOWLEDGE AND SOCIETY: STUDIES IN THE SOCIOLOGY OF CULTURE PAST AND PRESENT 6 (1986): 1-40.

8. Latour, 22.

9. Latour, 8.