Looking Glass: A Field Study on Noticing Interactivity of a Shop Window

Removed for blind review

ABSTRACT

In this paper we present our findings from a lab and a field study investigating how passers-by notice the interactivity of public displays. We designed an interactive installation that uses visual feedback to the incidental movements of passers-by to communicate its interactivity. The lab study reveals: (1) Mirrored user silhouettes and images are more effective than avatar-like representations. (2) It takes time to notice the interactivity (approx. 1.2s). In the field study, three displays were installed during three weeks in shop windows, and data about 807 persons interacting were collected. Our observations show: (1) Significantly more passers-by interact when immediately showing the mirrored user image (+90%) or silhouette (+47%) compared to a traditional attract sequence with call-to-action. (2) Passers-by often notice interactivity late and have to walk back to interact (the *landing effect*). (3) If somebody is already interacting, others begin interaction behind the ones already interacting, forming multiple rows. Our findings can be used to design public display applications and shop windows that more effectively communicate interactivity to passers-by.

Author Keywords

Interactivity, Noticing Interactivity, Public Displays, User Representation

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Miscellaneous

INTRODUCTION

A major challenge to create engaging public displays is that people passing by are usually not aware of any interactive capabilities. Unlike privately owned devices, such as mobile phones or PCs, people simply do not know or expect that they are interactive - an effect that has been amplified by displays having been used for static ads from their very advent. If public displays cannot communicate their interactivity, they will be hardly used and not fulfill their purpose. We believe that this issues will become even more apparent in the future as current LCD technology for public displays are likely to be replaced by technologies that more closely resemble traditional paper (e.g., e-paper [5]). As a consequence, passers-by might not even be able to notice that a surface is digital, unless the content is constantly moving.

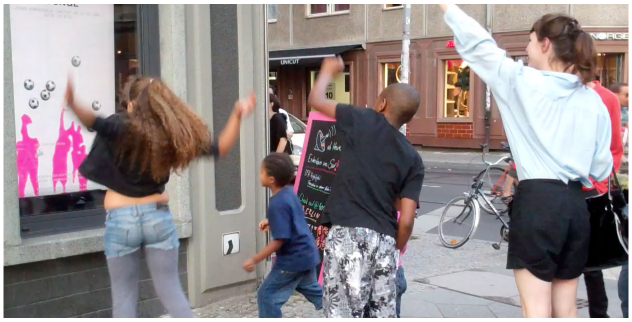

Figure 1. Two groups of users (lined up in multiple rows) having a social experience with the looking glass. Users are represented by their silhouette on the display.

To put this problem in context, passers-by of public displays need to (1) notice the display, (2) understand that it is interactive, and (3) be motivated to interact with it (not necessarily in this order). Mueller *et al.* [19] discuss the role of attention and motivation. However, relatively little is known about understanding interactivity (2), which is the focus of this paper. Previous solutions involve calls to action and attract loops [14]. A *call-to-action*, like a “Touch to start” label, can be effective. However, text or symbols are language and culture dependent and complex to understand subconsciously. *Attract loops*, such as a video of a person interacting, can create an atmosphere of an arcade game and be complex to understand in a similar manner.

In this paper we investigate how feedback to the passer-by’s incidental movements (e.g., a mirror image) can be used to communicate the interactivity of a display (see Figure 1). As humans are very efficient at recognizing human motion [2] as well as their own mirror image [18], this technique benefits from these perceptual mechanisms. After discussing psychological foundations, we report and discuss the results of a lab and a field study. In the initial lab study we were able to show that a real-time video image or silhouette of the user are equally effective for recognizing interactivity. Avatar-like and more abstract representations are less effective. We measured an average time of 1.2s people required to recognize interactivity for the mirrored video. In the subsequent field study we deployed and tested three displays in a

Submitted for review to CHI 2012

shop over the course of three weeks. Our observations show: (1) Significantly more passers-by interact when immediately showing the mirrored user image (90% more) or silhouette (47%) compared to a traditional attract sequence with call-to-action. (2) Passers-by often recognize interactivity after they already passed. Hence they have to walk back - we call this the *landing effect*. (3) Often passers-by notice interactivity because of somebody already interacting. They position themselves in a way that allows them to see both the user and the display, allowing them to understand the interactivity. If they start interacting themselves, they do so behind the person interacting hence forming multiple rows.

Our observations can be useful for designers of public displays who need to communicate display interactivity to passers-by, and more generally, for any designer of devices where users do not know in advance that the device is interactive.

RELATED WORK

Attracting *attention* with public displays and kiosks is not easy [8][11][19], and has been described as the 'first click problem' [11]. Huang *et al.* observed passer-by's attention towards (non-interactive) public displays and show that most displays receive little attention [8]. One solution is to use stimuli for attracting attention [8][19]. However, this is challenging as in public space, many other objects strive for the user's attention [8]. Another solution suggests using physical objects. For instance Ju *et al.* [11] show that a physical attract loop (animatronic hand) is twice as effective as a virtual attract loop (virtual hand). While physical objects seem to attract more attention than digital content, they are less flexible and more difficult to update with new content.

We conducted a literature review and identified 6 techniques used for *communicating interactivity* of both public displays and tabletops: (1) A *call-to-action* [14], often a simple text label such as "touch screen to begin" was used in [11], [14] and [16]. (2) An *attract sequence* which is originally described as a slideshow [14]. Some multitouch installations used constantly moving objects [22][7]. Arcade machines also use a video that either explains the interaction or shows a user performing the interaction. (3) Nearby *analog signage*, either with a simple call-to-action or a more complex manual, has been used in many deployments, e.g. [14][16][22]. (4) The *honeypot effect* [1] describes the effect of people being attracted by persons already interacting with a device. Brignull *et al.* observed this effect and divided the people around the display into the phases peripheral attention, focused attention, and interacting. Further observations of the honeypot effect are reported in [16][22][17]. (5) *Persons inviting passers-by to interact* can be either users who have already noticed the interactivity of the display and now motivate their friends [16][22], or researchers standing next to the device inviting users and explaining the interaction [9]. (6) *Prior knowledge* that a device is interactive can either be used if users pass by the same device multiple times, or if they know the device (e.g., the Microsoft Surface as in [16]).

After people noticed interactivity, *immediate usability* is important. The term has been introduced in the context of Shneiderman's CHI photo kiosk [14]. Users should be able to use the interface after observing others or using it themselves for a brief period of time (15-60s). Marshall *et al.* [16] observed that even a delay of a few seconds after touching an interactive tabletop is problematic. Users are likely to give up and think that the device is not interactive or broken.

Perceived affordances [20] are derived from Gibson's concept of affordances, which are properties of an organism's environment that have a certain relation to the body and skills of the organism. These properties make certain actions possible (afford them). While affordances exist independent of their perception, it is important how they can be perceived by users. More recently, Norman proposed the more general concept of signifiers [21]. Signifiers may be any information in the environment that indicate that a certain action is possible or appropriate. This is especially interesting in the context of public displays, as for example, smears on a screen may indicate that it is a touch screen.

Several researchers have proposed to use a *shadow or mirror image* of users of large displays to indicate and support interaction. They have been used in the context of artistic installations [13], pointing tasks on large displays [26], and interaction above a tabletop [6]. In the context of public displays, Michelis [17] deployed public displays showing a camera image of what was happening in front of the screen and augmented it with digital effects guided by motion, like clouds of numbers or growing flowers. The focus of this study was on the motivation to interact rather than noticing interactivity. Thus, no different user representations were compared and no baseline like call to action was tested. While these works explored various aspects of shadow and mirror metaphors, their application and properties to communicate interactivity of displays has not yet been explored.

Attract sequence and call to action are practical solutions to communicate interactivity. In the following we explore the representations through mirror images as an alternative.

PSYCHOLOGICAL CUES & INTERACTIVITY

When it comes to noticing interactivity, several concepts from psychology provide useful hints as to how such an interactive system should be designed. Table 1 shows that for a certain interaction it is possible to compare whether the manipulation has been intentional (or not) and whether the effect has been noticed by the user (or not). Dix *et al.* [3] discuss a continuum of intentionality between explicit and incidental interaction. Explicit interaction refers to the case where users intentionally manipulate an interactive system. Incidental interaction refers to situations where the interaction is neither intended nor the effect noticed after the fact, such as when a user enters a room and the temperature is adjusted automatically, without the user noticing. A similar concept is implicit interaction [24], which describes situations where the user interacts without being aware of interacting. As users become aware of the fact that they are interacting, implicit and incidental interaction turn into explicit interaction. To describe the situation where users manipulate a device incidentally, but become aware of the effect and thus the fact

	Noticed Effect	Unnoticed Effect
Intentional Manipulation	Explicit Interaction [3]	?
Unintentional Manipulation	**Inadvertent Interaction**	Incidental [3] / Implicit [24] Interaction

Table 1. While incidental / implicit interaction assumes that the user does not notice the effect, we can distinguish the case where the user inadvertently interacts and then sees the effect.

that the device is interactive, we use the term *inadvertent interaction*. When users perceive that a device reacts to their incidental movements, this reaction can be perceived in three ways. It can be perceived as (1) a *representation* of the user (e.g., a mirror image), (2) an effect being *caused* by the user, or (3) an *animate* being or thing reacting to the user. For all of these perceptions, powerful perceptual mechanisms exist. While the focus of this paper is on the *representation* of the user, we will also shortly review psychological foundations for perceptions of causality and animacy.

Representation: Recognizing Oneself

There are two ways how one could potentially recognize oneself in a mirror: appearance matching and kinesthetic-visual matching [18]. Appearance matching is based on a comparison of the image seen in a mirror with the knowledge of how oneself looks like. Kinesthetic-visual matching is based on the correlation between the own motion and the visual feedback in the mirror. The question whether some organism can recognize itself in a mirror has been a topic of investigation since the early work of Gallup [4]. They learned that only humans, chimpanzees, and orang-utans show this behavior. Humans can recognize themselves already in the first months of life [18]. For recognizing somebody else's reflection in a mirror, visual-visual matching can be used instead of kinesthetic-visual matching (if we can see both the person and the reflection). This is presumably easier than kinesthetic-visual matching (it is learned early in childhood).

When users control a representation of themselves on a display (e.g. mouse pointer or mirror image), they need to understand that they are in control. This is similar to the questions of psychology how humans perceive which part of the world is one's own body (ownership) and controlled by oneself (agency) [10]. From this we learn: (1) Visual feedback can override proprioceptive feedback, such that people feel agency for parts of the world which are not actually their own body. People might forget about their real surroundings when immersed in the virtual representation. (2) People assume more often that they control something that they do not actually control than vice versa (overattribution). People might assume that they control a representation even if they do not. (3) People can experience a continuum between more and less agency, depending on the correlation (amount of noise and delay) [10]. It is important to minimize noise and delay to improve the perception of agency.

Abstraction, Biological Motion, and Body Schema

Humans can not only use appearance matching, but also kinesthetic-visual matching, to recognize their mirror image. So it is possible to abstract the user representation and still have users recognize themselves. This gives the designer of a device much more artistic freedom in designing the user representation. Fortunately, humans have direct perception of the motion of humans and animals from minimal information. It was shown that a video of a dynamic array of point lights (at skeletal joints) is sufficient to see the presence of a walker [2]. For recognizing gender, the upper body joints are more relevant, and adding more points besides shoulders, elbows, wrists and hips (70% accuracy) does not improve accuracy [12]. From static images of point lights without motion however, not even the presence of a human can be seen. For this paper it is especially interesting that we can recognize ourselves and friends, and that we are more effective in recognizing ourselves (43% accuracy) than our friends (36%, 16.7% chance), despite the fact that we see our friends walking more often [2]. This is explained by the fact that both executed and perceived motion are represented in isomorphic representations (the body schema) and can easily be translated into each other.

Concluding, a system could use minimal representations similar to point light displays to represent users, but it is very important that the representation is dynamic. Upper body parts like wrists and torso might be most effective. In order to use the body schema for representation, however, the feedback needs to directly match to the movements of specific body parts (e.g., head or hand). More abstract feedback that cannot directly be matched to body parts (e.g., averages of the movements of multiple body parts) often needs more time to be recognized [27].

Perceptual Causality and Animacy

Besides for recognizing themselves, humans also have perceptual mechanisms for causality and animacy. This is impressively demonstrated by 2D movies of simple moving geometric shapes [25]. If an object 'hits' another, and this second object is 'pushed' away, humans have a strong impression that the first object caused the motion of the second. If there is more than a 50-100 ms delay between the two events, this perception starts to disappear. Similarly, objects that start from rest, change direction to avoid collision, or have directed movement towards a goal can appear to be 'alive' [25]. Perceptual causality and animacy can be used to communicate interactivity, and in these cases, known cues causing these perceptions should be used (e.g. collision). In particular, causality can be combined with mirror representations. As interacting with mirror representations alone is not very motivating, physics simulations provide motivating interaction and increase the perception of interactivity.

Relevance for this paper

In this paper we focus on the representation of the user as a cue to interactivity, because such a user representation is a very general tool to support multiple interaction techniques. From these psychological foundations, we learn the following: (1) There are efficient perceptual mechanisms that support this self-recognition (2) It is unclear how recognition of oneself degrades when the representation is abstracted. (3) It seems crucial that the correlation between the user's movement and feedback is high (low noise and delay). (4)

The feedback should be directly matchable to a certain body part, in order to use the efficient body-schema representation. (5) User representations can be combined with perceptual causality (or animacy) to strengthen the perception of interactivity and provide a more interesting application.

STUDIES

To explore how inadvertent interaction and representations of the user can be used to communicate the interactivity of public displays, we conducted a series of three user studies. We developed a series of prototypes that were successively refined based on the results of these studies. During these studies the focus was on noticing interactivity rather than attention or motivation. We simply relied on the motion of the user representation to capture attention and on a very simple game (playing with balls) to motivate users. More elaborate attention grabbing or motivating techniques would probably increase the total number and duration of interactions.

Hardware and Implementation

The system was deployed on large portrait oriented public display LCD screens of different dimensions ranging from 40” to 65”. To detect passers-by and users the Microsoft Kinect sensor was employed. The code runs on a recent linux workstation machine.

We use the 3d rendering capabilities of *OpenGL* to display the user’s mirrored image or silhouette and other virtual objects. For detecting users we rely on the *OpenNI* framework, which provides unique IDs and pixel masks to separate them from the background. The mirrored user representations are directly embedded into the scene to the lower half of the screen (see Table 2) and interacts with other virtual objects (balls). We use the *Bullet* physics library to simulate the behavior of these objects constrained to 2d plane. Since the simulation is optimized for rigid bodies, we approximate the users’ shape with small objects along their contour which are continuously tracked between frames. We record the depth image stream and user activities for later analysis.

Study Design

In the following we present a prestudy and 2 consecutive studies on noticing interactivity. We began with a *pre study* to see if and how passers-by are interacting with a public display. This was followed by a controlled *lab study* removing the attraction and motivation criteria. Hence, we could measure the time required to recognize if the test application was in an interactive or non-interactive (video playback) mode. The study further included the influence of the user representations for which we evaluated multiple levels of abstractions. Finally, in an “in the wild” *field study* we compared immediate, inadvertent interaction against an attract sequence combined with a call-to-action. We also again compare different user representations. The focus of this study however is on exploring the noticing of interactivity “in the wild”.

PRE-STUDY

Our prototype showed the silhouette of the passer-by on a 46” portrait LCD monitor. Passers-by could interact with a virtual ball using simulated physics. The display was installed for three days around lunchtime in front of a university cafeteria. Users were observed from a hidden position and interviewed on opportunity basis. Interrater reliability was satisfactory (Cohen’s Kappa=0.61) [15]. We observed 832 passers-by, of which 456 (54.8%) looked at the display, 171 (20.6%) interacted with the display, and 141 (16.9%) stopped walking to interact. People played for 2 to 182 seconds ($\mu = 26s$), and stated to mostly have left for time pressure. Interestingly, most persons interacted in groups – most single passers-by rather hurried past the display.

There are two important conclusions from this study. First, a large percentage of all passers-by interacted (in a university setting), so the design is very promising for our purpose. Second, almost no passer-by interacted alone. As our design supported only single-use, this posed problems as mostly groups of 2-5 users tried to interact simultaneously. Also, almost all passers-by stopped before interacting, while we expected more interacting while passing by.

LAB STUDY

The objective of this study was to determine the impact of the abstraction of the representation of the user on how quickly users can notice that a display is interactive. We compared the user representations *mirror, silhouette, avatar,* and *abstract*. In this study, we only focused on noticing interactivity. We asked participants to pay attention to the display and decide whether the display reacted to their movements or not. No additional virtual objects, that would potentially have biased the motivation of the participants, were shown on the screen. This lab study setup provided a baseline of how quickly users can decide whether a display is interactive under optimal conditions using the different representations. The lab design provided a high degree of control, while at the same time providing a lower degree of ecological validity. To counterbalance, the study was followed by a field study, which offers high ecological validity but less control.

Conditions

The conditions were (a) Mirror image: an interactive colored image of the user (on a black background), (b) Silhouette: a white filled silhouette of the user, (c) Avatar: a 2d avatar including head, torso, and hands, and (d) Abstract: just the head of the user, with abstract eyes and mouth.

All of these conditions can be directly matched to body parts by the user (see section Psychological Cues & Interactivity). For the expected interaction distance at the shop windows the camera could not capture both feet and head of the user. Based on the studies of point-light displays that show that upper body parts are most relevant, we decided to position the camera so that these parts were visible. Based on the same studies, we expect the gain in speed and accuracy from adding feet to the avatar to be low. Related work on stimulus-response compatibility [27] indicates that stimuli that can be directly matched to body parts are more effective than those which cannot. Therefore, we decided for the abstract condition to directly represent the head of the user (instead of, e.g., an average of multiple body parts). All four of these interactive conditions were also presented as non-interactive conditions. In this case, a video of another user interacting with the display was started as soon as the user stepped in front of the display. These non-interactive conditions should simulate situations where either just a random video was shown on a display, or a different user (e.g,. standing behind the participant) would interact with the display.

Task and Stimulus

Users were asked to walk past the display back and forth following a line on the ground placed at a constant 2m distance. On the display, one of the 4x2 different conditions was shown. Users carried a device (Logitech Presenter) and were asked to click on the left button when they believed the display to react to their movements, and the right button when they believed the display not to react to their movements. Users were asked to be as fast and accurate as possible. Time was measured from the moment when they entered the FOV of the camera (and thus appeared on the screen in the interactive conditions) until they pressed a button.

Apparatus and Design

A 82" portrait LCD display was used to present the content. The representation of the user was created using a Microsoft Kinect camera and software using OpenNI, NITE, and Processing. A within subjects design was used with n=16 participants recruited from a pool of non-computer scientists. Variables measured were time and accuracy.These 4x2 conditions were repeated in 10 blocks. The order was counterbalanced using a latin square within the participants, and randomized between the participants.

Results

The *selection time* was measured as the time from when the stimulus appeared (as the user entered the camera's field of view) to the time when the user made a choice. An ANOVA revealed a significant effect for *representation* on selection time ($F_{3,45} = 80.76, p < .0001$). It also revealed a representation * interactivity interaction effect on selection time ($F_{3,45} = 6.75, p < .0001$). A post-hoc Tukey test showed that Mirror (1.2s) and Silhouette (1.6s) are significantly faster than Avatar (2.8s) and Abstract (2.8s) in the interactive condition. In the non-interactive condition, Mirror (1.2s) is significantly faster than Silhouette (1.7s) and Avatar (2.1s) which is significantly faster than Abstract (2.8s).

An ANOVA also revealed a significant effect for representation on *accuracy* ($F_{3,45} = 43.09, p < .0001$). It also revealed a representation * interactivity interaction effect on accuracy ($F_{3,45} = 5.84, p < .0001$). A post-hoc Tukey test shows that Mirror (100%) and Silhouette (97.5%) are significantly more accurate than Abstract (84.3%) and Avatar (81.2%) in the interactive condition. In the non-interactive condition, Mirror (98.8%) and Silhouette (97.5%) are significantly more accurate than Avatar (86.3%) which is significantly more accurate than Abstract (73.1%). Finally, the ANOVA revealed a significant effect for *block id* on accuracy ($F_{9,135} = 5.84, p < .0001$). A post-hoc Tukey test shows that users are less accurate in the first block (74.2%) than in the other blocks (mean:91.6%).

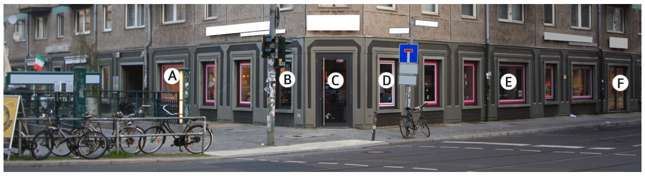

Figure 2. Study location: Displays were finally installed in three shop windows (B, E, F)

Discussion

From this experiment we learn that (1) the Mirror and Silhouette representation are similarly efficient, but both more efficient than the Avatar and Abstract representation, and (2) it takes considerable time to distinguish the interactive and the non-interactive conditions even in an optimal environment (1.2s vs. 1.6s). The fact that the Silhouette is efficient is good, because it provides much more artistic freedom for the designer of a display. While the lab study provided control, ecological validity was low. Therefore, we decided to compare the two most promising representations, Mirror and Silhouette, to a combination of two common traditional techniques, call to action and attract loop, and a purely causal technique in a field study.

FIELD STUDY

The objective of this study was to explore how users would notice interactivity and interact with public displays using different user representations "in the wild". We compared the two most effective user representations, image and silhouette, to the most common strategy in industry, call-to-action combined with an attract loop, and a merely causal condition without user representation. This comparison was regarding their ability to attract users to interact with the display as well as their general effect on the social situation in an urban place. A field study was chosen in order to maximize ecological validity, sacrificing the control of the lab.

Deployment

Three displays were deployed for three weeks in shop windows of a store in the city center of Anonymous (see Figure 2). Windows on one side of the store (D, E, F) were close to a well frequented sidewalk, windows on the other side (A, B) were near a subway entrance. To decide in which windows to install the displays we observed 200 passers-by of the street-facing side of the store (C, D, E, F) during afternoon until night. The observations showed that there are large differences in how many passers-by look into each shop window. The percentages are: Main door C (6%), small window D (12%), small window (13%), small bright window (19%), large window E (29%), small window (16%), large window (29%), second door F (large and bright, 33%). For people walking from right to left, for whom the second door was the first window they saw, even 66% looked into the window. It seems that the large and bright windows attract more attention, especially if surrounding windows differ. Also for people walking from right to left, we noticed a large percentage (17%) looking straight away from the last window. Apparently, they looked down a road at the crossing. For the deployment we used three LCD monitors in portrait format

(65”, 46”, 46”). Cameras (Microsoft Kinect) were installed below the monitors. For the first week of deployment we moved the displays between the windows A, B, C, D, E, F (see Figure 2). While window B had the advantage that people could play relatively undisturbed from passers-by, windows E and F had a larger number of passers-by and attracted most views. Therefore we decided to install the 65” display in window B, and two 46” displays in windows E and F. For the background image we initially tried different artistic contents, but could not observe a large influence of our contents on behavior. The final content was an advertisement for the store and was created by a professional advertising agency.

Conditions

In our study we tested two variables: user representation (image vs. silhouette vs. no-representation) and interactivity cue (inadvertent interaction vs. attract sequence with call to action). Regarding the application, we opted for a very simple ball game. Ten balls were displayed on the screen, and users could play with them (kick them) using the contour of their representation. The whole game took place in the 2d plane of the user representation. In the image condition, the user’s image from the color camera was extracted from the background and shown on the display. In the silhouette condition, the silhouette of the user was shown on the display, and in the no-representation condition, just the balls were rendered, but no user representation was shown (but interaction was as in the other conditions). In the inadvertent interaction condition, when nobody was in front of the screen, just the background image and balls were shown. The interaction started as soon as users entered the FOV of the camera. In the attract sequence with call to action condition, a video of a person demonstrating the interaction was shown together with a label “Step Close to Play” (see Figure 2. The video showed a person in the corresponding visualization (image, silhouette, and no visualization) stepping close to the camera and then playing with the balls. When the user entered the FOV of the camera with a closer distance (1m), the screen was switched to interaction, the user was represented in the corresponding visualization and could play with the balls. Conditions were counterbalanced and automatically switched every 30 minutes. This was done to minimize the influence of time of day on the results.

Data Analysis

We collected both qualitative and quantitative data. Qualitative data was gathered from observations, semi-structured interviews, and manual video recording. As quantitative data, complete interaction logs (from NITE person tracking) and videos from the depth camera were kept from each display over three weeks. For anonymity reasons we did not record the camera image, but only the (anonymous) depth image.

Qualitative data collection was conducted daily during three weeks. As displays worked best and most interaction occurred in the late afternoon and evening, at least two researchers were present during these times. Additional observations were conducted as needed. Observations were conducted from inconspicuous positions like the other side of the street or near the subway entrance, where it was common to see waiting people. During the observations, video was recorded using video cameras that looked similarly to mobile phones (FLIP HD). Further, field notes were kept. Every day interesting findings were presented and discussed in a meeting of the entire research team. Eventually, the team agreed on a specific focus for following observations.

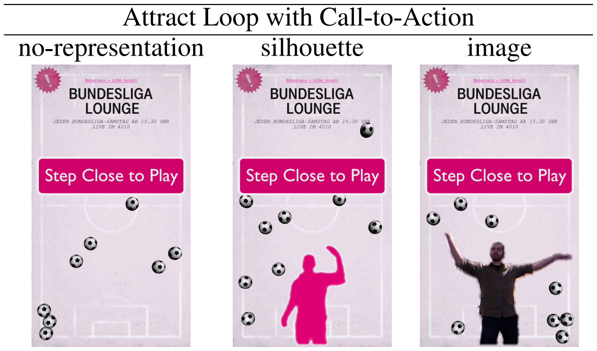

Table 2. Representations. We tested three user representations: no-representation, silhouette, and image. All three representations were tested in an “attract loop with call to action” as well as in a “inadvertent interaction” version. In this figure, the corresponding attract loops (a video of somebody stepping close to the display and starting to interact) are shown. In the inadvertent interaction condition, the person in front of the display was shown in the same representation, just without the call to action (“Step Close to Play”).

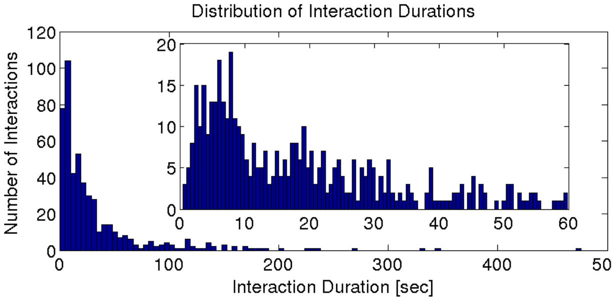

Figure 3. Interaction Durations: In order to investigate how well the different conditions communicate interactivity, we needed a large number of situations where nobody was currently interacting with the screen. Hence we intentionally designed the interaction not to be especially motivating for extended play. The mean duration of interactions was 31s, but many interactions only lasted for a few seconds. Surprisingly, some users seemed to be motivated to play for many minutes.

From the depth videos we recorded roughly 1500 hours of videos. We selected 11 consecutive days for manual coding. We implemented an analysis software that automatically searched the log files for scenes, in which a user was detected for more than 4 seconds. In accordance to [16] and [22], interactions which followed each other within less than 20s were merged to single sessions. All sessions were then manually reviewed and annotated. We observed 363 interactions. Inter-rater reliability was substantial (Cohen’s Kappa=.75) [15]. During the analysis, we grouped our findings in four categories: *Image, Silhouette, and Call-to-Action*, *the landing effect*, *dynamics between groups*, and *dynamics within groups*.

	No-representation	Silhouette	Image
Call-to-action	67	59	79
Inadvertent interaction	60	87	150

Table 3. Total number of interactions in the different conditions during 11 days of field study. Inadvertent interaction attracts significantly more interactions than call-to-action. Further, Image works significantly better than Silhouette and No-representation.

Findings

Image, Silhouette, and Call-to-action

The total number of interactions during the 11 coded days is shown in Figure 3. We compared the number of interactions per day. ANOVA reveals a significant effect for interactivity cue (call-to-action vs. inadvertent interaction) ($F_{1,11} = 12.6p < .001$). A post-hoc Tukey test shows that passers-by interact more with the inadvertent interaction condition than with the call-for-action. ANOVA also reveals a significant effect for user representation ($F_{2,22} = 13.1$). A post-hoc tukey test shows that Image is more efficient than Silhouette and No representation. Finally, ANOVA also reveals a user representation*interactivity cue interaction ($F_{2,22} = 6.8, p < 0.005$). As expected, there are no significant differences between user representation for call-to-action. User representations differ only in the inadvertent interaction condition. Many interactions with the display only lasted for a few seconds (see Figure 3). The interviews revealed different preferences for the user representations. The shop owner preferred the silhouette as people were covered in company colors. For users there was no clear preference, and many said that they liked the representation they discovered first. Users who preferred the image representation described it as more "authentic", more "fun", and they liked to see themselves and their friends. Users who preferred the silhouette representation described it as more "anonymous" and said that they liked it when bystanders could not see their image. Some also said that they did not like to see themselves, so they preferred the silhouette representation. In the image representation, also some users mentioned that they do not like to be observed by a camera, which they did not say for the silhouette representation. From our observations we found, that in the call-to-action conditions, people spent several seconds in front of the display before following the instructions ("Step Close to Play") (compare Figure 4). In this vignette, two girls observe the display for some time, before one steps close and activates the interaction in the image condition. They are surprised by seeing themselves and walk away. A few meters further, they notice a second display running the inadvertent interaction silhouette condition, where they start to play. When interviewed how they noticed interactivity, most people said that they saw themselves on the display. Some also said that they saw themselves and a friend / partner at the same time. Only very few stated to have seen the representation of another person walking in front of them.

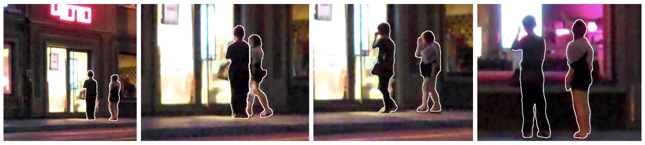

Figure 4. In the call-to-action condition people sometimes spent considerable time in front of the display (1) before stepping closer (2). In this case, the two women are surprised by seeing themselves and walk away (3). On the next window, they encounter inadvertent interaction in the silhouette condition and start playing (4).

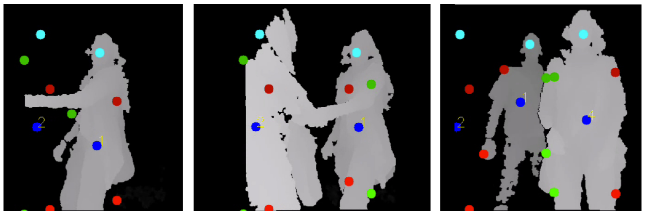

Figure 6. Landing effect for a couple: As the couple passes by, the woman notices the screen and stops. As her partner walkes on, she drags him back to the screen. Both start interacting. (The scene is from the depth video logs that were annotated)

When a crowd had already gathered around the display, it was sometimes very difficult to distinguish which effect was caused by whom. This was especially true for the silhouette and obviously the no representation conditions. In these cases we observed people copying the movements of other users and seemingly interacting with the screen, although they were not represented on the screen. Sometimes they were not even standing in the field of view of the camera. This can be an example of *overattribution* (compare section on psychological cues), where people assume they are causing some effects although they are not.

Over time, knowledge about the presence and interactivity had built up among people who pass the location regularly. In the third week of deployment, a number of people who interacted said that they had seen somebody else interacting, e.g., "a few weeks ago" or "earlier that day", but had not tried interaction themselves. There were also a few regular players. For example, we noticed from the logs that between 7-8am, there was considerable activity in front of the displays. Observations revealed that a number of children played regularly with the displays on their way to school. We observed them waiting expectantly at the traffic light, then crossing the street directly to the display to play with it. Such interaction is obviously different from situations where people encounter the displays for the first time.

Design Recommendations: Inadvertent interaction outperforms the attract loop with call-to-action in attracting interactions. The image representation also outperforms the silhouette and interaction without user representation. In contrast to the lab study, the image representation works significantly better than the silhouette. From this we learn that image representations are a powerful cue to communicate interactivity, although silhouettes may have some benefits like more artistic freedom in designing the content and provide more anonymity. As most people recognize themselves on the display rather than someone else, displays should be positioned so that people can see themselves well when passing by. Over time, as knowledge about the interactivite device builds up, these interactivity cues become less important.

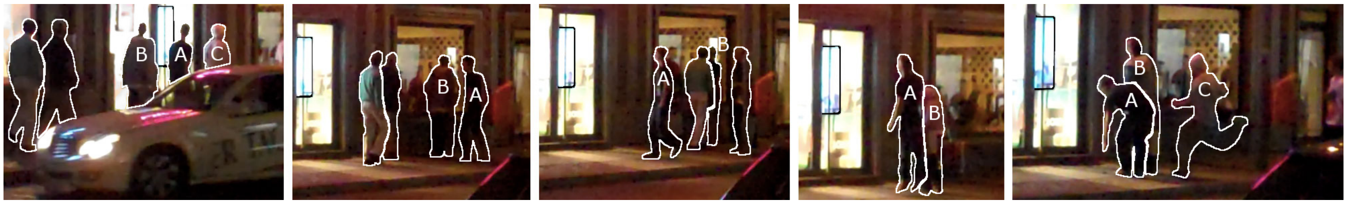

Figure 5. Landing effect for a group: A group of people passes the display (1). Only at the next shop windows person A stops (2), turns around, and walks back to the display (3). As he starts interacting (4) more and more people from the group join. (5)

The Landing Effect

One striking observation regarding the moment when people start to interact was that often, people stop late and have to walk back (see Figure 6 for this effect with a couple, and Figure 5 for this effect in a group). In Figure 5, a group of young men is passing the display. The seventh person in the group looks at the display but keeps on walking with the group. Some meters further the person suddenly turns around and walks back, followed by a second person. They then start to interact, and are soon joined by other group members. In this paper we refer to these cases as the *landing effect*.

Regarding the number of landing effects, interestingly ANOVA reveals a significant effect for interactivity cue ($F_{1,11=23.1}, p < 0.0001$). A post-hoc Tukey test shows that more landing effects are observed in inadvertent interaction (18.5% of all interactions) than in call-to-action (8%). There was no significant effect for visualization. We observed this behavior only for people passing by the displays (not waiting), when nobody was yet interacting with the displays, and who apparently did not know before that the displays were interactive (e.g., because they already interacted with them). The landing effect often led to conflicts when one person in a group noticed the interactivity. If the first persons in a group suddenly stopped and turned around, the following people would sometimes bump into them. More often, the whole group stopped rather than walking on. When a following person in a group however noticed interactivity, the first would usually walk on for some time before they noticed that somebody stopped and stop themselves. This situation created a tension in groups as to whether people who already continued walked back or whether the person interacting would abandon the display and join the group. In some cases the group simply walked on after some waiting, causing the interacting person to continue playing only for a short moment and then hurry (sometimes even run) to rejoin the group. Interviews revealed more details about this behavior. One man who had walked back (image condition) answered that he had seen from the corner of his eye two persons on the screen walking into the same direction. He was curious and walked back, accompanied by his wife. When he saw himself on the display, he understood that it was interactive and explained it to his wife. They both started to play with it. For another couple, the man stated that he saw something moving from the corner of the eye and walked back. His wife stopped, but did not follow him. He noticed that the display was interactive upon seeing himself, but only played very shortly before joining his wife. It is quite possible that users did not interact, because they only noticed interactivity after they had already passed the displays and did not want to walk back.

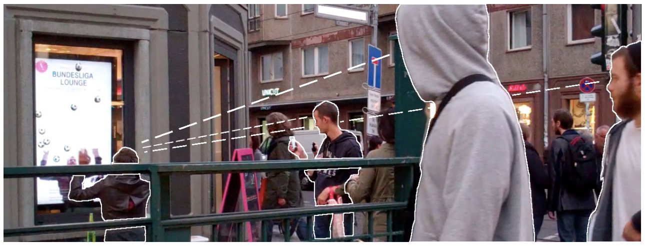

Figure 7. The Honeypot Effect: As people notice a person making uncommon gestures, they position themselves in a way allowing both the screen as well as the interacting person to be seen. They also often position themselves so that they are not represented on the screen.

Because we installed multiple displays along the same trajectory, passers-by had the option to notice interactivity on one screen, but then interact with another one. When they noticed the second screen, they already expected that it was also interactive and stopped earlier. One man for example said to have noticed the balls jumping away on the first screen, but then did not walk back. When he noticed the second screen, he decided to stop his friend. They saw their representations and played for a short moment. Often, after playing with one screen, people also searched the other windows of the shop for further screens. If they saw further screens, they often also played with those (see Figure 4).

Design Recommendations: The landing effect is in line with our observation from the lab, that people need approx. 1.2s (image) and 1.6s (silhouette) to recognize interactivity. They also need to notice the display first and be motivated to interact. With an average walking speed of 1.4 m/s, by the time passers-by have decided to interact, they have already passed the display. This effect is so strong that it should be designed for in any public display installation. Displays should be placed so that, when people decide to interact, they are still in front of the display and do not have to walk back. Optimally, when users stop friends walking in front of them, also the friends should still be able to interact with the display without walking back. This could be achieved by designing very wide displays (several meters), or more practically, a series of displays along the same trajectory. Another solution would be to place displays so that users walk directly towards them, but this is possible only for very few shop windows.

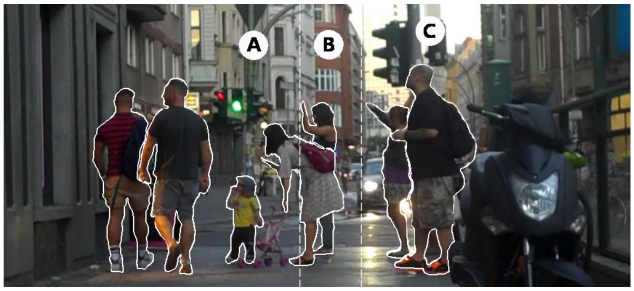

Figure 8. Multiple Rows: The girl from group A noticed interactivity first. Woman B positioned herself behind them to see what happens and also started interacting. Later, a couple C stopped behind them and started interacting in a third row.

Dynamics Between Groups

We observed many situations in which different groups started to interact. The first group (or person) usually causes what has been previously termed the "honeypot effect". We found that people passing by firstly observed somebody making unconventional movements while looking into a shop window (the manipulation [23]). They subsequently positioned themselves in a way that allowed them to see and understand the reason for these movements – usually in a location that allowed both the persons interacting as well as the display to be seen (see Figure 7). In this figure, a man interacting with the display with expressive gestures attracts considerable attention. The crowd stopping and staring at him and the display partially blocks the way for other passers-by. Newcomers seem to be first attracted by the crowd, then follow their gaze, then see the man interacting, follow his gaze, repositioning themselves so they can see both the man and the display. They also seem to prefer to stand a little bit to the side, so that they are not represented on the screen. The audience is mostly positioned behind the user. We observed this pattern regularly. When people in the audience decided to join the interaction, they accordingly regularly did so *behind* the ones already interacting, not next to them (see Figure 8). In this figure, the little girl in the front noticed the interactivity first, followed by her mother, who then stopped to explore the display together with the daughter (the father did not walk back and is standing behind the camera). The young woman behind them was attracted by their interaction and eventually also started interacting behind them. This again attracted the couple behind them, of which the girl finally also started interacting in a third row. In some cases, such multiple rows where then again observed by people in the subway entrance. In the few cases where other people started to interact in the same row as people already interacting, we were able to observe social interaction between the users, which we did not observe for different groups interacting behind each other.

People interacting with the screens were usually standing in the way of others. The resulting conflicts were solved in different ways. For the screen installed near the subway entrance, passers-by usually tried to pass behind the ones already interacting, not disturbing them. When multiple rows of people interacted, this was not possible however, and they passed in front of them (Figure 8). When a large group passed by, we sometimes observed that the person interacting abandoned the display. This again sometimes let someone from the coming group take the place and start to play. We also saw some occasions, where users deliberately moved between the display and the person interacting and interacted for a very short moment.

Design Recommendations: The honeypot effect is a very powerful cue to attract attention and communicate interactivity. Displays which manage to attract many people interacting will be able to attract more and more people. The honeypot effect even works after multiple days, as people who have seen somebody interacting previously may also try the interaction in the future (see subsection on image, silhouette and call-to-action). To achieve this, displays should be designed to have someone visibly interacting with them as often as possible. This can be achieved by improving motivation and persuading people to play longer. Because the audience repositions themselves such that they can see both the user and the display, the environment needs to be designed to support this. In our case, both the subway entrance and the narrow sidewalk limited the possible size of the audience. In order to support more audience, displays should be visible from a wide angle, or considerable space should be available directly in front of the displays. This is also necessary as different groups start to interact behind each other. This interaction behind each other should also be supported, e.g., by increasing the maximum interaction distance beyond the distance from where single groups normally interact.

Dynamics Within Groups

We discovered that the vast majority of interactions were performed by people being in a group. The only cases of single people interacting we observed personally were children before or after school, men after waiting for considerable time near the subway entrance, a man in rags, and a man filming himself while playing. One man for example waited for many minutes directly in front of one screen, while incidentally interacting with it through his movements. After some time, he was approached by an apparent stranger, who showed him the display and the fact that he was interacting. The man seemed surprised, and continued to play a little bit with the display. While a considerable number of single people passes by the store, they usually walk faster and look more straight ahead and downwards. When we interviewed some of them, only very few had noticed the screens at all, and nobody had noticed that the screens were interactive.

Between 1 and 5 people interacted simultaneously ($\mu = 1.5$). Often the first person in a group noticed the display first, while this was not always the case.

We discovered that people strongly engaged with the game and apparently identified more with their representation on the screen than the possible influence of their movements on people around them (see section on psychological cues). This sometimes led to situations where people were not aware anymore of their neighbors (people belonging to one group usually line up next to each other), even though they were able to see their representation on the screen. This focus on the virtual space led in some situations to that people accidentally hit ot bumped into each other. Another observation was that people usually started interaction with very subtle movements and continously increased the expressiveness of their movements. This process sometimes took just a few seconds and sometimes extended over many minutes. The subtle movements at the beginning were sometimes just slight movements of the head or the food. Later, people proceeded to extensive gesturing with both arms, jumping, and even acrobatic movements like high kicks with the legs.

Design Recommendations: The most important observation from this section is that very few persons who are alone interact. This observation is supported by the results of the prestudy. Therefore it is important to understand how groups notice interactivity, and public displays should always be designed to support groups. Even if just one person is interacting, the display must provide some value for the other group members. When users strongly engage with their representation on the screen, they may forget about their real surroundings. According to our observations, more slowly moving objects make users conduct also slower movements, which increases safety.

CONCLUSION

From this paper we learn that: (1) Using the mirror image of users such that passers by inadvertently interact with public displays is an effective way of communicating interactivity. Mirror images are more effective than silhouettes and avatars, and more effective than a traditional attract loop with a call-to-action. (2) Noticing interactivity needs some time, which leads to the *landing effect*. When passers-by decide to interact with public displays, they have often already passed them, so they have to walk back. This can be mediated e.g., by installing multiple displays in a row. (3) Users from a different group often start to interact *behind* the ones already interacting, forming multiple rows. Because also, the vast majority of interacting persons are in groups, public displays should support multiple users, in particular when interacting behind each other. We hope that mirror representations for inadvertent interaction will also be applied to other devices beyond public displays, e.g., tables or floors. Finally, we believe that public displays that effectively communicate their interactivity have the potential to make urban spaces all over the world more fun and engaging to be in.

REFERENCES

1. Brignull, H., and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *Proc. of INTERACT '03* (2003), 17–24.
2. Cutting, J. E., and Kozlowski, L. T. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society 9*, 5 (1977), 353–356.
3. Dix, A., Finlay, J. E., Abowd, G. D., and Beale, R. *Human-Computer Interaction (3rd Edition)*. Prentice-Hall, Inc., 2003.
4. Gallup JR, G. G. Chimpanzees: Self-recognition. *Science 167*, 3914 (1970), 86–87.
5. Heikenfeld, J., Drzaic, P., Yeo, J.-S., and Koch, T. A critical review of the present and future prospects for electronic paper. *Journal of the Society for Information Display 19*, 2 (2011), 129–156.
6. Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., and Butz, A. Interactions in the air: adding further depth to interactive tabletops. In *Proc. of UIST '09*, A. D. Wilson and F. Guimbretière, Eds., ACM (2009), 139–148.
7. Hinrichs, U., and Carpendale, S. Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In *Proc. of CHI '11*, ACM (New York, NY, USA, 2011), 3023–3032.
8. Huang, E., Koster, A., and Borchers, J. Overcoming assumptions and uncovering practices: When does the public really look at public displays? In *Proc. of Pervasive '08*. Springer, 2008, 228–243.
9. Jacucci, G., Morrison, A., Richard, G. T., Kleimola, J., Peltonen, P., Parisi, L., and Laitinen, T. Worlds of information: designing for engagement at a public multi-touch display. In *Proc. of CHI '10*, ACM (New York, NY, USA, 2010), 2267–2276.
10. Jeannerod, M. The mechanism of self-recognition in humans. *Behavioural Brain Research 142*, 1-2 (2003), 1–15.
11. Ju, W., and Sirkin, D. Animate objects: How physical motion encourages public interaction. In *PERSUASIVE'10* (2010), 40–51.
12. Kozlowski, L. T., and Cutting, J. E. Recognizing the sex of a walker from a dynamic point-light display. *Perception Psychophysics 21*, 6 (1977), 575–580.
13. Krueger, M. W. *Artificial reality II*. Addison-Wesley, 1991.
14. Kules, B., Kang, H., Plaisant, C., Rose, A., and Shneiderman, B. Immediate usability: a case study of public access design for a community photo library. *Interacting with Computers 16*, 6 (2004), 1171 – 1193.
15. Landis, J. R., and Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics 33*, 1 (1977), 159–174.
16. Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., and Davies, M. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *Proc. of CHI '11* (2011), 3033–3042.
17. Michelis, D., and Mueller, J. The audience funnel. *International Journal of Human-Computer Interaction* (2010).
18. Mitchell, R. W. Mental models of mirror-self-recognition: Two theories. *New Ideas in Psychology 11*, 3 (1993), 295–325.
19. Müller, J., Alt, F., Michelis, D., and Schmidt, A. Requirements and design space for interactive public displays. In *Proc. of ACM Multimedia '10*, ACM (New York, NY, USA, 2010), 1285–1294.
20. Norman, D. A. Affordance, conventions, and design. *interactions 6* (May 1999), 38–43.
21. Norman, D. A. The way i see it: Signifiers, not affordances. *interactions 15* (November 2008), 18–19.
22. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., and Saarikko, P. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In *Proc. of CHI '08*, ACM (New York, NY, USA, 2008), 1285–1294.
23. Reeves, S., Benford, S., O'Malley, C., and Fraser, M. Designing the spectator experience. In *Proc. of CHI '05*, ACM (New York, NY, USA, 2005), 741–750.
24. Schmidt, A. Implicit human computer interaction through context. *Personal and Ubiquitous Computing 4*, 2/3 (2000), 191–199.
25. Scholl, B. J., and Tremoulet, P. D. Perceptual causality and animacy. *Trends in Cognitive Sciences 4*, 8 (2000), 299 – 309.
26. Shoemaker, G., Tang, A., and Booth, K. S. Shadow reaching: a new perspective on interaction for large displays. In *Proc. of UIST '07*, ACM (New York, NY, USA, 2007), 53–56.
27. Wilson, M. Perceiving imitatible stimuli: Consequences of isomorphism between input and output. *Psychological Bulletin 127*, 4 (2001), 543–553.