Move Your Body: Engaging Museum Visitors with Human-Data Interaction

Milka Trajkova[1*]**, A'aeshah Alhakamy**[1,2*]**, Francesco Cafaro**[1]**,
Rashmi Mallappa**[1]**, Sreekanth R. Kankara**[1]

[1]Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, USA
[2]University of Tabuk, Tabuk, Saudi Arabia
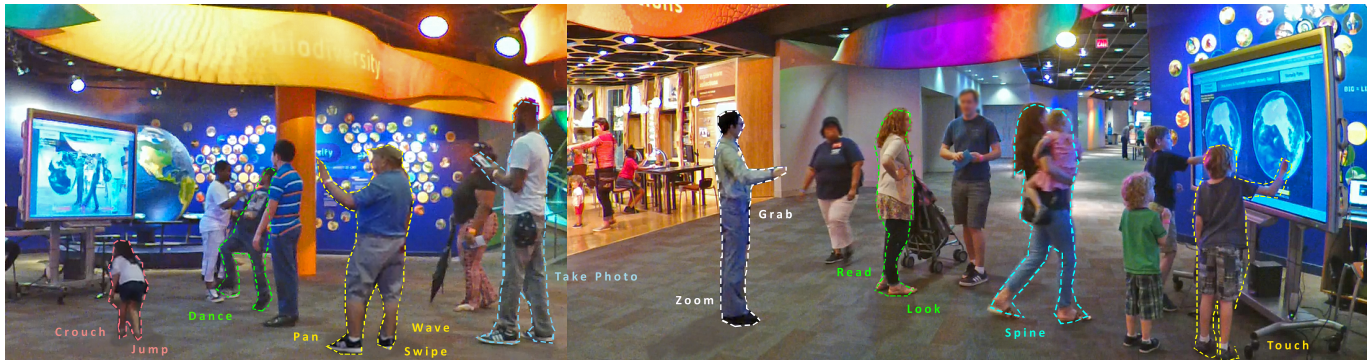milktraj@iu.edu, aalhakam@iu.edu, fcafaro@iu.edu, rmallap@iu.edu, skankara@iupui.edu

Figure 1. Museum visitors interact with our prototype Human-Data Interaction system using gestures and body movements.

ABSTRACT

Museums have embraced embodied interaction: its novelty generates buzz and excitement among their patrons, and it has enormous educational potential. Human-Data Interaction (HDI) is a class of embodied interactions that enables people to explore large sets of data using interactive visualizations that users control with gestures and body movements. In museums, however, HDI installations have no utility if visitors do not engage with them. In this paper, we present a quasi-experimental study that investigates how different ways of representing the user ("mode type") next-to a data visualization alters the way in which people engage with a HDI system. We consider four mode types: avatar, skeleton, camera overlay, and control. Our findings indicate that the mode type impacts the number of visitors that interact with the installation, the gestures that people do, and the amount of time that visitors spend observing the data on display and interacting with the system.

Author Keywords

Embodied Interaction; Human-Data Interaction; Public Displays; Informal Learning; Museums

---

*Both authors contributed equally to this research.

INTRODUCTION

The availability of commercial motion tracking devices (e.g., Microsoft Kinect) was initially met with great enthusiasm by early-adopters and by the research community, because of the promise to bring embodied interaction [15] [22] to the masses [17]. Such devices opened scenarios in which people no longer controlled interactive systems and installations using traditional input devices (such as keyboard and mouse), but directly with their hand gestures and body movements.

In particular, museums have embraced embodied interaction: its novelty generates buzz and excitement among their patrons and it has enormous educational potential. For example, children better understand and remember physics concepts when they are asked to "embody" a meteor in an interactive simulation (rather than when they are introduced to similar concepts using a more traditional desktop simulation) [34], and are able to more easily grasp the elements of rhythm, melody, and harmony when mimicking characters in music education [57].

This paper focuses on the design of a specific class of embodied interaction: Human-Data Interaction (HDI) [16] [9]. Today, we live in a world that has an unprecedented collection of and access to data. Therefore, there is an increasing need to facilitate understanding of data. Human-Data Interaction (HDI) allows people to directly interact with data (rather than just passively consume them) in order to gain knowledge from them [23]. Designing HDI installations that are actually engaging for museum visitors, however, is still a challenge. First, HDI displays that are installed in museums often compete with surrounding stimuli (other exhibits, people, signs, etc.); this limits the number of visitors who notice them (a phenomenon known as "display blindness" [12]) or understand

that the display is interactive ("interaction blindness" [26]). Second, museum visitors cannot consult user manuals before interacting with an exhibit, and often leave thinking that the system is broken if the installation does not quickly respond to their gestures and body movements [38] (an issue known as "affordance blindness" [14] or "discoverability problem" [10]). Third, museum visitors usually engage for two minutes or less with a successful exhibit [53], but exploring large datasets typically requires a longer time. Fourth, visitors may interact with gestures and body movements just for fun [38], without taking a break from the interaction to observe the visualization and reflect on the data [50].

In order to tackle these design challenges, we conducted a quasi-experimental study with 731 museum visitors (see Figure 1) that showed how different ways of representing the user next-to a data visualization impacts the way an HDI system attracts passer-bys, the gestures and body movements that users do, people's engagement with the interaction, and the amount of time visitors spend looking at the visualization.

BACKGROUND

Embodied Interaction
According to Dourish [15], we construct meaning through our "embodied" (i.e. physical) interaction with the world. More recently, the term "embodied interaction" has also been used to denote interactive installations that are controlled by hand gestures and body movements (e.g., [22], [11]).

Human-Data Interaction (HDI)
To the best of our knowledge, the wording Human-Data Interaction was used for the first time in data visualization by Zhu et al. in 2008 to refer to the analysis of multi-variate data [60]. HDI has also been used by Mortier et al. [41] to investigate how personal data are collected and shared, the ethical implications surrounding the collection and use of data, and issues related with privacy and consent. Mortier's work has been referenced in HCI literature, e.g. [24]. More recently, the term HDI has been used in visualization literature to investigate how different data visualizations alter how people make sense of data (e.g., [8]).

As noticed by Victorelli et al. in their literature review of Human-Data Interaction [56], the term HDI has been used to refer to a diverse range of research topics spanning from computer graphics to information science. Thus, we need to clarify that the work in this paper is positioned within the sub-field of embodied interaction research that was dubbed as "Human-Data Interaction" (HDI) by Elmqvist [16] and Cafaro [9]. These works investigate how users interact with and make sense of large datasets (that are visualized on gesture-controlled large displays or tabletops). Although digital data is the common thread, this line of work is profoundly different in aims and methodologies from the one that stems from Mortier's [41]: the focus is not on how to handle personal data or user-generated data, but on how embodied interaction can "facilitate the users' exploration of large datasets" [9].

RELATED WORK

Enticing People to Interact with Large Displays
Enticing people to interact with public displays can be challenging [7]. For example, a seminal work by Brignull et al. [7] analyze usage patters around the Opinionizer, a shared display where people could add their views and opinions using keyboard during social gatherings, and claim that people failed to interact in part because of the fear of social embarrassment. Interesting, the more people congregated around the screen, the more others tended to stop by ("honey-pot" effect [7]).

Alternatively, Claes et al. [13] suggest to include tangible elements to improve users' engagement with displays; in particular, this work explored the use of a non-interactive display paired with plastic plates that people could use to change the data visualized on the screen.

Display Blindness, Interaction Blindness, and Affordance Blindness
Three reasons why people do not interact with a display may be that they even do not notice the display at all ("display blindness" [12] [37]), or they do not know that they can interact with it ("interaction blindness") [51] [26], or they do not understand how to interact with the display ("affordance blindness" [14]).

Interaction blindness refers to the fact that people may not interact with a screen just because they "simply do not know that they can" [44]. To communicate interactivity, Ojala et al. [44] suggest placing a keyboard and mouse in front of the screen (because people are familiar with such interactive tools) -a solution, however, that does not fit the design of full-body interactions. Proxemic Interaction [5] [20] -i.e., interactive devices that adapt to the user's fine-grain position -has been used as a strategy to entice people to interact. For example, Hello.Wall [47] displays different light patterns depending on the user's proximity to the display. Similarly, Proxemic Peddler [58] cycles commercial advertisements depending on passerby's proximity to a large screen. Alternatively, Houben and Weichel [25] propose to place a "curiosity object" (a small wooden casket with a switch) near the interactive display as a gateway to the interaction with the display -however, people need to first notice the tangible object, and some people may only interacted with the curiosity object, not with the screen. An alternative approach is to add prompts that invite people to interact. For example, Kukka et al. [32] compared the impact of static and dynamic icons and textual prompts (e.g., "touch me") in making people interact with the display, and found that text is more effective. The literature on the topic is, however, conflicting: other studies suggest that such "call-to-action" techniques are not that effective overall in communicating the interactivity of displays [42].

In the context of museum exhibits, Cafaro et al. [10] recommend making data oscillate when a passerby is able to control them -but this only works when georeferenced data are represented as bubbles on a map. Gentile et al. [19] propose to place an avatar in the middle of a public display interface (which included ribbons with weather, news, and university information) as a strategy to communicate to passersby that the display is

interactive. In such interface, the avatar hands work as mouse cursor and allow the user to select interface elements -but this "mouse cursor" strategy may be difficult to extend to other interaction types. Similarly, Khamis et al. recommend including a non-abstract representation of the user (e.g., skeleton, avatar) [29]. Muller et al. [42] discuss a study that investigates how passers-by notice the interactivity of public displays, when they are represented as mirror image, skeleton, avatar, or in abstract in their interaction with a ball game. The study was conducted in-lab and on the field (a shop window). The lab study revealed that the Mirror image and skeleton are more efficient than the Avatar and Abstract in attracting people to the screen, while the in-situ study showed a preference for Mirror image over skeleton. Ackad et al. [1] compared skeletons vs. silhouettes using a public information display, and found that skeletons support longer interactions than silhouettes and facilitate play, while silhouette attract more passers-by and are better suited for more serious interactions. The difference between the in-lab and in-situ evaluation, however, highlights how these results may be context-dependent, and cannot be directly translated from an interaction with a ball game in a shop window into design recommendations for crafting HDI installations in museums.

Affordance blindness means that users of interactive displays often cannot guess how to interact with the screen. In other words, people cannot "discover" [10] how to control the installation. Gestures and body movements that can be used to control the display are either defined by designers, or using in-lab elicitation studies [59] [40] [10]. During elicitation studies, groups of potential users are exposed to animations that represent functionalities of the system, and asked to recommend how they would control that functionality. Although the agreement among participants in an in-lab elicitation study is used as a proxy of how well the gestures will be guessed, the surrounding context may alter the users' interaction patterns when the interactive installation is tested in a museum gallery. An alternative approach has been the use of the theory of embodied schemata [33] to guide the design of "intuitive" embodied interaction: designers rely on kinesthetic metaphors developed by humans at a very young age through constructivist explorations of the physical world (e.g., [4]). This line of work, however, implies a 1:1 relationship between one concept and one action (e.g., "fast" tempo - running "fast"), and is difficult to generalize to interactive installations that show data visualizations and support many functionalities.

In this paper, we take a different approach: our study explores how the way in which users are represented on the screen alters the gestures and body movements that people tend to do in front of a data visualization. This could help in reducing the affordance blindness, by making people converge towards specific groups of gestures.

SCENARIO: HUMAN-DATA INTERACTION WITH GLOBAL DATA IN MUSEUMS

The test bed for this study was a prototype HDI installation that we iteratively developed and tested at a science museum. Visitors can use this installation to explore and compare two datasets -see Figure 1. The visualization consists of two

interactive globes, displayed on a 65" screen. Data (from data.un.org) are visualized at a country level, using color gradations -i.e., darker colors mean higher data values. Each globe visualizes one dataset. After a preliminary brainstorming session with three museum directors, we included ten distinct pairs of datasets (20 datasets total) that we expected would have been thought-provoking for museum visitors. Above the globe, we included a scaffolding question that prompts visitors to reflect on possible correlations between the two datasets. For example, questions included: Does Fertilizer Consumption Influence the Number of Threatened Fish Species?; Does Firearm Ownership Influence the Number of Murders?; Does GDP Influence Deforestation?

The ultimate goal of this installation was to engage museum visitors in data "exploration" (not "analysis") tasks: because museum visitors spend generally two minutes or less interacting with a successful exhibit [53], we did not include more advanced, time-demanding data analyses tasks (e.g., cleaning or mining data). Rather, we incorporated three data "exploration" functionalities grounded in embodied HDI literature [10] (see Table 1). It is worth noting, however, that the evaluation of these learning goals goes beyond the scope of this study and is left to future work. Additionally, as recommended in [38], we implemented two alternative movements (jump and swipe) to change the data sets on display (visitors could explore 10 different pairs of data, as described below in the implementation section).

IMPLEMENTATION OF THE PROTOTYPE INSTALLATION

In this section, we provide a quick overview of our prototype HDI installation, i.e. a description of how the interactive prototype has been implemented. We used a 3D engine (Unity3D) in combination with a tracking camera (see hardware and software description below). The code was structured in three system components that work simultaneously in real-time: RGB Depth Sensor Controller, Gesture Manager, and Global Dataset Designer - see Figure 2.
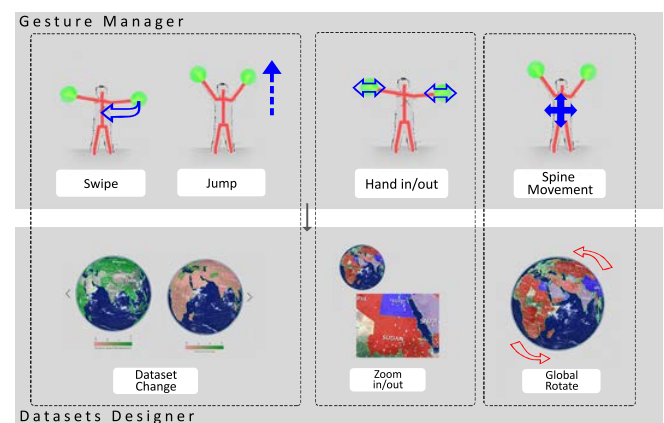

Figure 2. System overview. The two main components of our HDI prototype installation are: gesture manager, and global dataset designer.

First Component: RGB-Depth (RGB-D) Sensor Controller

The colors and depth data from the RGB-D sensors are the base of most gesture systems. Our system tracks the 3D location

Table 1. HDI Task that we support with functionalities for Human-Data-Interaction that we implemented in our prototype exhibit, and their STEM learning goal

HDI Task	Stem Learning Goal	System Functionality
Expose properties of data representation	See data from different viewpoints/perspectives (spatial reasoning)	Rotate Globes
Compare two data sets	See where many/few data elements are present (reason about quantities); Compare single data elements to whole (reason about proportions)	Zoom in and out
Change data sets	See different data sets which answer a different question each time that may indicate causation/correlation affect; the first globe represent a problem that effects other issue on the second globe	Jump or Swipe
Distill properties of a single data set	See specific data values (reason about quantities); Compare how data are represented (reason about data representation)	Select one data point "country" with a cursor

(x, y, z) of the user's joints that are then sent to the gesture manager.

Second Component: Gesture Manager
Four researchers iteratively designed, developed, and tested (in-lab) the code to recognize five gestures that people can use to control the interactive visualization. Specifically:

- **Spine movement.** When the user moves within the interaction space, we transform and rotate the angle of the two globes based on the spine direction As Table 1 indicates, the purpose of this functionality is the ability to explore and see the data from different perspectives on the globe to build a spatial reasoning.

- **Zoom in/out.** To get a closer look at the data presented on the globe, we coded a hand movement (moving hands closer together) to allow the user to zoom in up to the country level on the globe map. This gesture provides a comparison mechanism between two data sets. Each one is represented on a globe that has some correlation and causation reasoning, see Table 1.

- **Jump/Swipe.** Although these are two distinct gestures, we implemented both jump and swipe as continuous gestures. In other words, they are both required to record a certain joint movement (legs for jump, hands for swipe) for a period of time (window). Swipe starts when the user holds their hand above their shoulder, but this condition alone is not sufficient to trigger the gesture. The functionality of both gestures is changing the data sets presented on both globes. For instance, one of the data sets asks, "Does Fertilizer Consumption Influence Number of Threatened Fish Species?" where data about Fertilizer Consumption is represented on the first globe while the "Number of Threatened Fish Species" is displayed on the other. However, when a user jumps/swipes, the question and data sets change e.g, "Does Firearm Ownership Influence Number of Murders?", as shown in Table 1.

- **Grab.** The cursor and its functionalities are controlled based on the hand grab and release gestures. Such gestures are based on the tracking information from the hand and fingers joints. The gesture functionality provides the ability to explore each data set presented on the globe in depth based on each country or point on the globe.

Third Component: Global Datasets Designer
We designed an interactive globe map (by modifying a Unity3D globe map package) to provide realistic Earth and atmosphere settings where each country can be represented. When the user moves to the next pair of datasets, the datasets are automatically uploaded in the designated location on the global map and a new gradient color is applied to each country (to reflect the data value loaded from the dataset files). A normalization of the color value in the range [0-1] is performed for consistent gradient scheme.

Implementation of the Mode Type. We implemented the four mode types as follows: (1) Control - The user is not represented on the screen; (2) skeleton - A sketch of the user's joints is drawn in between the two globes; (3) Avatar - a robot resembles the user movements (i.e., the robot joints move like the user's joints); and, (4) Camera "Overlay" - the live video-feed from the camera is shown behind the globes. The four mode types are illustrated in Figure 3.

Hardware Description. The system runs on an Intel® Core™ i7-4710HQ CPU @ 2.50GHz 2501 MHz, 4 core(s), 8 logical processor(s) 16.0 GB RAM, and NVIDIA GeForce GTX 970 GPU. For the RGB-D sensor, we tested two devices in-lab: (1) Orbbec Astra 3D camera, and (2) Microsoft Kinect V2. For the experiment that we describe in this paper, we used the Microsoft Kinect v.2 because we noticed (in-lab) that it had a better tracking accuracy on the gestures that we implemented. The visualization was shown on a 65" TV screen mounted on a Peerless cart.

Software Description. We used a 3D engine (Unity3d) as the main platform for the entire system. For body tracking and gesture support, we used Kinect for Windows Software Development Kit (SDK) 2.0 for the Kinect camera, and developed the gestural interaction system in C#.

METHODOLOGY
The overarching aim of this study was to investigate the effect of different ways of representing the user (mode type) on how museum visitors engage and interact with an HDI installation. We focused on three alternative ways of representing the user: 1) skeleton, 2) Avatar, or the 3) full Camera overlay. Additionally, we considered a Control condition in which the user was not represented on the data visualization. Throughout the paper, we refer to these four experimental conditions as "Mode Type" -see Figure 3. In preparation for this study, we conducted five rounds of unstructured interviews with museum interpreters and directors to elicit feedback on the four representations that we designed with our research team (for

example, this process helped us to choose a more audience-appropriated avatar than the one we originally crafted).
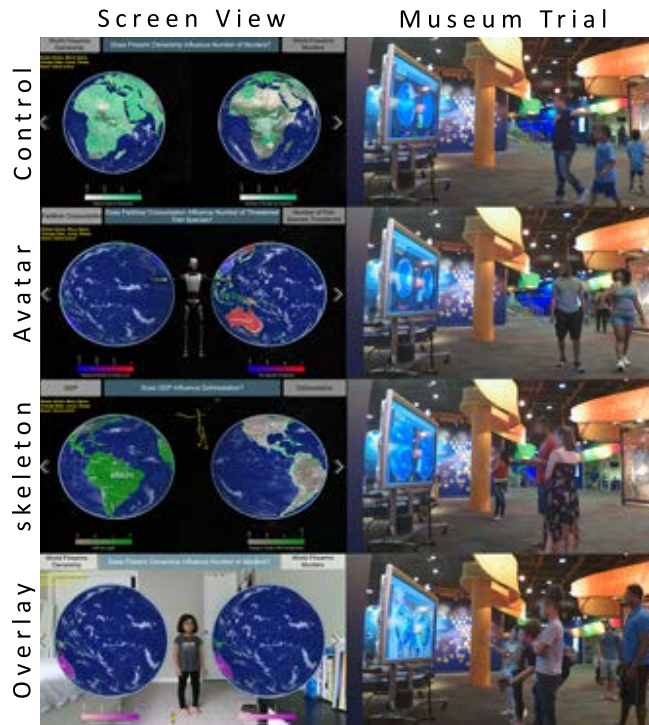

Figure 3. Screen view and example of museum trial for each of the four mode types (control, avatar, skeleton, and camera overlay)

All experimental sessions were video recorded.

Participants and Location
A total of 731 museum visitors participated in this study (406 children and 325 adults). The study was conducted at Discovery Place, a science museum in Charlotte, North Carolina, United States.

Experimental Procedure
We conducted a quasi-experimental study (as common in museum research [45] [10]): participants were not randomly assigned to an experimental condition, but interacted with the version of the system that was active at the time of their visit. Specifically, we left a Mode Type (see Figure 3) active for 30 minutes, after which we switched to the next mode type. The sequence in which the four mode types were activated was randomized (to avoid, for example, always starting with one condition at the same time of the day). This approach allowed us to observe visitors' interaction "in-the-wild," without a direct intervention by the researchers (which would have compromised the ecological validity of this study).

Research Questions, Hypothesis, and Data
Our study focused on four research questions.

(R.Q.1) Which mode types are the most effective in attracting people towards the display?

Significance. As we described in the related work section, display blindness is a challenge for the design of interactive installations. If we notice a connection between how the user is represented on the screen (mode type) and the percent of visitors that interact with the screen, we can recommend the mode type that works best at making people aware of the display and luring them to interact.

Hypothesis. Prior literature on the design of interactive public displays in shop windows [42] showed in-lab that using the full camera overlay or a silhouette may be more effective than using an avatar or another abstract representation. We cannot rely on the lab study portion of the work in [42] to make design recommendations for museum settings because the context of use may significantly alter these findings. We expect, however, that using the full video feed from the camera would attract more visitors, because it allows people to see themselves on the screen in real time.

Data and Variable Definitions. We reviewed the videos of our experimental sessions and counted: (1) the number of people who were in transit near the screen, and (2) the number of people who stopped by to interact with the screen. Two researchers counted the total number of visitors using Anvil [30]. These are people who entered and exited the camera field of view. We were situated in a hallway, so all passersby had to step in front of the camera. This allowed us to compute the percent of visitors who interacted with the installation in each experimental session (we called this dependent variable "Capture Rate" [49]). For the statistical analysis, we were interested in the effect of Mode Type (independent variable) on the Capture Rate (dependent variable) -see Figure 4.

Figure 4. R.Q.1 Effect of Mode Type on Capture Rate.

(R.Q.2) Given the mode type, can we predict the gestures that users will most frequently do?

Significance. Increasing the discoverability of hand gestures and body movements is another challenge for the design of interactive installations. If people tend to perform different gestures and body movements when they are exposed to different mode types, we can develop design recommendations on which control actions should be implemented.

Hypothesis. We expected different mode types may prime people towards using different gestures: for example, the work by Ackad et al. [1] suggests that skeletons facilitate play. We were, however, unsure about the specific gestures.

Data and Variable Definitions. Using Anvil, a video annotation tool [30], two researchers listed each gestures and body movements that visitors performed in front of the screen (the coding process is described in the next sections). For the statistical analysis, we look at whether the mode type can be used to predict the gestures ("interaction type") that people most frequently do in front of the screen.

(R.Q.3) What is the effect of mode type and interaction type (gestures that people perform in front of the display) on engagement (i.e., the amount of time users that spend interacting with the display)?

Significance. We used the amount of time that museum visitors spend interacting with an installation as a proxy for their engagement with the HDI system. We acknowledge that this assumption may not be true in other application contexts (e.g., an higher time on task may mean that the user is spending a considerable amount of time trying to figure out how to use a spreadsheet). In museums, however, people tend to leave thinking that the system is broken if they are not able to use it within 10 to 15 seconds [10], because they do not have a specific task to carry out. Furthermore, the more time museum visitors spend interacting with the screen, the higher are the chances that other people will notice the interactivity and try the installation because of the honeypot effect [7]: visitors tend to congregate with other visitors. Thus, we want to identify the mode types and interaction types that are able to engage the user for the longest time.

Hypothesis. We expected some interaction to last longer (e.g., jumping in front of the screen takes a longer time than waving). Additionally, we expected mode type to affect the length of the interaction: this would extend Ackad et al.'s [1] observation that skeletons support longer interactions than silhouettes.

Data and Variable Definitions. Using Anvil [30], two researchers coded the amount of time that each visitor spent performing each gesture or body movement (the coding process is described in the next sections). We used this amount of time as a proxy for "Engagement." In the statistical analysis, we looked at the effect of Mode Type (4 levels: Avatar, Control, etc.) and Interaction Type (11 levels: Spine, Swipe, Jump, etc.) on Engagement -see Figure 5.
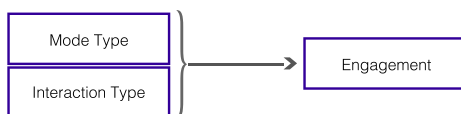

Figure 5. R.Q.3 Effect of Mode Type and Interaction Type on Engagement.

(R.Q.4) How does the mode type -moderated by the person type -impact attention (i.e., the amount of time spent looking at the display)?

Significance. The goal of an interactive installation for data exploration is to allow people to observe and discuss patterns in data. Ultimately, we want visitors to learn from what they see on the screen and from the informal conversations they have with each other [50]. In order to have conversations around the data on display, however, visitors need to spend time looking at data and data patterns. **Hypothesis.** We expected that less dynamic some mode types (e.g., Control) may be more conducive to looking and reflecting on the data.

Data and Variable Definitions. Using Anvil [30], two researchers coded the amount of time that each visitor spent looking at the display (the coding process is described in the

next sections). We used this amount of time as a proxy for "Attention." In the statistical analysis, we looked at the effect of Mode Type on Attention. We believe that the interaction between these two variables might depend on the Type of Person (child or adult) because people of different age groups have different attention spans [48],. Thus, we considered Type of Person as the moderator variable in the study - see Figure 6. A moderator variable, as defined by Baron & Kenny[6], is a "...qualitative (e.g., sex, race, class) or quantitative (e.g., level of reward) variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable. Specifically, within a correlation analysis framework, a moderator is a third variable that affects the zero-order correlation between two other variables. ...In the more familiar analysis of variance (ANOVA) terms, a basic moderator effect can be represented as an interaction between a focal independent variable and a factor that specifies the appropriate conditions for its operation" (p. 1174).
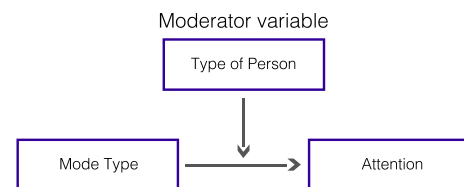

Figure 6. R.Q.4 Effect of Mode Type on visitor's Attention.

Coding

Two researchers reviewed and coded the videos of the experimental sessions using Anvil [30].

In the preliminary phase of the coding process, the two researchers individually coded one of the video files and then met to iteratively develop a common coding dictionary. The coding was structured using multiple tracks: mode type, person type, interaction type. The resulting data included the start and end time of each coded portion of the video file. The Mode Type track specifies the experimental conditions: Control, Avatar, skeleton, or Camera Overlay. We maintained eight "Person" groups, each of which has two main tracks (Interaction Type and Person Type). The Person Type has two main attributes [Adult, and Child]. During the coding, the researchers noticed that visitors frequently performed some gestures and body movements that were not actually implemented in the system (e.g., dancing in front of the screen was a recurrent movement). Thus, while coding the Interaction track, the two researchers agreed a posteriori on a dictionary of recurrent gestures and body movements that they coded even if the installation did not actually support them. This resulted in 11 interaction types: Spine, Swipe, Jump, Zoom, Touch, Dance, Wave, Pan, Crouch, Grab, and Other. Additionally, the researchers used the same track to code the amount of time people spent looking at the screen, while not interacting (we did not use a different track for Looking because we defined Look and Interaction Type as mutually exclusive events).

Inter-Rater Reliability

After researchers agreed on a set of tracks and on a coding dictionary, two researchers independently coded the videos

of the experimental sessions. Table 2 reports the result of the inter-rater reliability for the two coders.

The analysis includes three types of agreements: (1) segmentation agreement (i.e., whether the two coders agreed on the beginning and end time for a code), (2) category agreement (i.e., whether the two coders agreed on the actual code on a segment of the video -when considering only the slices that both coders annotated), and (3) overall coding agreement. We computed Cohen's kappa coefficient (κ) to measure the inter-rater agreement for these categorical items.

Table 2 depicts a high level of agreement on most of the three agreement types (even though the overall coding agreement is probably the most relevant for our analysis). Not surprisingly, the "ModeType" scores the highest value of agreement because it simply refers to which mode type was used (Control, Avatar, skeleton, or Camera). While coding the "Interaction," instead, we anticipated that some gestures could have been interpreted in different ways by different coders. The overall agreement, however, proved to be high also in this case.

	SEGMENTATION		CATEGORY		OVERALL	
Track.Attribute	%	κ	%	κ	%	κ
M.Modetype	99.9	0	100	1	99.99	0
P1.Interaction	97.82	0.94	86.26	0.82	94.47	0.87
P1.PersonType	98.31	0.95	100	1	98.31	0.96
P2.Interaction	98.24	0.94	89.35	0.86	96.17	0.89
P2.PersonType	97.71	0.93	100	1	97.71	0.93
P3.Interaction	97.72	0.89	88.56	0.83	96.37	0.84
P3.PersonType	97.72	0.89	100	1	97.72	0.90

Table 2. Results of the inter-rater reliability on all video slices are considered by two different coders, where (κ) is Cohen's kappa coefficient, P refer to Person, and M is short for mode (α). For brevity, we report only three "Person" tracks.

RESULTS
We collected, coded, and analyzed a total of 7.13 hours of videos. In this section, we present the results of our statistical analysis for each of the research questions that we investigated.

(R.Q.1) Effect of Mode Type on Capture Rate
A total of 731 museum visitors interacted with the display ("number of users") across the four mode types: 85 in the Control condition, 99 in the Avatar condition, 199 in the skeleton condition, and 348 in the Camera condition. Overall, we observed 2507 museum visitors who entered the field of view of the camera during our experimental sessions ("number of visitors"): 630 during Control, 333 during Avatar, 594 during skeleton, and 950 during Camera. Due to the varied number of visitors across experimental conditions (we used a quasi-experimental design, and the museum was more crowded during some hours of the day than others), we considered the "capture rate" [49] (i.e., the percent of users who interacted with the system) for our statistical analysis, defined as:

$$CaptureRate = NumberOfUsers/NumberOfVisitors \quad (1)$$

Precisely, the mode type (e.g., Avatar, Control) was the independent variable, and the capture rate the dependent variable. A one-way ANOVA was conducted to determine if the percent of museum visitors who interacted with the display ("capture rate") was different when using alternative mode types. The data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p < .05$). There was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .179$). The capture rate was statistically significantly different across different mode types, $F(3, 16) = 6.717, p < .0001$. The average capture rate in each experimental session increased from skeleton ($M = 0.17, SD = 0.06$) to Control ($M = 0.25, SD = 0.09$) to Avatar ($M = 0.37, SD = 0.11$) to Camera ($M = 0.38, SD = 0.105$). Tukey post-hoc analysis revealed that Avatar and Camera attracted a statistically significant higher percent of visitors than skeleton ($p = 0.038 and p = 0.003 respectively$). In other words, Avatar and Camera performed better than skeleton in luring visitors to interact with our embodied HDI system.

(R.Q.2) Predicting Gestures
Goodman and Kruskal's lambda (λ) is a statistical test that provides a measure of association between an independent nominal variable (in our case, the interaction that a visitor does in front of the screen, e.g. Swipe, Dance, etc.) and a dependent nominal variable (in our case, the mode type, e.g. Camera). What makes this test particularly relevant for our analysis is that λ can be interpreted as the reduction in the error when predicting the value of a dependent variable, given the knowledge of the independent variable. For example, the gesture that was used the most during all our experimental sessions was Spine (307 total occurrences) -see Table 3. Thus, without any knowledge of the mode type, we could wrongly conclude that Spine is always the preferred interaction. Table 3, however, reveals that Pan is actually the most used interaction in the Control condition.

Interaction	Mode (%)				Total
	Control	Skeleton	Avatar	Camera	
Spine	7.17	22.15	16.29	54.40	22.09%
Swipe	8.79	32.97	19.78	38.46	6.55%
Jump	3.13	19.79	25.00	52.08	6.91%
Zoom	7.84	23.53	23.53	45.10	3.67%
Touch	14.95	23.36	18.69	42.99	7.70%
Dance	4.95	15.32	22.07	57.66	15.97%
Wave	10.26	42.31	14.10	33.33	5.61%
Pan	13.00	26.91	4.48	55.61	16.04%
Crouch	4.55	13.64	4.55	77.27	1.58%
Grab	3.70	29.63	33.33	33.33	1.94%
Other	5.42	28.31	41.57	24.70	11.94%

Table 3. Percentage of visitors who used each interaction type in each experimental condition. The last column reports the total percent of people who used each interaction type across all the experimental conditions.

Goodman and Kruskal's λ was run with the interaction type as dependent variable to determine whether the interaction type could be better predicted by knowledge of the mode type. Goodman and Kruskal's λ was .26. This was a statistically significant reduction in the proportion of errors due to the

knowledge of mode type as a predictor of interaction type, $p < .0005$. In other words, by knowing how the user is represented on the screen (mode type), we can better predict which gestures museum visitors will most likely do in front of the screen. For example, people in the Camera condition predominantly use Spine, or Dance, while users in the Control condition predominately perform a pan gesture.

(R.Q.3) Effect of Mode Type and Interaction Type on Engagement (Interaction Time)

A two-way ANOVA was conducted to understand the effects of mode type and interaction type on the time spent interacting with the display (we used the interaction time as a proxy for the visitors' "engagement" with the installation). Mode type included four levels: 1) Control, 2) Avatar, 3) skeleton, and 4) Camera. Interaction Type included 11 levels (as listed in the coding section). Type of mode will be abbreviated as mode, type of interaction will be abbreviated as interaction, and interaction time as time for the rest of the results.

Time was not normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$) and the assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances, $p < .05$. This occurred due to the nature of the quasi-experimental research design. ANOVAs, however, are considered to be fairly "robust" to deviations from normality and heterogeneity of variance [35] [27]. There was a statistically significant interaction between mode and interaction on time, $F(20, 1269) = 2.603, p < .001$, partial $\eta^2 = .058$. Therefore, an analysis of simple main effects for interaction was performed with a statistical significance receiving a Bonferroni adjustment and being accepted at an alpha level of $.05/11 = .0045$. There was a statistically significant difference in mean "Time" duration between 11 types of interactions in the camera mode, $F(10, 1269) = 6.42, p < .001$, partial $\eta^2 = .048$.

As shown in Figure 7, users interacting in the camera mode spent more time *dancing* ($M = 7.48, SD = .503$) than jumping ($M = 2.04, SD = .809$), a statistically significant mean difference of $5.44(95\%CI, 2.27, 8.61), p < .001$. Users in the camera mode also spent more time *dancing* ($M = 7.48, SD = .503$) than moving their spine ($M = 3.55, SD = .438$), a significant mean difference of $3.93(95\%CI, 1.71, 6.15), p < .001$. Moreover, visitors spent more time *panning* ($M = 6.23, SD = .509$) than jumping ($M = 2.04, SD = .809$) in the camera mode, a significant mean difference of $4.17(95\%CI, 1, 7.36), p < .001$.

Significance was also found in the control mode. Users spent more time *touching* ($M = 14.72, SD = 1.42$) the display than dancing ($M = 5.91, SD = 1.71$), mean difference of $8.806(95\%CI, 1.43, 16.18), p = .004$, *panning* ($M = 5.03, SD = 1.05$), mean difference of $9.68(95\%CI, 3.82, 15.55)p < .001$, using their *spine* ($M = 3.98, SD = 1.21$), mean difference of $10.73(95\%CI, 4.54, 16.92), p < .001$, *waving* ($M = 3.92, SD = 2$), mean difference of $10.8(95\%CI, 2.64, 18.96), p = .001$, *swiping* ($M = 3.7, SD = 2$), mean difference of $11.02(95\%CI, 2.86, 19.17), p < .001$ and doing *other* types of interactions ($M = 3.43, SD = 1.88$)

,mean difference of $11.29(95\%CI, 3.44, 19.14), p < .001$. There was also a significant difference in mean "Time" duration between 11 types of interactions in the control mode, $F(10, 1269) = 4.82, p < .001$, partial $\eta^2 = .037$. In summary, these results show that different mode types such as camera and control afford specific types of interactions and affect the time spent engaging visitors.
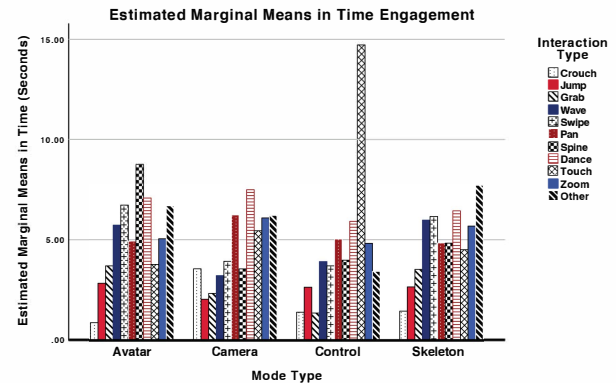
Figure 7. Statistical difference in the mean duration of time between the 11 types of interactions in each mode type

(R.Q.4) Impact of Mode Type on Attention (i.e., Time Spent Looking at the Display)

We were also interested in understanding which mode (control, avatar, skeleton, and camera -independent variable) made different types of visitors (adult or child -moderator variable) spend more time looking at the display (dependent variable). A two-way ANOVA was conducted to examine this effect. Assumptions of normality and homogeneity of variances were violated as well, following the described nature of the quasi-experimental study as mentioned before.

The reasons why we used a two-way ANOVA, rather than a different a statistical test (such as a MANOVA) with both Engagement and Attention as dependent variables, are that: (1) looking is not properly a gesture, rather we coded Look as the time spent without interacting with the screen (so we could not conceptually consider it as a 12-th level of "Interaction Type"); (2) we had a distinct, a priori set of hypotheses for R.Q.3 vs. R.Q.4; and, (3) only one dependent variable was captured (time) -we did not measure the time a visitor looked at the screen exactly after each interaction, because the act of looking occurred occasionally.

There was a statistically significant interaction between mode and visitor on time, $F(3, 1011) = 8.91, p < .001$, partial $\eta^2 = .03$. Therefore, an analysis of simple main effects for mode was performed with a statistical significance receiving a Bonferroni adjustment and being accepted at an alpha level of $.05/4 = .0125$, while for type of person was accepted at an alpha level of $.05/2 = .025$. There was a statistically significant difference in mean time between children and adults in the *avatar* mode, $F(1, 1011) = 59.83, p < .001$, partial $\eta^2 = .056$. *skeleton* mode, $F(1, 1011) = 14.67, p < .001$, partial $\eta^2 = .014$, and *camera* mode, $F(1, 1011) = 25.04, p < .001$, partial $\eta^2 = .024$. Results indicate that *adults* (M = 25.34, SD = 1.74)

spent more time in avatar mode than children (M = 8.16, SD = 1.38), p < .001, as well as skeleton ($M = 11.30, SD = 1.19$), and camera ($M = 11.07, SD = .89$) - see Figure 8.

Regardless of person type, there was a statistically significant main effect of mode, $F(3, 1011) = 18.13, p < .001$, partial $\eta^2 = .05$. Post-hoc comparisons were accepted at Bonferroni corrected alpha of $.05/4 = .0125$. Analysis indicates that visitors spent more time looking at the avatar mode ($M = 14.78, SD = 1.13$) than control ($M = 6.98, SD = 1.23$), skeleton ($M = 7.86, SD = .82$), and camera ($M = 7.6, SD = .58$), $p < .001$.
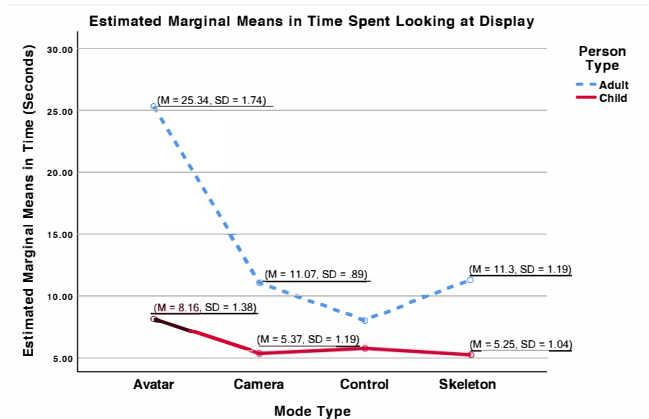

Figure 8. Statistical difference in mean duration of time spent looking at display in each mode type moderated by person type

DISCUSSION

Mode Type and Display Blindness
The results to (R.Q.1) confirmed our hypothesis that some mode types (i.e., different ways of representing the user near a data visualization) are more efficient than others in luring museum visitors into the interaction with an HDI installation. Our results that Avatar and Camera Overlay can be more engaging that Skeleton extend the work of Ackad et al. [1] by comparing the effect on display blindness of a larger number of mode types.

Specifically, our findings suggest that camera overlay and avatar (not control or skeleton) may be the most efficient designs for coping with the issue of display blindness in the context of museum installations. Future work, however, should be done to assess if and how our results can be generalized when using different data visualizations (e.g., a bar chart rather than a globe map) and when the installation is moved to different locations within the museum or even to a different museum (e.g., a children's museum rather than a science museum).

Implications for Human-Data Interaction
The results of our study suggest a design feature and a design guideline for crafting engaging Human-Data Interactions.

Design feature: Using the user's body as a cursor. Both the Avatar and Overlay conditions (the most engaging in our study) provide a full representation of the user's body. Thus, portions of the userâĂŹs body (e.g., hands) could be used as navigation

cursors (e.g., [51]) or integrated with design elements (like the magnifying lenses in Kister et al.'s BodyLense [31]) to provide access to specific data points.

Design guideline: Matching the mode type with the system purpose. Considering that visitors tend to perform different movements depending on the mode type, the way in which the user is represented on the screen needs to match the system purpose. For example, if the visualization is designed to enable fine-grained data analysis tasks, it should use a mode type such as Control that promotes pan-like gestures; vice-versa, more casual data exploration activities and play can be better supported by Skeleton [1] or Overlay.

Moving Beyond In-Lab Elicitation Studies: Implications for the Design of Gestures for Embodied Interaction
Our findings indicate that the mode type (i.e., the way in which museum visitors are represented on the screen) can be used to improve the prediction on which gestures and body movements people will most likely do. This is particularly interesting because we did not instruct museum visitors on how to use the system: the prototype installation was on display without researchers and moderators; visitors were not actively recruited for this study, nor they were asked to guess how to operate the system. Thus, our results provide guidelines for the design of gestures and body movements for HDI: for example, camera overlay seems to promote more active gestures (like moving spine and jumping) than control. Furthermore, our findings have implications for the design of elicitation studies (e.g., [59] [40]) in general: researchers need to be mindful that alternative visualizations (including alternative ways of visualizing the users) -not just the system functionality/effect -have an impact on the gestures and body movements that people may want to do. Similarly, when designers use approaches that recommend incorporating priming during the elicitation process (e.g., [10]), they should be aware that the visualization itself may provide an implicit priming to the users, and lead them into recommending specific classes of gestures and body movements.

Ultimately, we believe that future work should investigate if and how elicitation studies should move out of research labs. In a different application context, Ali et al. suggests running "crowd-powered elicitation study" [2]. In museum settings, researchers need to craft strategies for conducting elicitation studies in-situ: researchers may want to assess if conducting elicitation studies in-situ, with actual museum visitors (rather than in-lab, generally with participants recruited in an academic setting), will produce gestures and body movements that are more "intuitive" and easier to be discovered. This would provide a strategy to incorporate visitors' spontaneous and/or playful gestures [54] into the design process.

Touching the Screen
Although we did not use a touch screen for our study, we observed 107 instances of museum visitors trying to touch the display (16 in Control, 25 in skeleton, 20 in Avatar, and 46 in Camera). Unfortunately, this could derail the interaction, if visitors assume that they are in front of an unresponsive touch screen (rather than a tracking camera that recognizes gestures

and body movements). This problem is in line with what has been reported in existing literature on pervasive displays (e.g., [51]), although in different application contexts (an interactive art jigsaw). Humans, ever since they are babies, are attuned to directly seeing their image at every turn, from mirrors, images, and video. This phenomenon is pertinent more than ever with the advent of social media and the tenet of modern society, the selfie, providing an ever ending means of recording and reflecting on one's self. Historically, the image of the body is an integral part of the development of symbolic activity: the child in front of the mirror moves from the perception of the fragmented images of the body to the recognition of its unity [36]. This might be why users tend to touch the screen more in the control condition due to the assumed belief that they cannot use their gestures and body to interact with the data: their representation of self is not synthesized in their mental space due to a lack of visual reinforcement. In our opinion, however, this phenomenon is bound to disappear with time: it is interesting, for example, how back in 2008 Hornecker [21] reported that museum visitors had a difficult time using a touch-screen in a museum. Just few years later, basics touch-screen gestures (such as pinch and zoom) had become part of our lexicon on how we interact with technologies up to the point that they generate legacy biases during elicitation studies [39] [52]. In the meantime, designers may use two alternative strategies to cope with this phenomenon: either implementing hybrid screens that can be controlled by both mid-air gestures and directly by touching the screen itself; or, researching which visualization snippets could be included to communicate to the users that the display is not touch-enabled (by extending, for instance, the avatar strategy discussed in [51] from interactive art to HDI systems).

Impact on Interaction Time and Look
Our findings revealed that visitors' engagement with the display is the simultaneous effect of the mode type and interaction type. As expected, some interactions such as dancing and panning lasted longer than jumping. Something particularly interesting is that users in the camera mode spent more time *dancing* in front of display than moving their spine alone. A possible explanation may be that dancing is a social activity [55] that may capture visitors for a longer time which, in turn, increases the probability that other people will start dancing because they see other interacting (the honeypot effect) [7]. Future work should investigate which gestures and body movements are better suited to be performed in a social setting.

In regards to time spent *looking* at the display, results show that, regardless of the person type, most visitor's attention was drawn to the avatar, which is not what we expected. A possible explanation may be that seeing an additional object on the screen (i.e., the avatar) [28] (rather than just the data, as in the control condition) and certain kinds of contrast/luminance changes [18] that are more visible on the avatar than on a simpler skeleton can contribute to attract people's attention towards the screen.

Adults spent more time looking at the display significantly than children, which leads us to assume that they read and reflected on the data presented in front of them. Perhaps the types of questions were more attuned to their age and development than the children present at the time, so additional work needs to be done to develop strategies on how to visualize multiple HDI-relevant questions that may be engaging for both adults and children [46]. Future work should also analyze visitors' conversations and remarks to explore if avatar (or other mode types) enabled visitors, especially adults, to discuss insights on the data -an essential component of the learning process in museums [3].

Implementation Challenges
The depth of field in the RGB-D camera which is used to develop the system has some limitations that designers should consider, including only allowing six users at the same time. This may restrict the interaction opportunities when larger groups of users (e.g., a class during a school trip) enter the field of view of the camera. A possible solution could be to use multiple tracking cameras -although this approach would require calibrations [43]. Also, although the tracking algorithms are becoming more and more robust, there is still a chance that the identity of a user can be "stolen" if another person occludes that user or if the user accidentally leaves and re-enters the interaction space [11].

Finally, we want to remind the reader that only a subset of gestures that we considered during the coding process were implemented, while others were spontaneous ones that visitors frequently performed in front of the screen. On the grounds of how our system was implemented, some of the visitors' spontaneous gestures actually controlled the data. For example, while we did not explicitly implement "dancing," dancing makes the globe rotate because they are linked to the user's spine position. In our view, this partially loosens the difference between gestures that we implemented a-priori and spontaneous ones.

CONCLUSION AND FUTURE WORK
In this paper, we tackled four challenges for the design of interactive HDI installations: limiting display blindness and attracting people towards the screen; aiding visitors in guessing the gestures and body movements that control the system; keeping users engaged in the interaction; and, sustaining visitors' attention on the data. Our results indicate that the design of the visualization and, in particular, how we represent people on the screen (mode type) impacts the number of visitors that interact with the installation, the gestures that people do, and the amount of time that visitors spend observing the data on display and interacting with the system. Future work should investigate whether mode type and interaction type also influence the way in which visitors discuss the data sets on the screen (probably by taking a more qualitative approach in terms of research design).

ACKNOWLEDGMENTS

REFERENCES

[1] Christopher Ackad, Martin Tomitsch, and Judy Kay. 2016. Skeletons and Silhouettes: Comparing User Representations at a Gesture-Based Large Display. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16)*. Association for Computing Machinery, New York, NY, USA, 2343–2347. DOI: `http://dx.doi.org/10.1145/2858036.2858427`

[2] Abdullah X. Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. 2019. Crowdlicit: A System for Conducting Distributed End-User Elicitation and Identification Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 255, 12 pages. DOI: `http://dx.doi.org/10.1145/3290605.3300485`

[3] Sue Allen. 2003. Looking for learning in visitor talk: A methodological exploration. In *Learning conversations in museums*. Routledge, UK, 265–309.

[4] Alissa N Antle, Milena Droumeva, and Greg Corness. 2008. Playing with The Sound Maker : Do Embodied Metaphors Help Children Learn ? *Proceedings of the 7th international conference on Interaction design and children IDC 08* USA, USA (2008), 178–185. `http://portal.acm.www.library.gatech.edu:2048/citation.cfm?id=1463689.1463754`

[5] Till Ballendat, Nicolai Marquardt, and Saul Greenberg. 2010. Proxemic Interaction: Designing for a Proximity and Orientation-aware Environment. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, 121–130. DOI: `http://dx.doi.org/10.1145/1936652.1936676`

[6] Reuben M Baron and David A Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51, 6 (1986), 1173.

[7] Harry Brignull and Yvonne Rogers. 2003. Enticing people to interact with large public displays in public spaces. In *Proceedings of INTERACT*, Vol. 3. INTERACT, USA, 17–24.

[8] Federico Cabitza, Angela Locoro, Daniela Fogli, and Massimiliano Giacomin. 2016. Valuable Visualization of Healthcare Information: From the Quantified Self Data to Conversations. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '16)*. ACM, New York, 376–380. DOI: `http://dx.doi.org/10.1145/2909132.2927474`

[9] Francesco Cafaro. 2012. Using embodied allegories to design gesture suites for human-data interaction. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. ACM Press, New York, New York, USA, 560. DOI: `http://dx.doi.org/10.1145/2370216.2370309`

[10] Francesco Cafaro, Leilah Lyons, and Alissa N. Antle. 2018. Framed Guessability: Improving the Discoverability of Gestures and Body Movements for Full-Body Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 593, 12 pages. DOI:`http://dx.doi.org/10.1145/3173574.3174167`

[11] Francesco Cafaro, Alessandro Panella, Leilah Lyons, Jessica Roberts, and Josh Radinsky. 2013. I see you there!: developing identity-preserving embodied interaction for museum exhibits. In *Proceedings of the 2013 ACM annual conference on Human factors in computing systems*. ACM, New York, 1911–1920. DOI: `http://dx.doi.org/10.1145/2466110.2466252`

[12] Victor Cheung, Diane Watson, Jo Vermeulen, Mark Hancock, and Stacey Scott. 2014. Overcoming Interaction Barriers in Large Public Displays Using Personal Devices. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)*. ACM, New York, NY, USA, 375–380. DOI: `http://dx.doi.org/10.1145/2669485.2669549`

[13] Sandy Claes and Andrew Vande Moere. 2015. The Role of Tangible Interaction in Exploring Information on Public Visualization Displays. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis '15)*. ACM, New York, NY, USA, 201–207. DOI: `http://dx.doi.org/10.1145/2757710.2757733`

[14] Jorgos Coenen, Sandy Claes, and Andrew Vande Moere. 2017. The concurrent use of touch and mid-air gestures or floor mat interaction on a public display. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays*. ACM, ACM, USA, 9.

[15] Paul Dourish. 2001. *Where the action is: the foundations of embodied interaction*. MIT Press, Cambridge, MA, USA.

[16] Niklas Elmqvist. 2011. Embodied human-data interaction. In *ACM CHI 2011 Workshop "Embodied Interaction: Theory and Practice in HCI"*. ACM, USA, 104–107.

[17] David England. 2011. Whole Body Interaction: An Introduction. In *Whole Body Interaction An Introduction*, David England (Ed.). Springer London, UK, 1–5. DOI: `http://dx.doi.org/10.1007/978-0-85729-433-3_1`

[18] James T Enns, Erin L Austen, Vincent Di Lollo, Robert Rauschenberger, and Steven Yantis. 2001. New objects dominate luminance transients in setting attentional priority. *Journal of Experimental Psychology: Human Perception and Performance* 27, 6 (2001), 1287.

[19] Vito Gentile, Salvatore Sorce, Alessio Malizia, Dario Pirrello, and Antonio Gentile. 2016. Touchless Interfaces For Public Displays: Can We Deliver Interface Designers From Introducing Artificial Push Button Gestures?. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI*

'16). ACM, New York, NY, USA, 40–43. DOI:
`http://dx.doi.org/10.1145/2909132.2909282`

[20] Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob
Diaz-Marino, and Miaosen Wang. 2011. Proxemic
Interactions: The New Ubicomp? *Interactions* 18, 1
(Jan. 2011), 42–50. DOI:
`http://dx.doi.org/10.1145/1897239.1897250`

[21] E. Hornecker. 2008. "I don't understand it either, but it
is cool" - visitor interactions with a multi-touch table in
a museum. In *2008 3rd IEEE International Workshop on
Horizontal Interactive Human Computer Systems.* IEEE,
USA, 113–120. DOI:
`http://dx.doi.org/10.1109/TABLETOP.2008.4660193`

[22] Eva Hornecker. 2011. The role of physicality in tangible
and embodied interactions. *interactions* 18, 2 (2011),
19–23.

[23] Heiko Hornung, Roberto Pereira, M Cecilia C
Baranauskas, and Kecheng Liu. 2015. Challenges for
human-data interaction–a semiotic perspective. In
*International Conference on Human-Computer
Interaction.* Springer, Springer, USA, 37–48.

[24] Steven Houben, Ben Bengler, Daniel Gavrilov, Sarah
Gallacher, Valentina Nisi, Nuno Jardim Nunes, Licia
Capra, and Yvonne Rogers. 2019. Roam-IO: Engaging
with People Tracking Data through an Interactive
Physical Data Installation. In *Proceedings of the 2019
on Designing Interactive Systems Conference (DIS'19).*
Association for Computing Machinery, New York, NY,
USA, 1157–1169. DOI:
`http://dx.doi.org/10.1145/3322276.3322303`

[25] Steven Houben and Christian Weichel. 2013.
Overcoming Interaction Blindness Through Curiosity
Objects. In *CHI'13 Extended Abstracts on Human
Factors in Computing Systems (CHI EA'13).* ACM, New
York, NY, USA, 1539–1544. DOI:
`http://dx.doi.org/10.1145/2468356.2468631`

[26] Elaine M Huang, Anna Koster, and Jan Borchers. 2008.
Overcoming assumptions and uncovering practices:
When does the public really look at public displays?. In
International Conference on Pervasive Computing.
Springer, Springer, USA, 228–243.

[27] James Jaccard and Jim Jaccard. 1998. *Interaction effects
in factorial analysis of variance.* Sage, USA.

[28] John Jonides and Steven Yantis. 1988. Uniqueness of
abrupt visual onset in capturing attention. *Perception &
psychophysics* 43, 4 (1988), 346–354.

[29] Mohamed Khamis, Christian Becker, Andreas Bulling,
and Florian Alt. 2018. Which One is Me?: Identifying
Oneself on Public Displays. In *Proceedings of the 2018
CHI Conference on Human Factors in Computing
Systems (CHI '18).* ACM, New York, NY, USA, Article
287, 12 pages. DOI:
`http://dx.doi.org/10.1145/3173574.3173861`

[30] Michael Kipp. 2014. Anvil: A universal video research
tool. *Handbook of corpus phonology* USA, USA (2014),
420–436.

[31] Ulrike Kister, Patrick Reipschläger, Fabrice Matulic,
and Raimund Dachselt. 2015. BodyLenses: Embodied
Magic Lenses and Personal Territories for Wall Displays.
In *Proceedings of the 2015 International Conference on
Interactive Tabletops & Surfaces (ITS âĂŹ15).*
Association for Computing Machinery, New York, NY,
USA, 117âĂŞ126. DOI:
`http://dx.doi.org/10.1145/2817721.2817726`

[32] Hannu Kukka, Heidi Oja, Vassilis Kostakos, Jorge
Gonçalves, and Timo Ojala. 2013. What Makes You
Click: Exploring Visual Signals to Entice Interaction on
Public Displays. In *Proceedings of the SIGCHI
Conference on Human Factors in Computing Systems
(CHI '13).* ACM, New York, NY, USA, 1699–1708.
DOI:`http://dx.doi.org/10.1145/2470654.2466225`

[33] George Lakoff and Mark Johnson. 2008. *Metaphors we
live by.* University of Chicago press, USA.

[34] Robb Lindgren and Mina Johnson-Glenberg. 2013.
Emboldened by Embodiment Six Precepts for Research
on Embodied Learning and Mixed Reality. *Educational
Researcher* 42, 8 (2013), 445–452. DOI:
`http://dx.doi.org/10.31020013189X13511661`

[35] Scott E Maxwell, Harold D Delaney, and Ken Kelley.
2017. *Designing experiments and analyzing data: A
model comparison perspective.* Routledge, UK.

[36] Sabine Melchoir-Bonnet. 2014. *The mirror: A history.*
Routledge, UK.

[37] Nemanja Memarovic, Sarah Clinch, and Florian Alt.
2015. Understanding Display Blindness in Future
Display Deployments. In *Proceedings of the 4th
International Symposium on Pervasive Displays (PerDis
'15).* ACM, New York, NY, USA, 7–14. DOI:
`http://dx.doi.org/10.1145/2757710.2757719`

[38] Swati Mishra and Francesco Cafaro. 2018. Full Body
Interaction Beyond Fun: Engaging Museum Visitors in
Human-Data Interaction. In *Proceedings of the Twelfth
International Conference on Tangible, Embedded, and
Embodied Interaction (TEI '18).* ACM, New York, NY,
USA, 313–319. DOI:
`http://dx.doi.org/10.1145/3173225.3173291`

[39] Meredith Ringel Morris, Andreea Danielescu, Steven
Drucker, Danyel Fisher, Bongshin Lee, Jacob O
Wobbrock, and others. 2014. Reducing legacy bias in
gesture elicitation studies. *interactions* 21, 3 (2014),
40–45.

[40] Meredith Ringel Morris, Jacob O Wobbrock, and
Andrew D Wilson. 2010. Understanding Users'
Preferences for Surface Gestures. In *Proceedings of
Graphics Interface 2010 (GI '10).* Canadian Information
Processing Society, Toronto, Ont., Canada, Canada,
261–268.
`http://dl.acm.org/citation.cfm?id=1839214.1839260`

[41] Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft. 2014. Human-data interaction: The human face of the data-driven society. *Available at SSRN 2508051* 000, 00, Article 00 (2014), 00 pages.

[42] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking glass: a field study on noticing interactivity of a shop window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, ACM, USA, 297–306.

[43] Matteo Munaro, Filippo Basso, and Emanuele Menegatti. 2016. OpenPTrack: Open source multi-camera calibration and people tracking for RGB-D camera networks. *Robotics and Autonomous Systems* 75 (2016), 525–538.

[44] T. Ojala, V. Kostakos, H. Kukka, T. Heikkinen, T. Linden, M. Jurmu, S. Hosio, F. Kruger, and D. Zanni. 2012. Multipurpose Interactive Public Displays in the Wild: Three Years Later. *Computer* 45, 5 (May 2012), 42–49. DOI:http://dx.doi.org/10.1109/MC.2012.115

[45] Scott Pattison, Josh Gutwill, Ryan Auster, and Mac Cannady. 2019. Experimental and Quasi-Experimental Designs in Visitor Studies: A Critical Reflection on Three Projects. *Visitor Studies* 22, 1 (2019), 43–66. DOI:http://dx.doi.org/10.1080/10645578.2019.1605235

[46] Deborah L Perry. 2012. *What makes learning fun?: principles for the design of intrinsically motivating museum exhibits*. Rowman Altamira, USA.

[47] Thorsten Prante, Carsten Röcker, Norbert Streitz, Richard Stenzel, Carsten Magerkurth, Daniel Van Alphen, and Daniela Plewe. 2003. Hello. wall–beyond ambient displays. In *Adjunct Proceedings of Ubicomp*, Vol. 2003. USA, USA, 277–278.

[48] Samantha Punch. 2002. Research with children: the same or different from research with adults? *Childhood* 9, 3 (2002), 321–341.

[49] Jessica Roberts, Amartya Banerjee, Annette Hong, Steven McGee, Michael Horn, and Matt Matcuk. 2018. Digital Exhibit Labels in Museums: Promoting Visitor Engagement with Cultural Artifacts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, ACM, USA, 623.

[50] Jessica Roberts, Leilah Lyons, Francesco Cafaro, and Rebecca Eydt. 2014. Interpreting Data from Within: Supporting Humandata Interaction in Museum Exhibits Through Perspective Taking. In *Proceedings of the 2014 Conference on Interaction Design and Children (IDC '14)*. ACM, New York, NY, USA, 7–16. DOI:http://dx.doi.org/10.1145/2593968.2593974

[51] Elisa Rubegni, Vito Gentile, Alessio Malizia, Salvatore Sorce, and Niko Kargas. 2019. Child-display interaction: exploring avatar-based touchless gestural interfaces. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays*. ACM, ACM, USA, 23.

[52] Jaime Ruiz and Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3347–3350. DOI:http://dx.doi.org/10.1145/2702123.2702583

[53] Cody Sandifer. 1997. Time-based behaviors at an interactive science museum: Exploring the differences between weekday/weekend and family/nonfamily visitors. *Science Education* 81, 6 (1997), 689–701.

[54] Martin Tomitsch, Christopher Ackad, Oliver Dawson, Luke Hespanhol, and Judy Kay. 2014. Who Cares about the Content? An Analysis of Playful Behaviour at a Public Display. In *Proceedings of The International Symposium on Pervasive Displays (PerDis âĂŹ14)*. Association for Computing Machinery, New York, NY, USA, 160âĂŞ165. DOI:http://dx.doi.org/10.1145/2611009.2611016

[55] Milka Trajkova, Francesco Cafaro, and Lynn Dombrowski. 2019. Designing for Ballet Classes: Identifying and Mitigating Communication Challenges Between Dancers and Teachers. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. ACM, ACM, USA, 265–277.

[56] Eliane Zambon Victorelli, Julio Cesar Dos Reis, Heiko Hornung, and Alysson Bolognesi Prado. 2019. Understanding human-data interaction: Literature review and recommendations for design. *International Journal of Human-Computer Studies* 000, 00, Article 00 (2019), 00 pages.

[57] Gualtiero Volpe, Giovanna Varni, Anna Rita Addessi, and Barbara Mazzarino. 2012. BeSound: Embodied reflexion for music education in childhood. In *Proceedings of the 11th International Conference on Interaction Design and Children*. ACM, ACM, USA, 172–175.

[58] Miaosen Wang, Sebastian Boring, and Saul Greenberg. 2012. Proxemic Peddler: A Public Advertising Display That Captures and Preserves the Attention of a Passerby. In *Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12)*. ACM, New York, NY, USA, Article 3, 6 pages. DOI:http://dx.doi.org/10.1145/2307798.2307801

[59] Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. 2005. Maximizing the guessability of symbolic input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1869–1872. DOI:http://dx.doi.org/10.1145/1056808.1057043

[60] Yitan Zhu, Huai Li, David J Miller, Zuyi Wang, Jianhua Xuan, Robert Clarke, Eric P Hoffman, and Yue Wang. 2008. caBIG VISDA: Modeling, visualization, and discovery for cluster analysis of genomic data. *BMC bioinformatics* 9, 1 (2008), 383.