Social Immersive Media
Pursuing Best Practices for Multi-user Interactive Camera/projector Exhibits

Scott S. Snibbe[1] and Hayes S. Raffle[2, 3]

[1]Sona Research	[2]Nokia Research Center	[3]Tangible Media Group
1073 Howard Street	955 Page Mill Road #200	MIT Media Lab, E15-350
San Francisco, CA 94103	Palo Alto, CA 94304	20 Ames St. Cambridge, MA 01239
scott@sonaresearch.org	hayes.raffle@nokia.com	hayes@media.mit.edu

ABSTRACT
Based on ten years' experience developing interactive camera/projector systems for public science and culture exhibits, we define a distinct form of augmented reality focused on social interaction: *social immersive media*. Our work abandons GUI metaphors and builds on the language of cinema, casting users as actors within simulated narrative models. We articulate philosophical goals, design principles, and interaction techniques that create strong emotional responses and social engagement through visceral interaction. We describe approaches to clearly communicate cultural and scientific ideas through the medium. And we demonstrate how practitioners can design interactions that promote specific social behaviors in users.

Author Keywords
Augmented reality, camera-based interaction, embodied interaction, social computing, computer vision, animation, cinema, learning

ACM Classification Keywords
H.5.1 Multimedia Information Systems: Artificial, augmented, and virtual realities

INTRODUCTION
Over the past three decades, interface designers have developed increasingly more immersive strategies to engage viewers beyond the GUI metaphor in collaboration with interactive graphics systems. Two extremes of this development are virtual reality systems, whose ultimate aim is to artificially generate all channels of sensory input for a single person by placing her in a wholly synthetic world; and what we define here as *social immersive media*: immersive media that favors interaction in a shared social space using a person's entire body as the "input device," unencumbered by electronics or props.

Social immersive media allows users' bodies to control projected interactive graphics and video by collocating a camera or other remote sensors with interactive projected

graphics, accompanied by audio and other media. A subset of augmented reality, these experiences are distinguished by unencumbered full-body interaction using invisible sensing infrastructures [26,27,31,20], rather than wearable technologies, to augment a user's senses [16].

We have created over a hundred social immersive exhibits for science, history and art museums since 1998. Museums incorporate interactive projected exhibits because visceral interaction is particularly suitable for creating enthusiasm and engagement with challenging or unfamiliar topics. Reflecting on the limits of static media, Tom Rockwell, Director of Exhibits for the Exploratorium states: "There are huge areas of science that are too big, too small, too fast or too slow. Snibbe's work lets museum visitors enter these inaccessible realms in new and more immersive ways than ever before." [43] While single channel video sometimes serves the role of emotionally engaging visitors, museums increasingly include exhibits that physically engage visitors, since research indicates that interaction improves the learning process [39]. Confucius succinctly summarizes the benefit of interactivity for learning in his well-known 5th century BCE quote: "I hear and I forget. I see and I remember. I do and I understand."

The public and social context of museums and their increase in civic popularity demands new interaction designs for multiple users that depart from interactive kiosks that were once a mainstay of museums. Social immersive media leverages the public and social nature of the museum context, where pairs, groups or crowds often approach an exhibit at once. Unlike interactive kiosks, social immersive media accommodates the public, social, and informal learning that museums champion. Social immersive media exhibits' success is due in part to meeting young visitors on their own terms with experiences that are *emotional*, *social* and *physical*. These experiences engage people's whole bodies to viscerally "operate" an exhibit, and elicit strong emotional responses: while one would never expect to see a visitor in a science museum jumping up and down and laughing in front of a text panel or movie, it is a common occurrence with social immersive media.

This paper confronts the basic questions of design for this medium: what is the language of this social medium; how do we control and modulate people's responses and

behavior; and how can we design experiences for the greatest educational and cultural impact?

CONTEXT AND RELATED WORK
Our work builds on camera-based interactive research in interactive arts, tangible interfaces, and interactive games.

Interactive Arts
Kruger's Videoplace pioneered full-body interaction by combining video cameras with collocated projected animation [26]. In Videoplace, a user steps into a room with a camera pointed directly at him. Facing the user is a video projection screen upon which his colored silhouette appears as an actor in a simulated two-dimensional virtual world. This playful world includes interactions with animated characters, and remote collaboration with another user of a Videoplace system.

Since Videoplace, other artists and interaction designers have contributed to computer vision interface research. Rokeby's Very Nervous System is a physical environment that responds to full body movements with a changing musical composition [44]; Simpson's Shadow Garden [48] uses shadows to interact with simulated water and butterflies; Utterback's Text Rain [56] drops words onto users' video images; Rozin's mirrors [45] create graphical and electromechanical reflections of the user's body; and Levin's Manual Input Sessions [30] combine real hand shadows with computer graphics.

Tangible Interfaces
Wellner's Digital Desk pioneered pointing a camera and projector at the same surface for unencumbered interaction. On a tabletop in front of a user, the system responds to two-handed gestures with simulated graphics, simultaneously integrating real and virtual documents and tools [60].

Like the Digital Desk, Underkoffler's IO Bulb situates camera/projector systems in an intimate physical context to produce interactive surfaces [55]. For instance, an application for urban planning augments building models with dynamically changing projected light-studies and wind simulations. Underkoffler establishes interaction techniques for camera/projection systems on a small interactive surface using physical props manipulated by one or more users, as well as exploring users' hands themselves as parts of the interface.

Tangible interface researchers draw on and extend the interface principles of direct manipulation for interactive graphics [22]: particularly the principles of real-world metaphors for graphical interfaces and rapid incremental feedback to continuous user input.

Interactive Games
Freeman's computer vision techniques use the whole body as an input device to a videogame with a camera and monitor facing the user [17]. Sony and Nintendo adapted Freeman's ideas commercially for full-body interactive games with the EyeToy [15], which also uses a camera; and with the Wii [37], which, instead of a camera, uses an inertial sensor for full-body interaction.

Exertion Interfaces [36] use camera/projector systems to create collaborative games that require physical exertion. Although the researchers' focus is improving remote communication via full-body interaction, their findings on how to encourage social interaction, physical exertion and enjoyment are relevant to our work.

Several systems expand interaction to audiences of thousands. Cinematrix allows an audience to interact with interactive games via red/green colored paddles [9]; Interactive Audience Participation allows audiences to collaboratively control on-screen video games by leaning and other gestures [32]; and Squidball tracks the 3D location of large inflated balls over an audience to affect interactive projected games [5].

Commercial examples of camera-based interaction include the Reactrix and Gesturetek systems [41,19], which project interactive advertisements onto the floors and walls of public spaces. These commercial systems' interfaces often treat the body as a pointing device, activating discrete interactive projected "buttons" or "sliders". Such designs are based on interaction techniques inherited from GUIs. In contrast, the previously cited examples dynamically respond to the continuous, rather than discrete, gestures of a user or audience. This is also the focus of our work.

In the following sections, we outline a philosophy and design principles for social immersive media, and review exhibit case studies to illustrate how our interaction designs can direct people's behaviors and experiences (Figure 1). Our work is substantially influenced by cinema, casting users as actors within a simulated dynamic narrative model.

SOCIAL IMMERSIVE MEDIA
"Our body applies itself to space like a hand to an instrument, and when we wish to move about we do not move the body as we move an object. We transport it without instruments as if by magic, since it is ours and because through it we have direct access to space."
— Maurice Merleau-Ponty

Philosophy
Our approach to media is motivated by the philosophy of phenomenology, pioneered by Edmund Husserl and elaborated by Maurice Merleau-Ponty [33]. This philosophical view asserts that "reality" only exists as an interplay between consciousness, the perceived world, and our bodies. We each generate a pre-conscious understanding of reality that comes from sensorimotor perception. In simpler terms, we each have a "gut" visceral mode of perception that precedes the analytic, symbolic and linguistic. Language appears to be a higher level and more powerful means to affect the world. Yet we cannot affect our most prized possession, our body, with words at all.

We strive to engage users at this phenomenological level by creating interactive media that is first understood by the

body and later understood rationally. What does it mean to make media that engages with the visceral part of a person's mind? This question leads to an abandoning of language and its arbitrary signifiers, in favor of non-symbolic and enactive representations of knowledge [6].

Our visceral design philosophy has two important positive side effects. First, these experiences successfully engage users of all ages. Younger viewers 4-10 years old absorb the educational and cultural message only as a secondary experience to their main experience of physical play, since children this age experience the world primarily through full-body interaction [6,39]. In contrast, older users may try to understand what an exhibit is "about" intellectually before physically engaging. The second side effect is that, by focusing on visceral experience and avoiding language, these experiences can be understood across cultures without modification or translation.

Multi-touch interfaces face the same challenge of socially integrating multiple people's bodies, but in the more intimate arena of a tabletop. In many multi-touch interfaces, multiple users' hands are interpreted merely as a collection of discrete cursors that interface using GUI concepts of pointing, touching, scaling, and mode selection to create a series of individual experiences on the same surface [7]. In these cases, multi-touch fails to be social and to leverage the principles of direct manipulation that we see both in the physical world and in intuitive computer interfaces [22]. However, when multi-touch meaningfully responds to the continuously varying gestures of simultaneous multiple users, as in Wilson's work [61], then its full social and visceral potential is realized. As in the best multi-touch work, we have developed techniques that abandon discrete interactions in favor of continuous ones, and that focus on visceral interrelationships between multiple users.

In designing social interactive media, we look as much to the history of cinema as to the history of HCI and interaction design. Cinema is a mature time-based medium for emotional engagement and storytelling. We have found principles of cinema and animation to be overwhelmingly useful in designing for social immersive media [25, 29]. However, there is one significant case where this is not true: with rare exception, we have found that the cut, one of the most powerful narrative tools of cinema [3], often disengages a user from an immersive experience.

Design Principles
We have developed the following design principles to reward users' participation by being visceral, responsive and personalized while supporting social interaction. Social immersive media are:

- *Visceral*. The media is experienced physically and emotionally, through whole-body interactions, before it is experienced symbolically or rationally.

- *Responsive*. The media responds immediately, clearly and predictably to users' actions.
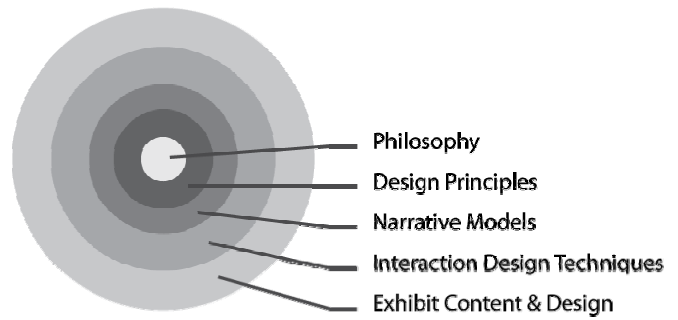

Figure 1. Guidelines for designing users' interactions with social immersive media.

- *Continuously variable*. The media, like our natural environment, is continuously changing with infinite variability, usually through simulated response to the user's image, silhouette, sound, location, gestures or other uniquely identifiable features.

- *Socially scalable*. Interactions are designed to share with others. Furthermore, interaction, representation, and users' engagement and satisfaction should become *richer* as more people interact. The unbreakable rule is that if the exhibit fits more than one person, it must work with more than one person.

- *Socially familiar*. The media should augment and reinforce existing collocated social behaviors.

- *Socially balanced*. Interaction equally emphasizes a user's awareness of herself, other users, and the media itself.

We highlight each principle in the case studies below to demonstrate their application.

NARRATIVE MODEL CASE STUDIES
The following case studies, drawn from a decade of work by Scott Snibbe and Snibbe Interactive, present interaction designs that cast users as actors improvising within a specific temporal structure. We call these temporal structures *narrative models*, as the choice of model affects the visitor's perception of an exhibit's "story." Each case study demonstrates a narrative model and highlights the implementation of social immersive media design principles. Following the case studies we generalize and formalize a set of design strategies (Figure 1).

Experiential Narrative: *Boundary Functions*
Boundary Functions (Figure 2) projects lines between each of the people walking upon its 4m x 4m floor [46]. These lines demarcate each person's personal space: the space closer to him than to anyone else. As more and more people walk onto this floor, each person's tile changes dynamically to create an overall pattern mathematically described as a Voronoi diagram. The Voronoi pattern relates to phenomena of nature at every scale: from bubbles to the gravitational influence of stars. When people on the floor touch each other, the line between them momentarily disappears, viscerally removing the border between them.
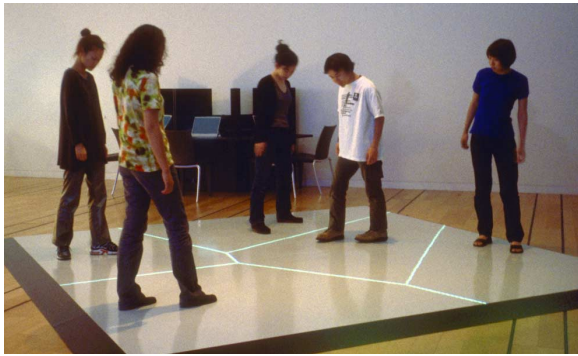
Figure 2. Boundary Functions: experiential narrative.

Figure 3. Deep Walls: performance narrative.

Boundary Functions demonstrates the principles of social immersive media. The exhibit is *visceral*, by responding to a user's whole body movements and each body's relationship to every other nearby body. The exhibit is *responsive*, updating at 60 Hz and responding to subtle changes in posture. It is *continuously variable* in that each tile changes continuously to reflect sub-pixel shifts in the unique location and posture of each person standing on the floor. It is *socially scalable*: the exhibit becomes more visually and socially complex as more people step on its surface. The exhibit is *socially familiar* by making visible our innate sense of personal space. And the exhibit is *socially balanced,* in that people view and interact with the projection as frequently as the do with each other. In fact, Boundary Functions often enlivens social interaction and provokes strangers to converse. Through all of these features, we argue that the work is a truly welcoming, yet not sensorially overwhelming social interface that people modulate in and out of their attention to seamlessly blend with natural social interactions in a manner suggested by the principles of ambient media [23]. In Boundary Functions, social communication flows more easily, achieving the effects of a cocktail party without the alcohol!

Beyond visceral play, Boundary Functions contains an educational message and has been included in exhibitions on mathematics, geometry, sociology, psychology, physics and animal behavior. Museum planners have chosen this exhibit not merely for education, but to create enthusiasm for topics that, to many children, seem uninteresting or irrelevant. Finally, Boundary Functions also contains an element of culture and philosophy, appealing to the more sophisticated adult user or art enthusiast. By showing that what we call "personal" space is only defined by our relationship to others and changes without our control, the exhibit suggests a view of reality as interdependent rather than individualistic, provoking thought and conversation.

Boundary Functions can also be seen as a cultural psychological experiment in the domain of proxemics [2], which studies how closely people stand to each other in different circumstances and cultures. We have observed trends in how people in different countries behave differently in the exhibit. In Germany, people move far from each other. In Japan, people immediately come close

to each other and strangers even touch. In New York City, people stand on the floor and do not move—perhaps because space is so scarce in New York and users are more hesitant to give it up. We believe these differences may be attributable to differences in cultural psychology.

We describe the narrative model of Boundary Functions as *experiential*. There is a continuous "reality" created by the exhibit whose behavior is predictable, yet continuously variable. The analogue to this experience is a natural environment like a lake, where the responses of the water's ripples are predictable, yet never exactly the same.

One common critique of Boundary Functions is that there is no response with a single user. In one sense this is a failure to fully satisfy the *socially scalable* criterion. One argument in favor of this technique is that a work explicitly about social interaction should indeed have no response except in a social environment. Another argument is that, like the power of silence in music, there is a power in non-responsiveness or a blank screen if used to convey meaning.

Performance Narrative: *Deep Walls*

Deep Walls [51] asks how you can make a movie with your body alone. A naïve approach suggests a photo booth interface, where users watch a countdown to begin; have a set time duration during which they act; and enjoy their performance during a playback period. However, this model makes the user subservient to the media, forcing him to keep track of modes and violating the *socially familiar*, *socially balanced*, and *responsive* criteria of social immersive media. Our approach puts users' bodies in charge of both the performance *and* its duration.

In Deep Walls (Figure 3), users record silhouette movies by performing with their shadows in front of a projection screen. The duration of a performance is determined by the most natural of dramatic gestures: entrances and exits. When users walk in front of the projection screen, the exhibit begins to record their shadows. Upon the last user's exit, the resulting movie loops in one of sixteen cells. If the users danced for five seconds, they see a five second movie. If they acted for five minutes, they see a five-minute movie. The piece collects the last sixteen movies recorded by visitors in a grid, and the cinematic contents of the grid are limited only by users' own imaginations.
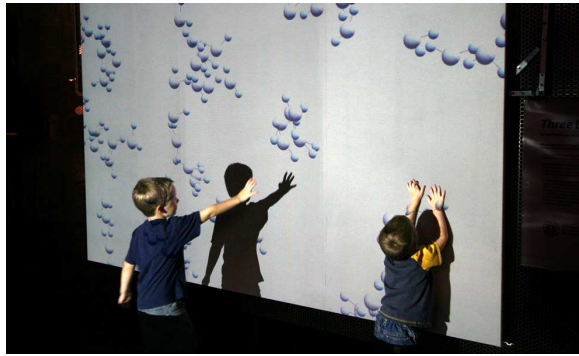
Figure 4. Three Drops: episodic narrative.

Figure 5. Fear: game narrative.

Deep Walls is inspired by Alexander's architectural concept of "thick walls" in which architects design walls that can be carved into shelves and cabinets, creating a home that becomes custom-tailored by its inhabitants [1]. The piece becomes an interactive wall that absorbs and "remembers" the acts taking place before it.

Users describe surprise and excitement in seeing what others have performed in the piece. People collaborate not only with other actors in front of the projector, but also with movies of previous visitors, in a recursive interaction loop. Despite the simplicity of this interaction technique—where all content comes from the users themselves—users sometimes engage with the piece for hours on end, and return to interact again and again over the course of years.

Deep Walls is: *visceral*, using the whole body as interface; *responsive* by immediately replaying movies at 24-30 Hz; *continuously variable* by creating a blank canvas to record any duration or type of performance; *socially scalable* by accommodating group performances without any mode change; *socially familiar* by echoing the culture of a dance floor; and *socially balanced*, alternating between engaged performance and animated dialogue as users switch from performer to viewer.

Deep Walls is structured as an improvised *performance narrative*: both the users and the piece as a whole are constantly moving forward like an improvisational dance or sketch comedy. The eventual erasure of each episode, far from reducing the energy of the interaction, increases it, as people strive to create a new performance that dramatically "comments" on the prior performances, or that re-enacts a performance that has disappeared.

A limitation of the performance narrative approach is that it is difficult to impose a more structured narrative because people are focused on their own movements and driven to increasingly energetic and flamboyant acts. Delaying the playback of movies until after users leave the projection creates confusion for many users and, in later versions of this piece, we initiate playback *immediately* within one of the frames, which dramatically improves users' ability to find their performance and understand the interaction.

Episodic Narrative: *Three Drops*

Three Drops [52] allows science museum visitors to become participants in a water simulation at three distinct scales, in order to gain familiarity with differences in behavior at the micro- and nanoscale (Figure 4). In the first episode of Three Drops, users take a virtual shower under a simulated showerhead. As people stand in front of the projection the water splashes, puddles, and flows around them. In the next phase of Three Drops, single drops of water drip one at a time onto the visitors. Here, water is magnified a hundredfold. At this scale, the surface tension of water dominates; water behaves so differently that concept of getting "wet" no longer make sense: a drop of water becomes a giant beach ball to catch and throw. Finally, zooming in further, users encounter individual water molecules. These molecules clump together in chains, and stream by from left to right, attracted to users' shadows as if they were positively charged impurities.

Visitors engage with this exhibit using full body movements: playfully showering, cooperatively bouncing simulated water droplets, and even standing still—a rare sight in a museum filled with children—to make water molecules coalesce around their bodies. Studies describe people engaging with the piece for over two minutes—a long time for science museums—and indicate improved learning compared to briefer interactions with passive video documentation of the same material. We believe these longer interactions are due to the *visceral* nature of users' interactions: people apply their whole bodies to the exhibit; the *responsiveness* of the simulation (30 Hz) to users' subtle movements; the *continuously variable* interactions with physical and molecular simulations rather than pre-recorded animations; the *socially scalable* design that responds to every user in frame, and, with the drop, the transformation into a game of catch with multiple users; *socially familiar* actions that echo everyday activities; and a *socially balanced* narrative that alternates between watching the simulation and playing with other users.

We went through several iterations to arrive at Three Drops' successful final form. We originally planned three side-by-side installations to represent three different scales, however a survey of museums indicated that this amount of space and equipment was impractical, so instead we cut between the three episodes on a single screen. A user study

of the first version [47] found that while people loved interacting with the exhibit, the abrupt cuts from one episode to another (i.e. from human scale shower to 100x scale of water droplets) confused them. People did not understand that they were still interacting with simulated water. This prompted us to add zoom/dissolve transitions between scales in the style of Eames' *Powers of 10* [14], and to add ambient sound to each episode to emphasize that we are still manipulating water. The sound for each episode is, in turn, the sound of a shower, a drip, and, for the molecules, a flowing river. At the end of the molecular episode, the exhibit zooms back out to human scale.

The narrative model of Three Drops is *episodic*. The benefit of this model is that we can tell a more complex story, bringing users through narrative episodes that have different imagery and interaction. Unlike a movie, since users arrive continuously, the "narrative" must be designed to remain understandable if entered at any time. The pitfall of this model is losing attention at the transition between episodes, which requires careful design to make a transition brief enough that attention is not lost, or interactive in its own right to maintain engagement.

Game Narrative: *Fear*
Exhibits that engage users in goal-oriented narratives borrow from interactive video games, a genre that is rapidly moving towards full-body, immersive interaction. Recent studies of Nintendo's Wii [38], Sony's EyeToy [28], exertion interfaces [36] and other visceral games [58] describe the social and physical benefits of integrating immersive media with game content. While turn-taking games include a social element, we feel a central concern of our designs should be to create content that instead allows simultaneous, collocated collaboration and competition among users. We return to our design principle of *social scalability*: if the exhibit fits more than one user, it should support more than one user. We challenge designers of physically active video games to reach beyond two players alternating in turn, and focus on users' collocated interactions with each other.

Fear (Figure 5) is an interactive exhibit on the science behind feeling scared developed for the Science Museum of Minnesota and the California Science Center. The exhibit creates a projected virtual environment in which users' synthetic shadows appear in a jungle environment via a camera pointed directly at them. In addition to their own shadows, users see the shadows of a pacing jaguar, and fruits dropping from a tree above. The visitors' goal is to collect the fruit without being "eaten." If children move towards the tree when the jaguar is not looking, their shadows appear in grey. Kids collect the tree's fruit in their silhouette arms and hands. When the jaguar turns to look at the players, each player that freezes sees his shadow now in outline, indicating that the jaguar can't see that person, because he is standing still. When someone starts to move while the jaguar looks at him, his shadow turns pink, then red, to indicate that the jaguar has seen him. If he doesn't

immediately freeze, the jaguar pounces and attacks the user's shadow, which collapses as the jaguar gnaws, eliciting squeals of mock fear and delight.

We believe adherence to the principles of social immersive media make this an educationally successful and popular exhibit teaching the freeze or flee response by being *visceral*; *responsive*; and *continuously variable* through incorporating visitors' actual shadows, having a simulated rather than pre-recorded "death," and a continuously variable simulation for collecting fruit. The exhibit is *socially scalable* by allowing users to pass fruit to each other, and, by letting them compete with each other by pushing one another while the jaguar is looking to cause an attack; the *social familiarity* of the exhibit comes from its clear children's book-like representation; and it is *socially balanced*, as kids talk, jostle and laugh together turning their attention from the screen to each other.

After a player has "died" in the exhibit, his shadow no longer appears on the screen. Unlike the previous exhibits, as a game narrative, Fear has a clear end point and indicates this to the user with the disappearance of his shadow. People then leave the exhibit, which is an important requirement of the museum to avoid bottlenecks or too long of a wait for other visitors.

In designing this experience, we wanted to prevent users from reaching into the tree branches, or stepping into the jaguar's home behind the tree, because we thought interaction there would be nonsensical and distract from the game goals. By pointing the camera at the users rather than at the screen, we can add an important constraint to this exhibit: we shrink down the visitors' silhouettes and clamp their interaction to a smaller rectangle beneath and to the left of the tree: if a user reaches too high, or walks beyond the tree, his or her projected silhouette elegantly disappears.

TOWARDS A DESIGN LANGUAGE
In the case studies, we have shown how our philosophy, design principles, and use of narrative models guide our approach to creating social immersive media. We now summarize specific interaction design techniques and cinematic, environmental, and software design principles, in the service of developing a cohesive design language for this medium.

Interaction Design Techniques
The following interaction design techniques can be applied to engage users and direct their behavior in social immersive media exhibits (Table 1):

Choice of Narrative Model
We have found the deliberate choice of a narrative model important to shaping an overall social immersive media experience. As discussed in our case studies, the choice of narrative model promotes dwell times varying from minutes to hours. *Experiential* (Boundary Functions) and *performance* (Deep Walls) models promote open-ended experiences with long total interaction time; while *episodic*

(Three Drops), and *game* (Fear) narrative models encourage constrained, shorter experiences.

Continuous vs. Discrete Interaction

A natural, but misguided approach to social immersive media is to translate the GUI metaphor to social space, where bodies behave like cursors and the screen holds a series of discrete buttons. This approach can be seen in the Eyetoy and Reactrix systems. The problem with this approach is that human social and physical interaction is naturally not discrete, consisting of singular specific events, but rather *continuous*, comprising continuously varying movements and sounds. Researchers in direct manipulation for screen interfaces have also observed the importance of continuous fine-grained feedback [22]. In unmediated full-body interactive experiences, objects should respond continuously and directly to the changing full-body gestures of users, rather than restrict the body to act as a pointer that activates buttons and widgets.

As mentioned earlier, we have found that discrete changes in temporal sequence—cinematic cuts—can disengage a user from the flow of his interactive experience. The artists of *structuralist cinema*, who explored cut-less cinema, inspired this response. With multi-minute long single takes, structuralist films can become for the viewer an experience in their own right rather than the depiction of someone else's experience we find in traditional cinema [49].

Recording and Replaying

In the service of creating *continuously variable* and *socially familiar* experiences, we often create a recording of the user either as a shadow or a motion recording of her gestures [53]. Replaying user movements creates a lively alternation between users' recording and observing, and promotes attention. The educational and communicational benefits of high fidelity human movement has also been explored and evaluated in tangible media research [4,18,40].

Shadow and Silhouette

Using a silhouette or shadow of a user has surprising advantages over a full-color representation. We observe that a picture of a shadow is a quite faithful representation of the shadow (a flat absence of light), while a picture of a person (a 2D variation of color) is substantially dissimilar from our 3D bodies. Thus, paradoxically, a shadow may appear more "real" to the user than a video image. Furthermore, we have observed people's discomfort with a color video representation of themselves, concerned and distracted by their hair, skin, clothes and weight. This may result from self-consciousness or the "uncanny" [34] likeness of a live video projection of oneself. A shadow representation immediately alleviates these concerns. Finally, users are able to immediately recognize a silhouette of themselves and their friends; however, their shadow remains anonymous to strangers. This ambiguity is not only socially disinhibiting, but it also satisfies laws about privacy, particularly those regulating the capture and display of children's images. Other researchers have also noted the power of shadows in user interfaces [21,27, 35,50]. Our

Interaction Technique	Boundary Functions	Deep Walls	Three Drops	Fear
Narrative Model	Experiential	Performance	Episodic	Game
Continuous Interaction	X	X	X	X
Record & Replay		X		
Shadow / Silhouette		X		X

Table 1. Interaction design techniques of the case studies.

work focuses on the use of silhouettes to create a sense of immediate personalization in an interactive experience.

Principles of Character Animation

Like movies, social immersive media are time varying projected experiences. As we design the movement and layout of interface objects, we principally look to the mature time-based medium of cinema, and particularly Thomas and Johnson's seven principles of animation, to create emotionally expressive animated responses [24]. Other user interface professionals have noted the value of adhering to these principles [10]:

- *Easing in and easing out*. Natural actions ease in slowly and come to a halt gradually. Ignoring this principle produces robotic movement and abrupt change.

- *Overlapping action*. Natural movements overlap in time. For example, a person does not walk toward a door, halt, then reach for the doorknob. Rather, he simultaneously extends his arm while walking towards the door.

- *Follow through*. Natural actions do not abruptly end, but rather follow through, in the same way a baseball swing does not abruptly end after hitting the ball.

- *Staging*. Since users can only pay attention to one thing at once, the designer must direct users attention to different portions of an experience at different times. This is accomplished by choosing when and where on-screen elements change. In general, there should be one dominant action upon which a user's attention is focused.

- *Squash and stretch*. Real objects deform as they move and interact with the world, either by squashing and stretching as fleshy bodies do, or by motion blur, when captured by a camera.

- *Exaggeration*. Since animated representations are usually simplified compared to the natural world, in order to get the same or greater emotional impact, the designer must exaggerate features, motions, and reactions.

- *Timing*. Human beings are very sensitive to the timing of movements. The same animated poses can be interpreted very differently depending on the time taken to interpolate between them. For example, a head turning in 30ms could mean a character heard a gunshot, while a head turning over 500ms indicates he is feeling pensive.

Environmental considerations

As discussed by Buxton [8] and others, scale is a powerful tool to engage different user behaviors. In large-scale museum installations, a principal architectural consideration is to physically situate an exhibit to invite interaction. We have developed spatial guidelines that assure a successful exhibit. First, the exhibit should be in an area of natural human traffic: it is still unnatural for users to expect a video projection to be interactive, so we place the exhibit where they will naturally walk into it. Exhibits should have significant extra space around or in front of the exhibit, so that users can experiment easily and safely with entering and exiting. For that reason, a projection in a cul-de-sac is the least desirable situation. For social immersive media, exhibit lighting should not be too dark, which contradicts what designers might first expect. It is just as important for users to see each other as it is for them to see the exhibit, facilitating social interaction and group learning. To avoid damage, exhibit equipment should be out of reach in the ceiling or well protected by transparent sheet or other materials. Finally, it is essential, and often a legal requirement, to make exhibits accessible to the disabled by incorporating gently sloping ramps to raised areas and wide passageways to accommodate wheelchairs.

Multiple Software Representations

The choice of software model of the user's body and gestures, and of the graphical response, must change to the most appropriate for a given interface element. Employing multiple software representations of users is inspired by interactive graphics and vision systems for real-time interaction [31,62] and by non-photorealistic animation techniques [54]. Both fields have recognized that certain problems are simple to solve in one domain (such as collision detection in a raster image), while complicated or slow to solve using another representation (vector). We transform camera images into simultaneous raster and vector representations, and both discrete and continuous sets. We capture and analyze user gestures as 1d, 2d, 3d and continuity models, and as higher-level abstractions of movement (e.g. hand/head, fast/slow, concave/convex).

The insight not only to employ multiple representations, but also to change representation on the fly, comes directly from the history of cinema and animation. Starting in the 1920s with pioneers like Emile Cohl, animation and special effects artists have changed cinematic representation from frame-to-frame to the technique most convenient to accomplish an effect; switching from live action to animation, stop-motion, or optical printing [11].

DISCUSSION: DESIGNING SOCIAL BEHAVIOR

When we first installed Boundary Functions we noticed that, in addition to describing people's social relationships, the piece also changed peoples' ways of relating to each other. In particular, by drawing a line on the floor, users seem to almost universally want to step on that line. Due to the *responsiveness* of the exhibit, however, the line slips away. This produces a positive feedback loop where, the quicker the line slips away, the harder and harder a user tries to step on that line. Since making this first observation, we have learned to control user behavior by choosing different interaction techniques. Examples include:

- *Energizing.* By providing interactive elements that slip away (such as lines on the floor), or by repeating users' own movements, we can increase the amount of movement and positive emotion in a social situation.

- *Calming.* By rewarding stillness, as in *Three Drops'* last segment, where molecules coalesce around viewers' shadows, we can induce stillness and mindfulness [59].

- *Competing.* By having a finite resource or score that users compete for (such as collecting fruit); a penalty they must avoid (being killed, or halting interactivity); or side-by-side comparison of personalized media elements, we can provoke competition.

- *Performing.* By incorporating a representation of the user, particularly an animated one as in *Deep Walls*, we encourage performance even from people who describe themselves as shy. This is highly dependent upon the ability of the representation to preserve privacy.

- *Disinhibition.* By rewarding gross movements either by responsiveness, or by incorporating a representation of the user, behavior can become increasingly uninhibited. Darrell observed this in his *Magic Mirror*, which distorts people's faces via computer vision. While observing the direct video-out from the exhibit, the authors noted that people twisted their faces into grotesque forms because the exhibit was already distorting their face [13].

- *Learning.* Rather than operating a simulation using remote controllers, users become part of the simulation itself, exploring and internalizing new ideas. Our installations may serve as what Vygotsky refers to as a *pivot* [57]: a transitional object that helps users move towards internalizing abstract concepts such as those demonstrated by our four case studies: personal space (Boundary Functions), cinematic recording (Deep Walls), the nanoscale (Three Drops), and the freeze and flee response (Fear). We are encouraged by Crowley's observation that, "the complexity, extended time-scale, and socially-embedded nature of *in vivo* learning and development are not just annoyances to be controlled by an experimenter; they are fundamental, irreducible characteristics of how learning actually occurs." [12]

- *Dwell time.* Museums often have very specific requirements for dwell time—in some cases requiring an exhibit to push people along after less than a minute; in others wishing to engage visitors indefinitely. We have discussed some of the techniques for limiting dwell time above: by incorporating a cinematic cut; by removing the visitors' avatar; or by demarking a clear beginning and end. The key techniques for increasing dwell time are to be *continuously variable,* and to be personalized by

users' own movement. Another is by inviting competition with an unlimited cap on the scoring mechanism.

Designers of social interactive exhibits need to remain aware that they are designing behavior. These social behavior design principles establish a theoretical foundation that combines with our specific interaction design techniques to support the design process. Designers must inevitably complement these principles with the creative processes of intuition, feedback and experimentation.

CONCLUSION AND FUTURE WORK

The design principles outlined in this article serve as a series of best practices for creating highly effective and engaging augmented reality experiences that promote not just individual but *social* engagement. Our philosophy, approach, and design principles formalize a language for large-scale multi-user exhibits that other designers may build on.

We have participated in several user studies and are interested in how to evaluate social immersive media both for mechanically measurable criteria such as dwell time, and also for less easily measurable social criteria such as learning. Reeves wrote an excellent summary of the challenges in evaluating this medium and differences between user interface, art, and museum approaches to evaluation [42]. Integrating the disparate ways of evaluating interactive media across the fields of HCI, psychology and museology is an important challenge for the field.

ACKNOWLEDGMENTS

REFERENCES

1. Alexander, C., *A Pattern Language*. New York, Oxford University Press, 1977.

2. Argyle, M. *Bodily Communication*. International Universities Press, 1977.

3. Bowser, E. *The Transformation of Cinema, 1907-1915*, University of California Press, 1994.

4. Brave, S. and Dahley, A., inTouch: A Medium for Haptic Interpersonal Communication. *Ext. Abstracts CHI '97*, ACM Press, 363-364.

5. Bregler, C. et al., Squidball: An Experiment in Large-Scale Motion Capture and Game Design. *Intelligent Technologies for Interactive Entertainment*, 2005, 23-33.

6. Bruner, J. *Towards a Theory of Instruction*. Belknap Press: 1966.

7. Buxton, W., Multi-Touch Systems that I Have Known and Loved. Microsoft Research, 2007.

8. Buxton, W., Fitzmaurice, G., Balakrishnan, R., Kurtenbach, G. Large displays in automotive design. *IEEE Computer Graphics & Applications*, July 2000.

9. Carpenter, L., Video Imaging Method and Apparatus for Audience Participation, US Patent 5,365,266, 1994.

10. Chang, B. and Ungar, D., Animation: from cartoons to the user interface. *Proc. UIST*, 1993.

11. Crafton, D., *Before Mickey: The Animated Film*, 1898-1928, University of Chicago Press, 1993.

12. Crowley, K. et al., *Designing for Science: Implications From Everyday, Classroom, and Professional Settings*. Lawrence Erlbaum Associates, Inc., 2001.

13. Darrell, T. et al., Magic Morphin' Mirror: Person Detection and Tracking. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998, p. 964.

14. Eames, C. and Eames, R. *Powers of Ten*, 1977. http://www.powersof10.com

15. EyeToy. Sony Corporation. http://www.eyetoy.com

16. Feiner, S. K., Augmented Reality: A New Way of Seeing: Computer scientists are developing systems that can enhance and enrich a user's view of the world. *Scientific American*, April, 2002.

17. Freeman, W. T. et al, Computer Vision for Interactive Computer Graphics, *IEEE Computer Graphics and Applications*, Vol. 18, No. 3, May-June 1998.

18. Frei, P. et al., Curlybot: designing a new class of computational toys. *Proc. CHI 2000*, 129-136.

19. GestureTek. http://www.gesturetek.com

20. Hämäläinen, P., Ilmonen, T., Höysniemi, J, Lindholm, M., Nykänen, A. Martial Arts in Artificial Reality. *Proc. CHI 2005*.

21. Herndon, K., Zeleznik, R., Robbins, D., Conner, D., Snibbe, S., Van Dam, A. Interactive shadows, *Proc. UIST 1992*, 1-6.

22. Hutchins, E. L., Hollan, J. D., and Norman, D. A. Direct Manipulation Interfaces. In *Human-Computer interaction: A Multidisciplinary Approach*, R. M. Baecker, Ed. Morgan Kaufmann Publishers, San Francisco, CA, 468-470. 1987.

23. Ishii, H. et al., ambientROOM: integrating ambient media with architectural space. *Proc. CHI '98*, 173-174.

24. Thomas, F. and Johnson, O. *Disney Animation: The Illusion of Life*, Abbeville Press, 1981.

25. Klein, Norman M., The Electronic Baroque: Jerde Cities. in *You Are Here: The Jerde Partnership*

International, ed. Frances Anderton. London, Phaidon Press Limited, 1999, 112-121.

26. Krueger, M., Gionfriddo, T. and Hinrichen, K. Videoplace - An Artificial Reality. *Proc. CHI 1985*.

27. Krueger, M., Environmental Technology Making the Real World Virtual. *Communications of the ACM*, 36,7 (July 1993), 36-37.

28. Larssen, A.T. et al., Understanding Movement as Input for Interaction. A Study of Two Eyetoy Games. *Proc. CHI, 2004.*

29. Laurel, B. *Computers as Theatre*. Addison-Wesley Pub. Comp. Reading, Massachusetts, 1991.

30. Levin, G. and Lieberman, Z. Sounds from Shapes: Audiovisual Performance with Hand Silhouette Contours in The Manual Input Sessions. *Proc. NIME 2005*.

31. Maes, P. et al, The ALIVE System: Wireless, Full-Body Interaction with Autonomous Agents, *ACM Multimedia Systems*, 1996.

32. Maynes-Aminzade, D., Pausch, R. and Seitz, S. Techniques for Interactive Audience Participation. *Proc. ICMI 2002*.

33. Merleau-Ponty, M. *Phenomenology of Perception* (Transl. Colin Smith). Routledge, 2002.

34. Mori, Masahiro. Bukimi no tani / The uncanny valley (MacDorman, K. F. & Minato, T. Trans.). Energy, 7(4), 33–35, 1970

35. Minomo, Y., Kakehi, Y., Lida, M. Transforming Your Shadow into Colorful Visual Media – Multi-Projection of Complementary Colors. *Proccedings of Advances in Computer Entertainment (ACE) 2006*. ACM Press.

36. Mueller, F., Agamanolis, S., Picard, R. Exertion Interfaces: sports over a distance for social bonding and fun, *Proc. CHI 2003*.

37. Nintendo Wii. http://www.nintendo.com/wii

38. Parker, J.R., Games for physical activity: A Preliminary Examination of the Nintendo Wii. *Sixth International Symposium on Computer Science in Sport*, 2007.

39. Piaget, J. *The Grasp of Consciousness*. Cambridge: Harvard University Press, 1976.

40. Raffle, H., Parkes, A., and Ishii, H. Topobo: a constructive assembly system with kinetic memory. *Proc. CHI 2004*, 647-654.

41. Reactrix. http://www.reactrix.com

42. Reeves, S. Research techniques for augmented reality. Technical Report, University of Nottingham, 2004.

43. Rockwell, T. San Francisco Exploratorium. Videotaped interview. http://www.snibbeinteractive.com, 2005.

44. Rokeyby, D. Transforming Mirrors: Subjectivity and Control in Interactive Media. In *Critical Issues in Interactive Media*. Simon Penny, SUNY press, 1996.

45. Rozin, D. http://en.wikipedia.org/wiki/Daniel_Rozin and http://www.smoothware.com/danny/

46. Schenck et al, *Cyberarts 98: International Compendium Prix Ars Electronica*. Ars Electronica, 1998.

47. Science Museum of Minnesota, Comparative Report Formative Evaluation Exhibit Package Prototypes. May, 2007.

48. Simpson, Z. Shadow Garden. *SIGGRAPH 2002 Electronic Art and Animation Catalog*.

49. Sitney, P., *Visionary Film*, 2d edition, New York, 1979.

50. Shoemaker, G., Tang, A., and Booth, K., Shadow Reaching: A New Perspective on Interaction for Large Wall Displays. *Proc. UIST 2007*.

51. Snibbe, S. Body, Screen and Shadow. *San Francisco Media Arts Council (SMAC) Journal*, 2003. http://snibbe.com/scott/SMAC/smac_jan03.html.

52. Snibbe, S. Three Drops. *ArtNano: New Approaches for Visualizing the Nanoscale*, 2006. http://www.nisenet.org/artnano/artists/snibbe/artwork/.

53. Snibbe, S. and Levin, G. Interactive Dynamic Abstraction, *Proc. NPAR 2000*, Annecy, France: ACM, 21-29.

54. Strothotte, T. and Schlechtweg, S. *Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation*. Morgan Kaufmann, 2002.

55. Underkoffler, J., Ullmer, B., and Ishii, H., Emancipated Pixels: Real-World Graphics In The Luminous Room. *Proc. SIGGRAPH '99*, pp. 385-392.

56. Utterback, C. Text Rain. http://www.camilleutterback.com/textrain.html

57. Vygotsky, L.S. *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press, 1978.

58. Warren, J. *Unencumbered Full Body Interaction in Video Games*. Master's Thesis, MFADT Program, Parsons School of Design, 2003.

59. Weiser, M., Brown, J.S. The coming age of calm technology. In Denning, P.J., Metcalfe, R.M. (eds.), *Beyond Calculation: The Next Fifty Years of Computing*. Copernicus, 1998.

60. Wellner, P. Interacting with paper on the DigitalDesk, *Communications of the ACM*, v.36 n.7, 87-96, July 1993.

61. Wilson, A. D. et al. Bringing physics to the surface. UIST '08. ACM, New York, NY, 67-76. 2008.

62. Zeleznik, R. et al. An object-oriented framework for the integration of interactive animation techniques, *ACM SIGGRAPH Computer Graphics*, v.25 n.4, 105-112, July 1991.