Lyons, L., & Mallavarapu, A. (2021). Collective Usability: Using Simulation Tools to Explore Embodied Design Challenges in Immersive, Shared Mixed-Reality Experiences. *Educational Technology & Society*, *24* (2), 120–135.

Collective Usability: Using Simulation Tools to Explore Embodied Design Challenges in Immersive, Shared Mixed-Reality Experiences

Leilah Lyons[1,2*] and Aditi Mallavarapu[2]
[1]New York Hall of Science, New York, USA // [2]University of Illinois at Chicago, Illinois, USA //
llyons@nysci.org // amalla5@uic.edu
[*]Corresponding author

ABSTRACT: In this paper we define the concept of collective usability, a complex systems perspective on usability that positions an entire group, not an individual, as the unit of analysis. Shared XR experiences have inherent temporal and spatial properties that produce emergent, collective impacts which can impede learners' engagement. Assembling large groups of users to test multiple design configurations is both logistically and financially impractical, however. We demonstrate the practical value of exploring the design space of an XR experience with a simple observation-informed Agent-Based Model. We used the model to explore how changes in the number of simultaneous users, and in the size, placement, and interaction duration of the proffered interactives, could affect collective access to a large-scale, mixed-reality, multi-user museum exhibit. (Collective access, an element of collective usability, is the degree to which users can gain access to each of the different interactives.) With this simple model, we explored (1) how the bottom-up propagation of individual-level design properties can affect collective outcomes, as when certain interactives' linger times cause a bottleneck, and (2) how the top-down propagation of collective design constraints can be used to guide individual-level design, as when we determined thresholds for the "stickiness" and "repeat allure" of an interactive to improve collective access. The final design of the exhibit implemented many of the design guidelines uncovered by the model. We argue that collective usability models could be useful for addressing a range of collective usability issues, beyond collective access, for temporally and spatially sensitive XR learning environments.

Keywords: Collective usability, Agent-based models, Usability methods, Informal learning, Shared XR

1. Introduction

Improved technologies are allowing designers to realize the dream of creating immersive, multi-user experiences akin to science fiction "holodecks." The embodied nature of XR experiences introduce new design challenges, as such systems are acutely sensitive to both temporal and spatial relationships between people and interactive objects or areas. We argue that in addition to traditional individual-level usability concerns, designers of multi-user XR experiences need to design for collective usability, a term we define as the degree to which a *group* of simultaneous users can make use of an interactive experience where the human-computer and human-human interactions combine to form a complex system. An individual user is the unit of analysis in traditional usability, whereas a group of users is the unit of analysis for collective usability. Where traditional usability gives primacy to the interaction between an individual human and a computer, collective usability embraces a complex systems perspective, wherein the individual interactions between humans and computers (and between humans and humans) combine to generate emergent effects that can only be measured at the group level. Attending to collective usability is important because designs with high individual usability (like a public transit ticket machine that ensures users pay the correct amount for their destination) may have inadvertently poor collective usability (e.g., a long line forms). Collective usability is especially critical for ensuring equitable learning environments, where all learners should get a chance to meaningfully engage with the proffered experiences.

The goal of this paper is to describe some of the challenges faced and the analysis techniques used when designing a large-scale Mixed Reality (MR) experience for a museum. For this study we videotaped 5 hours of more than 100 visitors' interactions with an early pilot exhibit, which consisted of six interactives that each supported a different type of full-body interaction (some gesture-based, some tangible-based). We used this data to empirically compute the statistical distributions for visitor engagement with these six interactives. We used these distributions to develop an Agent-Based Model (ABM) simulation to explore how changes in the number of simultaneous users and in the duration, size, and number of the proffered interactives could affect collective usability. We were specifically interested in a form of collective usability we dub collective access: the degree to which visitors could try each of the different interactives. We propose that collective usability models, of which our ABM is an example, could be used to represent a wide range of embodied usability challenges and thus become part of the toolkit of XR designers.

2. Background and prior work

2.1. Modeling users to support design processes

Using model-based simulations to support usability design and evaluation is not new — as early as the 1980s the GOMS model was used to engineer the usability of single-user interactives, by accounting for how human cognitive performance impacts the execution of software tasks (Card, Moran, & Newell, 1983). These "engineering models" should be "deliberately approximate … [including] just the level of detail necessary to do the design job … while keeping the modeling effort tractable" (John & Kieras, 1996, p. 4).

A number of researchers have explored how to model the behavior of users in collaborative scenarios to support the design process (Ivory & Hearst, 2001). For example, extending UML (a modeling language used by software developers) to describe group processes within design specifications (Garrido & Gea, 2001), or embedding the relational roles between group members into task models (Herrmann et al., 2004). The problem with many of these models is that, like GOMS, they assume a task model — i.e., that users will be engaged in specified tasks while working together. Unfortunately, there is often a gap between the intended tasks and how group interaction actually unfolds, suggesting models need to be based on how people actually act, not on idealized task execution sequences (Poltrock & Handel, 2010).

XR Learning Environments (LEs) tend to support open-ended use, so it is especially critical for collective usability models to be based on observations of how learners actually behave. We argue that for most XR LEs, minutely modeling learner cognition is not necessary for creating an engineering model. All that is needed is a minimal representation commonly seen behaviors, like the "foraging" behaviors of visitors in museums. It *is* important, however, to model elements of the shared domain context that can afford or constrain users' interactions with the system or their human companions (Wurdel, 2009). Because the interaction in XR LEs is stretched across a real or imagined physical setting, a modeling approach is needed that takes into account physical aspects like the spatial arrangement of interaction opportunities and other users.

2.2. Model-based methods in multi-user system design

Multi-user system researchers also use models to inform design, albeit in the form of highly formalized models used to validate competing system designs. They too have acknowledged the cyclical nature of the usability of multi-user systems, wherein "not only will the individual behavior of a user be affected by changes to the system through its collective use, but the system can also have an effect on the collective behavior of the users" (Massink et al., 2008). Researchers have predominantly used models to examine timing complications, making use of an array of formal models augmented with timing information and constraints — like finite automata, specialized algebras, and petri nets — to validate their designs (Bernardo & Corradini, 2004). Because most researchers were validating software like shared file management systems (e.g., Massink et al., 2008), the physical or spatial aspects of shared interaction were not modeled. One exception is a research group that represented spatial locations within a building to model a building's dynamic signage system (Harrison et al., 2008). They later used an ABM elaborated with force calculations to validate how the dynamic signage system performed against a static signage during evacuations (Langner & Kray, 2014).

Our work differs from this specific approach, and from the work on multi-user systems in general, in that we use an ABM as a *design tool* for collective usability. As a design tool, we employ ABM to explore the design space and to derive creative insights and guidelines to drive further design processes, not to formally validate designs.

2.3. Design challenges of full-body XR Learning Environments (LEs)

Museums and researchers exploring classrooms of the future are increasingly interested in the educational potential of sensing technologies that allow learners to use their bodies to interact with XR LEs. There are many rationales for using full-body movements to support learning activities. Full-body interactions can be designed to capitalize on *embodiment* by activating the proprioceptive aspects of cognition; they can take advantage of how cognition is *embedded* in the physical and social structures of the environment; or they can *extend* beyond individual organisms to support distributed cognition (Roberts & Lyons, 2017). One exhibit that relies on the *embodiment* aspect of whole-body interaction asks a visitor to control the path of a simulated meteor by running at different angles and speeds, using body sensations to build an understanding of force and motion (Lindgren, Tscholl, Wang, & Johnson, 2016). Many museums have developed camera-based XR interactives where visitors jump or dance, where usability designers need to design for how performances are *embedded* in the audience's spectator experience (Peltonen et al., 2008; Reeves, Benford, O'Malley, & Fraser, 2005). Sometimes *embedding* whole-body interactions within a spectator experience leads to learners coordinating and coaching one another, which *extends* the interaction into the realm of distributed cognition, as with an interactive slide (Malinverni, Ackermann, & Pares, 2016) or a classroom game about friction (Enyedy, Danish, & DeLiema, 2015).

The exhibit in this work employs full-body interactions to *extend* learner cognition to embrace shared management of a dynamic complex system. *Extended* full-body interaction design poses special usability challenges in the form of needing to manage potential interaction effects between different users. In prior work that *extended* cognition through whole body interactions, designers put the onus of managing interaction effects back on the learners — making learners responsible for monitoring one another and ensuring their actions don't negatively impact the collective experience. We argue that as XR experiences get more complex, designers need tools to examine the design challenges that arise from multiple users' actions, and to experiment with how different designs might impact those interaction effects.

2.4. Agent-based models

The review of prior work shows that there is a need for an "engineering model" (John & Kieras, 1996) to support design processes for XR LEs. Specifically, a modeling approach is needed that allows both spatial and temporal aspects of the design space to be represented. These requirements led us to use an ABM, defined as "(1) a computational method that enables a researcher to create, analyze, and (2) experiment with (3) models composed of (4) agents that interact within (5) the environment" (Gilbert, 2020). ABMs model a complex system by defining basic scenario properties (like the placement of food sources near an anthill) and the rule-based behaviors of independent agents (like the pheromone-governed movements of ants), to represent the complex behaviors that emerge (like coordinated food gathering). ABMs have been used to model a wide range of scientific and social collective activities, like traffic flows, military scenarios, movements of customers in stores, supply chain logistics, and administrative workflows (Abar et al., 2017). To the best of our knowledge, however, ABMs have never been promoted as a tool for usability design, apart from their use to design the behavior of artificially intelligent agents for ubiquitous computing (e.g., Mangina et al., 2010).

3. Design of pilot exhibit

The New York Hall of Science received a grant to develop a collaborative MR exhibit to teach visitors about sustainability and complex systems thinking in the context of a human-natural system. The exhibit needed to serve all ages, but the learning goals placed a special focus on middle school–age visitors. Complex systems thinking is challenging, and middle-school learners especially struggle with identifying the causal relationships between system components (Grotzer, 2012; Goldstone, 2006, Hmelo-Silver et al., 2007). One way to help learners develop a sense of a system's causal mechanisms is to have them take on first-person perspectives of different system agents (Jacobson & Wilensky, 2006). For this reason, it was decided to allow visitors to try out the roles of multiple agents within the system, allowing them to develop a firsthand appreciation for why, say, both farmers and city residents might draw down a water supply without realizing their actions could impact others. To encourage learners to try out multiple roles and thus broaden their understanding of the system, each role needed to be executed via an interactive that didn't require too much linger time. Moreover, because museums are free-choice learning environments, the operation of the interactive components could not be tightly coupled, i.e., *require* ongoing tight coordination between visitors (Salvador et al., 1996), although visitors could certainly choose to coordinate their actions. The design firm, DesignIO, created an initial functional prototype to explore some of these design issues.

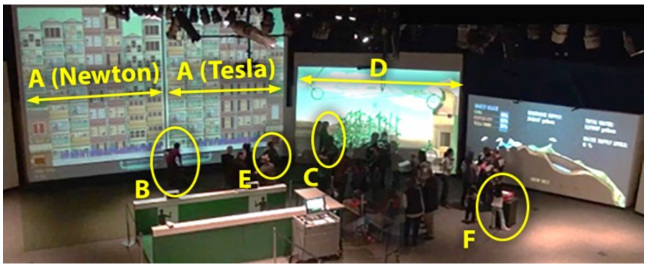

Figure 1. Pilot exhibit. Whole-body interactives are marked with letters and described in Table 1

The simulation consisted of two cities, Newton and Tesla (left screen in Figure 1), a farm where visitors could raise corn crops (middle screen in Figure 1) and a water-source selection screen (the screen on the far right in Figure 1). Water was shared among the farm and the two cities, meaning that if visitors didn't carefully manage urban growth there would be a water crisis, and subsequently a food crisis. The interactives available to individual users are illustrated in Figure 2 and described in Table 1.

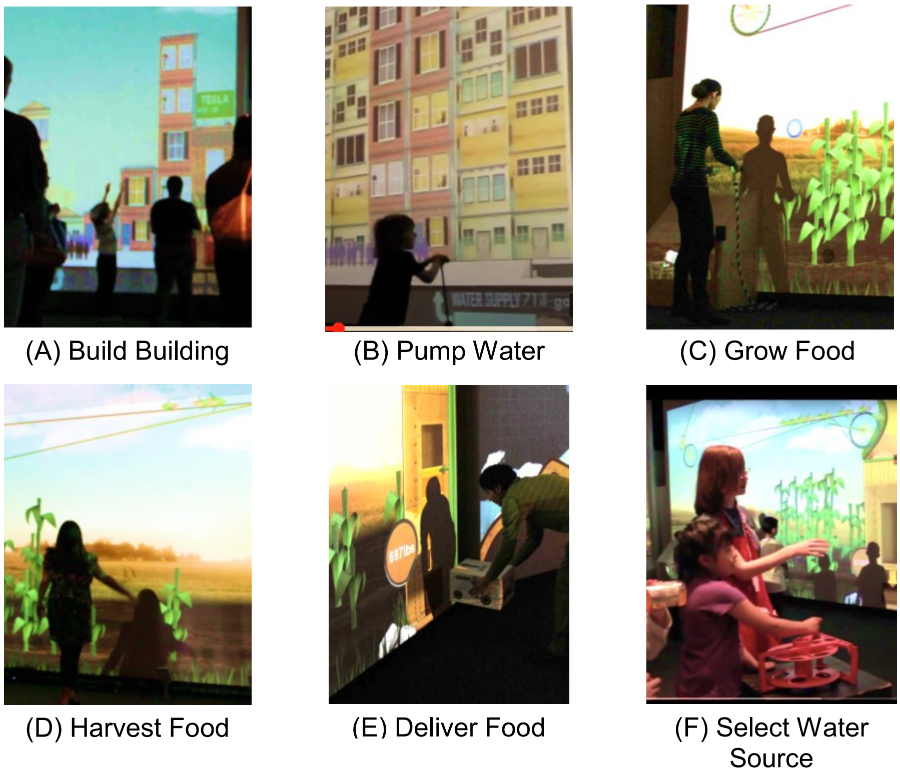

Figure 2. Snapshots of the six full-body interactives in the pilot exhibit

Table 1. Interactives available in the exhibit pilot

Action	Site of action	Max users	Interdependencies	Description of action enactment
(A) Build Building	Newton/ Tesla screen	7/7	Increases need for (E) Increases need for (B) Competes with other city	Visitor raises hands straight up above head to add another story to a building
(B) Pump Water	Pump near Newton/ Tesla screen	1	Affected by (F) Competes with (C)	Visitor operates a bicycle pump to pump water from the selected water source to both cities
(C) Grow Food	Hose in front of Farm screen	1	Affected by (F) Permits (D) Competes with (B)	Visitor uses a sprayer hose to irrigate farmlands (blue circle on screen indicates spray target), which sprouts corn
(D) Harvest Food	Farm screen	8	Requires (C) Permits (E)	Visitor uses a swiping arm motion to harvest mature corn, which gets deposited in silo
(E) Deliver Food	Between Farm and Newton/ Tesla screens	1	Requires (D) Choice in delivery city can induce competition	Visitor uses a cardboard "truck" box to pick up food from the Farm silo and deliver it to a depot in either Newton or Tesla
(F) Select Water Source	Rotary dial in front of Water Source screen	1	Permits (B) Permits (C) Is affected by (B) and (C)	Visitor rotates a dial, which sets the source of water used in the farm and the cities (e.g., aquifer, captured rainwater, river, etc.). When drained a new source must be chosen.

4. Methods

The purpose of this research was to explore how different design decisions might affect visitors' collective access. Collective access, the degree to which visitors could try each of the interactives, was important because allowing visitors to try on a breadth of different roles was theorized to support the exhibit's learning goal of encouraging systems thinking. To build such a model, we made the simplifying assumption to represent a visitor's engagement with just a single metric: linger time, based on distributions of how long real visitors would use a given interactive.

4.1. Participants

Participants were a mix of regular daily visitors to the museum who were told of the testing sessions at admission and/or saw signs outside the test gallery, and museum members who had received an email advertising the pilot testing sessions. The test sessions were held on several weekdays during spring break, so local families were the primary attendees. The apparent age ranges and cultural backgrounds of attendees were in keeping with typical visitorship for our institution, which tends to skew slightly young (average age of children being around 12) and which tends to be quite diverse. We used an implied consent procedure approved by our IRB, wherein a lollipop sign indicated that entering the test space constituted consent to being videotaped. Because collective usability takes the group as the unit of analysis, we did not collect information from individual visitors. We thus did not collect statistics on the number of visitors who used the exhibit, but there seemed to be about 20 visitors present at any given time, and both museum attendance data and the staff observations suggest that more than 100 individual visitors experienced the exhibit during our data collection.

4.2. Collecting visitor linger time statistics

The pilot exhibit was installed in a temporary gallery in the museum for several days, allowing us to videotape 4 hours and 58 minutes of visitors' interactions with the exhibit from a gallery on the rear wall (see Figure 1). Six research personnel coded the videos, noting when a visitor began using one of the interactives in Table 1, and when they stopped, capturing these start/stop events as timestamps. Using those timestamps, we then computed the average and standard deviation of visitor linger time (see Table 2). We did not compute inter-coder reliability, as the events were very obvious.

Table 2. Observations of visitors' interaction linger times

Interactive	Number of observations	Mean linger time, in seconds	Standard deviation
(F) Select water source	$n = 283$	$M = 7$	$SD = 9$
(A) Build building (Tesla)	$n = 103$	$M = 13$	$SD = 13$
(A) Build building (Newton)	$n = 83$	$M = 16$	$SD = 19$
(D) Harvest food	$n = 642$	$M = 20$	$SD = 28$
(C) Grow food	$n = 528$	$M = 21$	$SD = 33$
(B) Pump water	$n = 165$	$M = 42$	$SD = 53$
(E) Deliver food	$n = 132$	$M = 120$	$SD = 100$

These results show that the "stickiest" activity is food delivery (E) — visitors would run back and forth between the Farm silo and the Newton and Tesla food depot points for an average of 120 seconds before relinquishing the truck to another user. The least sticky activity is the water source selection wheel (F) — which makes sense, since the action is the closest one to being atomic (meaning that once the user chooses a water source, they are done, like clicking a button).

4.3. Simulating visitor use of exhibit interactives

The informal qualitative observations of the designers and researchers found that the pilot exhibit did not support collective access well, with numerous instances of people waiting around interactives to get a turn. This left a number of questions concerning how the exhibit could be modified to ensure that visitors could get a chance to try each of the interactives. We thus created an ABM to explore design variations virtually, without the cost and logistical overhead of revising and testing the prototype.

4.1.1. Building a simulation of the pilot exhibit in NetLogo

We selected NetLogo (Wilensky, 1999) to build our model. Originally developed for educational purposes, NetLogo has since been adopted by a number of professionals like urban planners for its convenience in quickly roughing out models of the intersection of people and places. We measured the screen sizes and angles and location of the tangible interactives to reproduce the pilot exhibit space (see Figure 3). For the tangible-based, single-user interactives (B, C, E, F), we set up rules to prevent other agents from using them if they were already in use. For the wall-based interactives (A, D), we used the maximum observed number of simultaneous visitors to determine the minimum wall space occupied per user, 1.875 feet, and used that to cap the number of simultaneous agents. Apart from (D), the interactives were designed to permit visitors to operate them regardless of their interdependencies, so we opted not to model the system's interdependencies. This simplification prevented us from going down the slippery slope towards fully modeling how visitors would reason about and respond to system interdependencies, which would be useful for modeling visitor learning, but not needed to inspect access issues. Such parsimony is recognized as good practice when constructing ABM (Gilbert, 2020).

4.1.2. Modeling agent behaviors

It is considered good practice to use the simplest possible representations of how agents behave to better explore emergent effects (Gilbert, 2020). We thus used Brownian motion as the simplest possible model of how visitors move within an exhibit space. With Brownian motion, agents are assigned a random trajectory they will follow until colliding with another agent or a physical object. Upon collision they get assigned a new random trajectory or (if the object is an available interactive) they will stop and interact. But as a check on the robustness of our results, we implemented two other plausible simplifications of how visitors could choose to move in the space: a Nearest model, where idle agents select as a target the nearest unoccupied interactive, and an Interest model, where idle agents select the unoccupied interactive that they have had the least experience with, with ties being broken by the nearest interactive. Regardless of the movement model, once an idle agent reaches an unoccupied interactive, they would occupy it for a duration of time drawn from the normal distribution for that interactive (gathered from observations of real users, see Table 2).

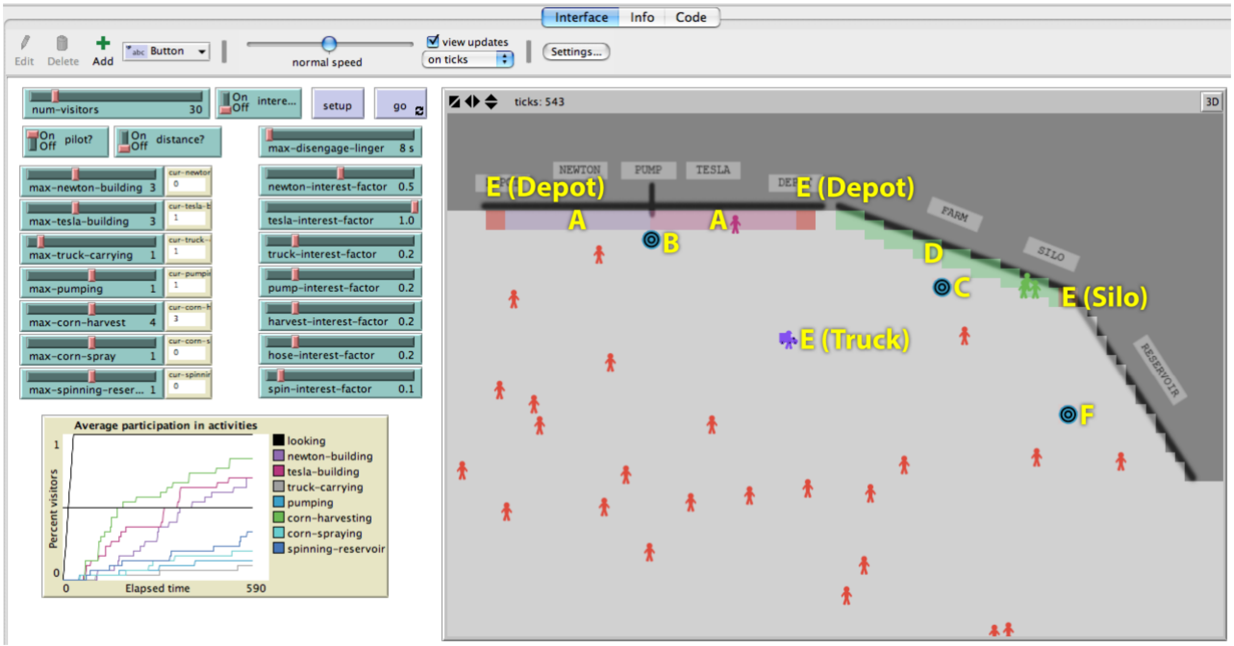

Figure 3. Agent-based simulation of visitors' participation in the Connected Worlds pilot. Colored squares mark where visitors can engage with an interactive

The Nearest movement model accounts for visitor "laziness" — the interaction space is quite large, and so visitors are likely to interact with a closer interactive if it is available. The Interest movement model represents the effect of novelty on visitor choices - all else being equal, visitors do prefer to try new or under-explored interactives. With the Interest model, for each agent, all interactives are initially assigned the same interest level of 1. The agent's interest level for each interactive is decreased by a fixed amount after using it. Visitors preferentially select available interactives with the highest remaining interest levels, with ties broken by selecting the closest interactive. This is a highly simplified analogue to more sophisticated value-cost models of visitor movements (Bitgood, 2006). We used the expert judgment of our research team, a mix of researchers and practitioners with extensive experience observing visitors, to select an interest decrement that produced plausible simulated visitor movements, eventually settling on a 10% decrease in interest. (Our choices in movement models, and implications for how much attention researchers should give to validating same, are tackled in the Conclusion).

5. Simulation results

The goal of this simulation-based exploration was to reduce the amount of time it would take for visitors to gain exposure to the majority of the exhibit interactives, i.e., to improve collective access. We thus used as our outcome the "half-used" metric: the time it would take at least 50% of the visitors to use each of the interactives. To generate half-used metrics we used a Monte Carlo approach. For each experimental configuration of the simulation model, we ran it 100 times, and averaged the results together. Each model would run until reaching the half-used state, or until 60 minutes of elapsed in-simulation time (60 minutes is well above the time visitors would be permitted to stay at the exhibit).

5.1. Baseline scenario under varied attendance

We first established a baseline simulation for 10, 20, 30, and 40 visitors to examine how the exhibit scaled with more users, and ran it under all three movement models to check for inconsistencies. Figure 4 shows that while the Nearest and Interest models do produce lower half-used times (which makes sense — there would be less aimless wandering), all three movement models produce linear regressions with similar slopes, suggesting that for each additional visitor, another 30-60 seconds of delay can be expected to be added to the half-used metric. This monotonic consistency is reassuring, suggesting that while in real life, visitors may exhibit a mix of movement characteristics (sometimes wandering, sometimes seizing a near interactive, sometimes seeking out a desired interactive), the collective access (as measured by the half-used metric) will fall within clear bounds.

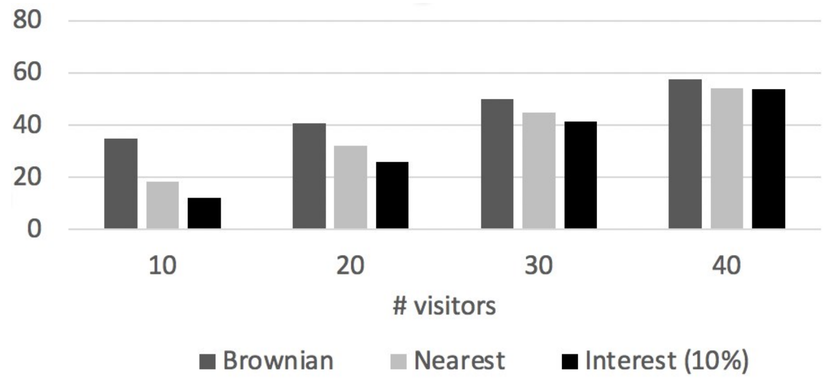

Figure 4. Time to reach the half-used state for all interactives under three different movement models, averaged across 100 Monte Carlo runs for each model and each number of visitors

5.2. Finding the source of half-used delays and extrapolating design variations to test

Our next step was to discover the major sources of delay — to determine this, we computed half-used metrics for each of the 7 possible interactives, under four different numbers of visitors (10, 20, 30, and 40) and the three different movement models (see Figure 5). The first thing to notice is that the overarching trends (in *magnitude* of half-used time, and in the *growth* in half-used times as the number of visitors increase) are consistent regardless of movement model, again suggesting that for our metric and problem space, the way we model visitor movement patterns is not a critical factor. The only major differences are that under the Interest model, the (F) Select Water Source time is reduced (because its low linger time allowed interested visitors easier access), and under the Nearest model, the (D) Harvest Food time is increased (because the farm is central to the exhibit, so it is often the nearest interactive and thus under more demand).

One strong pattern to notice is that the three screen-based activities that permit simultaneous users, (A) Newton Building, (A) Tesla Building, and (D) Harvest Food, do not show an increase in time to reach the half-used state, even as the number of visitors grows. At a collective level, it is this flat growth that we are seeking, as it shows that the interactives can flexibly scale to support larger groups. The screen sizes in our pilot were dictated by the gallery wall sizes, however, which were quite large — would we need to make the screens just as large in the final exhibit to preserve this scalability? This is a factor we wanted to explore further with our simulation.

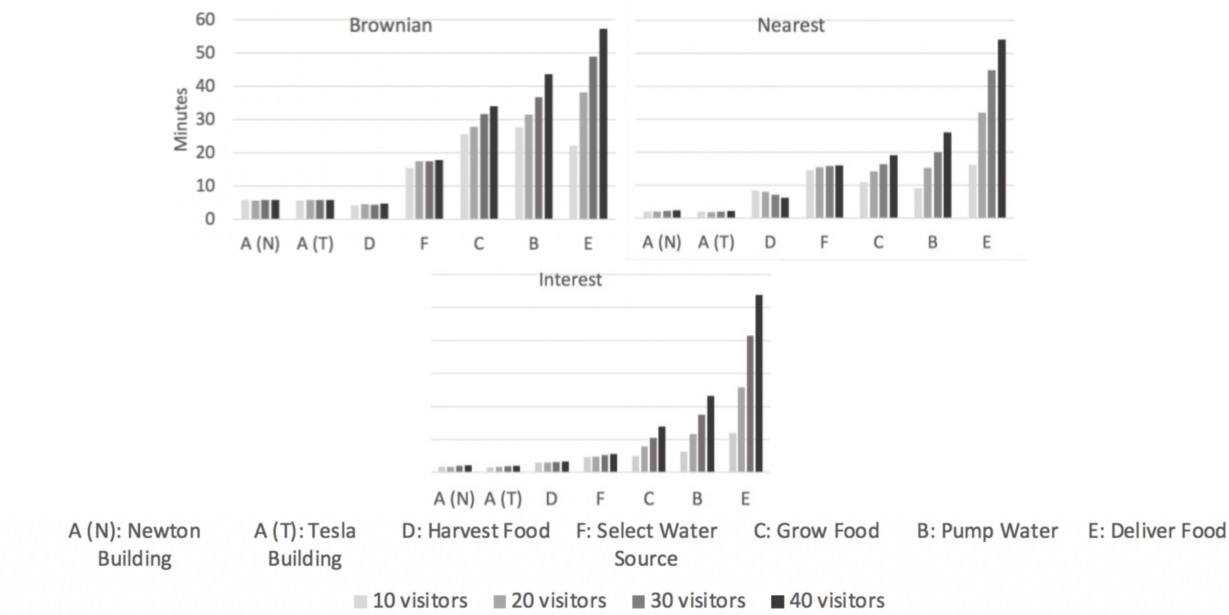

Figure 5. Time to reach the half-used state, in minutes, for each of the interactives as the number of visitors increases, under each of the movement models

The single-user tangible interactives, on the other hand, all show pronounced growth in half-used delay with increasing numbers of visitors, with one notable exception, the (F) Select Water Source activity. What differentiates the (F) from the other tangible interactives is its linger time (see Figure 6), which is closest to an atomic action (an instantaneous action-effect pairing, like a button click). This suggests that reducing the "stickiness" of individual tangible interactives (i.e., the amount of time a user will want to linger while operating it) is a design variation to explore for improving the collective user experience. However, it may be impossible to reduce the linger time of certain tangible interactives below a certain point — for example, carrying the food truck between the farm and the cities incurs a minimum linger time. Other alternatives to explore for improving collective access are replicating a tangible interactive with a long linger time (e.g., supplying multiple delivery truck interactives), or reducing how much a user would want to use an interactive again ("repeat allure").

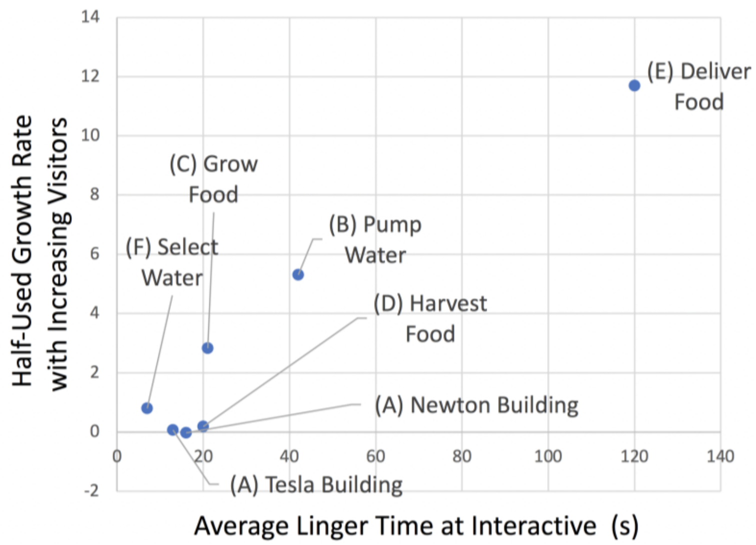

Figure 6. The relationship between the average linger time of each interactive, and the growth in half-used time as the number of visitors increases

5.3. Testing design variations with the ABM

The exploration of delay sources suggested that we should test four types of design variations for their impact on collective access: (1) changing the screen size for screen-based interactives, (2) decreasing the stickiness (i.e., linger time) of tangible interactives, (3) replicating tangible interactives, and (4) decreasing the "repeat allure" of the tangible interactives. We tested these parameters with all three movement models. All models produced similar results, so we present only the Nearest results for succinctness (the Nearest outcomes fell between the Brownian and Interest outcomes).

5.3.1. Manipulating screen width under varied attendance

The baseline results showed that the multi-user widescreen interactives were least impacted by growing numbers of visitors, so we wanted to see how important it was to use screens that were at least 13 feet wide. Would making the screens smaller hurt access, or would making them even larger benefit collective access? We wanted to hold as much else constant about our model as possible, including the relative positions of the interaction opportunities, so we opted only to manipulate the sizes of the Tesla Building and Newton Building screens (manipulating the farm screen width would have changed the entire layout). The Narrow Screen condition was half the size of the screen used in the baseline (6.5 feet), and the Wide Screen was 1.5 times the baseline (19.5 feet). We can see from Figure 7 that increasing the screen size does not appreciably impact the half-used time, while decreasing the screen size does remove its ability to scale with the number of visitors.

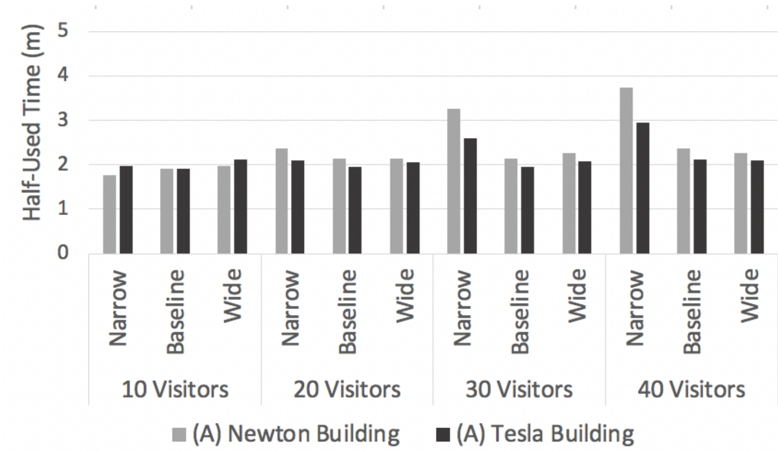

Figure 7. The impact of varying the screen size on the half-used time, in minutes, as the number of visitors increases

5.3.2. Reducing the "stickiness" of single-user tangible interactives

The baseline results showed that the single-user tangible interactives scaled least well when increasing the number of visitors, with the exception of the low-linger (F) Select Water Source wheel. Would decreasing the linger time of the other single-user interactives improve collective access? No designer wants to intentionally decrease the pleasurability of an interaction design, but it may be the case that the collective good demands it. There is research evidence that when visitors get highly engaged in single-user interactions within a shared exhibit, it negatively impacts the group experience (Lyons, 2009).

We reduced the linger time to a fixed 5 seconds (approximately the time needed to approach and execute an atomic action like a button click) to see how low stickiness would impact the half-used time. We implemented atomic versions of the (C) hose and the (B) pump (we did not test an atomic version of the (E) truck because its linger time depends on the distance between the farm's silo and the food delivery depots at Newton and Tesla). Figure 8 shows that if both interactives could be made less sticky, we could attain the flat growth with increased visitors as was seen for the (F) wheel. The (B) Pump Water interactive benefitted the most from a conversion to an atomic action — likely because its less-central location made it less well-trafficked, and thus more available for visitors.

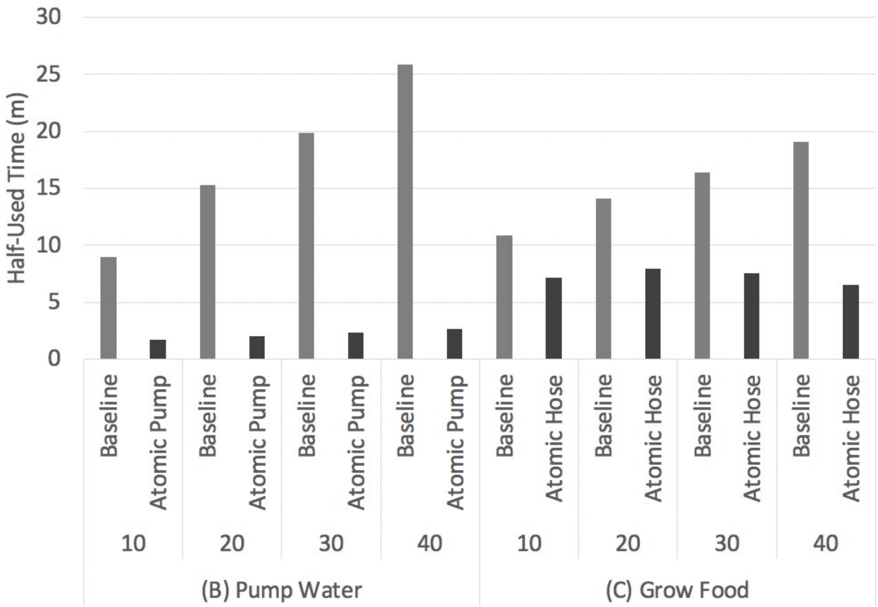

Figure 8. The impact on the half-used time, in minutes, of assigning a tangible interactive to have a fixed, atomic amount of linger time, as the number of visitors increases

5.3.3. The effect of replicating single-user tangible interactives

We replicated the three interactives with the largest growths in half-used time (B, C, and E) and present the results in Figure 9. An extra hose reduces the growth in (C) half-used time to be nearly flat, although the magnitude of reduction was less than what was seen when the hose was converted to an atomic action (Figure 8). While the rate of growth in half-used time remains high for the replicated (B) water pump and (E) truck, the *magnitude* of the half-used times decreased by around 50% as compared to baseline, regardless of visitor count. For tangible interactives with long linger times, it may be preferable to duplicate them than to remove stickiness, as removing stickiness may well involve removing the pleasurable aspects of interaction. However, a single replication may not be enough to meaningfully impact collective access — for example, the half-used time triples from a 10-visitor replicated truck scenario to a 40-visitor replicated truck scenario. This raises the question: could more replicated trucks attain a half-used measure that doesn't vary with increased numbers of visitors?

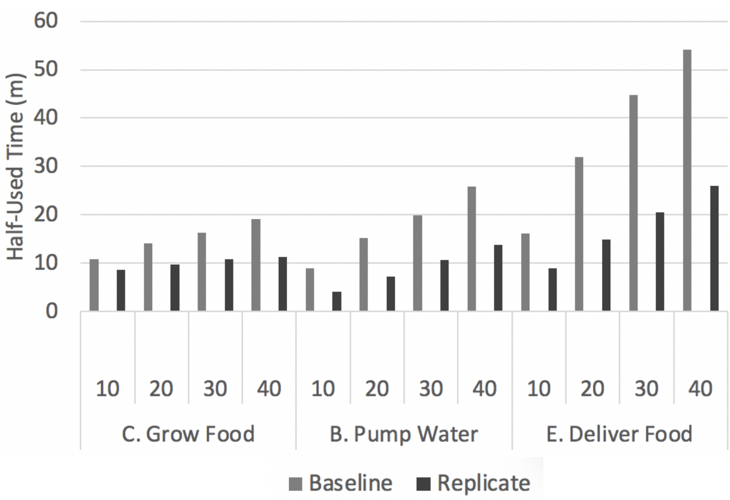

Figure 9. The impact on the half-used time, in minutes, of replicating the tangible interactives as the number of visitors increases

The results show that up to 8 trucks need to be present before the (E) ceases to account for more than 50% of the half-used time for the exhibit as a whole (see Figure 10). The half-used growth rate still doesn't flatten, however. Could it be that because the trucks are located in a highly-trafficked area, the same visitors end up picking up the trucks again and again, making it harder for other visitors to use them? The next section explores making that kind of repeat-use less appealing.

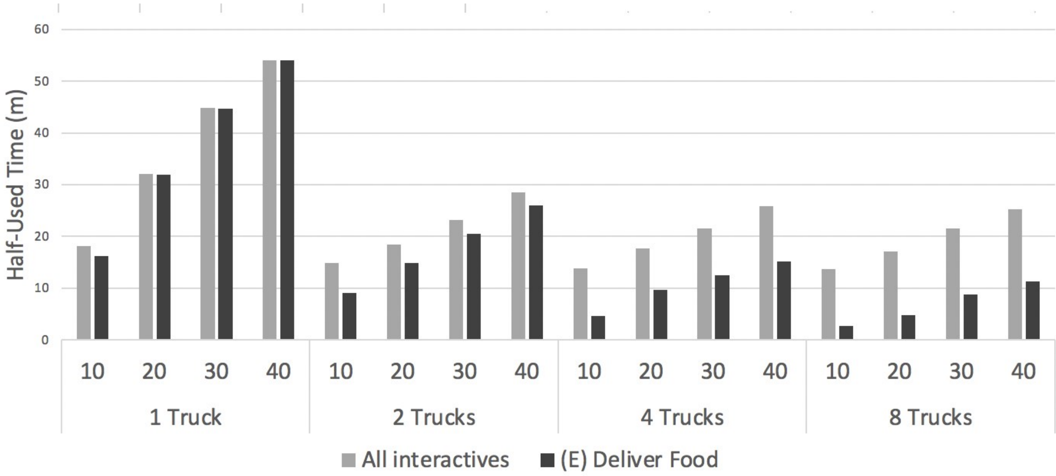

Figure 10. Time to reach the half-used state for all interactives and for the (E) Deliver Food interactive, in minutes, as the number of trucks and the number of visitors increases

5.3.4. Reducing the "repeat allure" of food delivery

Much like reducing "stickiness" by making interactives less engaging, designers can also make it less likely for users to opt to return to an interactive. Designers do not need to make the experience overtly unpleasant, but can instead think about ways to build in gentle impediments to make users less likely to immediately re-engage. With tangible user interfaces there are an array of ways of negatively manipulating their affordances, from literally adding friction (like making a wheel hard to turn) to adding weight (visitors might avoid picking up a heavy object again unless necessary).

We simulated a "repeat allure" decrease by manipulating the Interest-based movement model. In addition to the 10% drop used as the default for the Interest model, we experimented with three different degrees of decreasing the repeat allure for the Truck — a 25% drop, a 50% drop, and an extreme 100% drop in interest, compared to the Nearest baseline (see Figure 11). Decreasing the repeat allure does decrease the half-used magnitude, but still does not flatten the growth in half-used time. This suggests that while designers can incorporate negative affordances into the design of interaction opportunities that have unavoidably long linger times (e.g., tangible interactives that need to be carried from place to place), this alone may not be enough to scale with increased users.

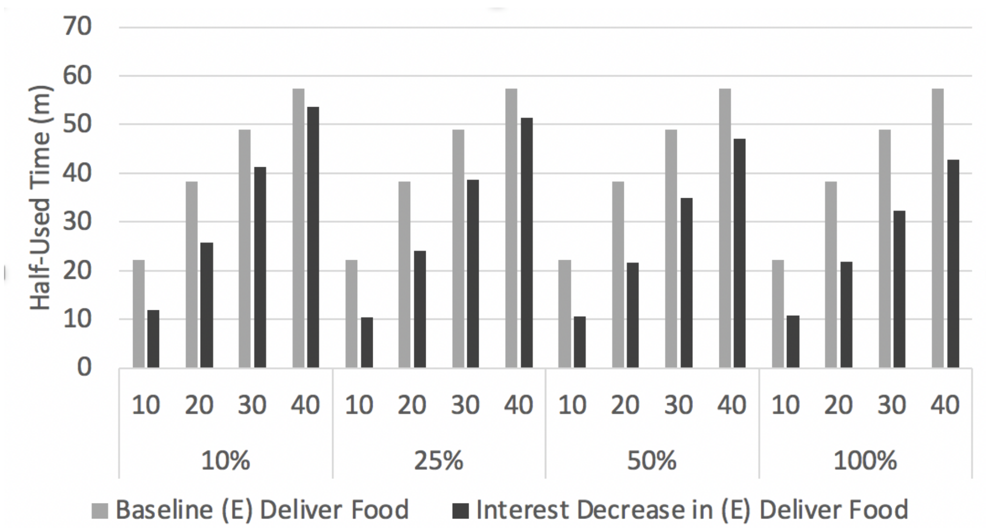

Figure 11. Time, in minutes, to reach the half-used state for the Food Delivery interaction type when visitor interest in using the Truck drops by growing percentages, as the number of visitors increases

5.4. Putting it all together — the kitchen sink simulation

One value of using ABMs for exploring design spaces is that they can reveal nonlinearities when multiple design variations are combined. In other words, sometimes combining two seemingly desirable changes can result in an unpredictably bad outcome. The final step in our analysis was thus to combine the best-result recommendations from each of the prior design variations into a single model and test them together (a "kitchen sink" simulation). The kitchen sink simulation combined the baseline screen width for (A) (we saw issues with decreasing the width, but no improvements with increasing width); the atomic (5 seconds linger time) versions of the (B) pump and (C) hose; and eight replicated (E) trucks with a 25% "repeat allure" decrease (larger repeat allure decreases did not dramatically reduce the half-used time, so we wanted to be conservative). The results, shown in Figure 12, reveal no negative interaction effects. To the contrary, combining these design recommendations finally flattened the growth in half-used rate. The magnitude of the decrease in exhibit-wide half-used time is also notable — a tenfold decrease from baseline when 40 visitors are present.

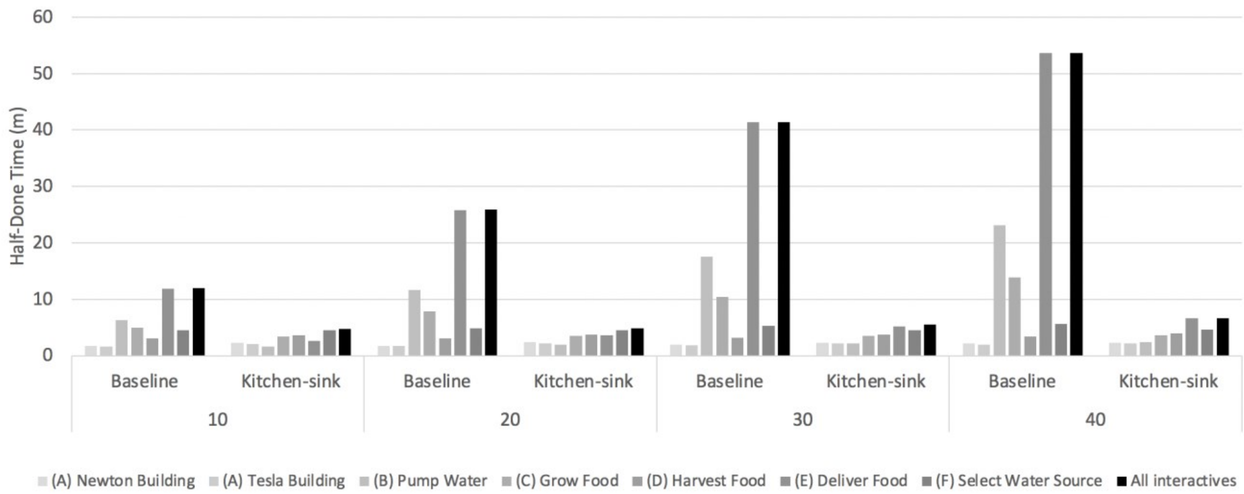

Figure 12. Time to reach the half-used state, in minutes, as the number of visitors increases, when all design recommendations are used (the "kitchen sink") compared against the baseline, using the Interest movement model

6. Discussion

The results show that ABM-based collective usability models can be used to profitably explore the impact of individual design changes on collective usability in immersive, shared XR LEs. It is worth emphasizing that while most laypeople assume all simulations are used to make precise predictions for specific outcomes, ABMs are uniquely useful for exploring the space of possible outcomes, and relative trends in those outcomes (Gilbert, 2020). Collective usability models are *thinking tools* for exploring a design space more effectively and efficiently. We demonstrated above how a collective usability model can be used to generate ideas for further model-based investigation. Below we describe the guidelines our analysis generated, and to what extent these guidelines were represented in the final exhibit design.

6.1. Simulation-derived collective design guidelines for MR exhibit

ABMs capture properties of complex systems like emergent effects, and thus can be used to generate design guidelines in two directions: bottom-up propagation of effects, and top-down propagation of design constraints.

6.1.1. Guideline derived from bottom-up effect propagation

With our model, we studied the bottom-up propagation of the linger times of individual interactives to measure the impact on collective access in terms of the half-used time. One might expect the interactives with longer linger times to impede collective access more, but we only found this to be true for the single-user tangible interactives. The multi-user screen-based interactives showed no problems scaling to support more users (see Figure 6). This suggests the design guideline:

DG1. When possible, situate interactive experiences on large, shared screens.

6.1.2. Guidelines derived from top-down propagation of design constraints

What we mean by "top-down propagation" is that we use the emergent outcome (the half-used time) to help set bounds on the design properties for individual interactives. Our simulation trials revealed several design guidelines applicable to our learning environment:

DG2. Keep the width of shared screen-based interactives to be at least 13 feet.
DG3. Reduce the "stickiness" of single-user interactives to reduce the linger time to be closer to an atomic button click in duration.
DG4. Replicate single-user tangibles that have irreducible linger times to avoid access issues. More copies than one might think (>=8) may be needed to prevent the interactives from being the limiting factor on collective access.
DG5. Mild friction may be introduced to the operation of single-user tangibles with irreducible linger times to reduce repeat allure.

6.2. Implementing the collective design guidelines in practice

Moving from prototype to final implementation can be a messy, nonlinear, and creative process. Collective usability design guidelines cannot and should not be the final word on how to design an immersive, shared XR experience, as collective usability is just one aspect of making such experiences successful. Designers need to balance collective usability against more intangible factors like the charm and pleasure that an individual interactive design can bring. While the final implementation of the exhibit did not follow all guidelines to the letter, they are present in various ways.

Figure 13. Photograph of the final exhibit. A six-story simulated waterfall (center), a river valley (middle left) and a reservoir (middle right) deliver water to the exhibit floor, which visitors can divert to four different simulated biomes projected on the exhibit walls

The final exhibit's learning goals are the same, but rather than maintaining two cities, visitors instead maintain the diversity of four different biomes: the Desert, the Grasslands, the Jungle, and the Wetlands (see Figure 13). These four biomes are each visualized on 22-feet wide projection screens (which far exceeds DG2's recommendation, but which adds a "wow" factor to the experience). Rather than positioning single-user tangible interactives in front of screens to support interactions, the design situates the interactions within the screens themselves as much as possible via gestures, per DG1. Visitors can plant seeds in these four biomes by holding their hands up in front of the screens (see Figure 14) and letting their hands quickly drop, and can prune dead plants by swiping their arms (the same action used for (D) in the pilot). These interactions are very short in duration, in keeping with DG3. Visitors supply water to the biomes by dragging large (5-feet-long) stuffed "logs" covered in an IR-reflective fabric around the floor of the exhibit (see Figure 14), diverting the flow of water from the sources. The logs are the only single-user tangible interactives in the exhibit, but manipulating the positions of these logs can take time, as they need to be moved to different spatial locations. There are 10 logs in the exhibit, per DG4, and the logs are a bit heavy, which encourages visitors to move them only when needed, per DG5. We should note that this approach did not decrease visitors' apparent enjoyment of using the tangibles — in fact, they seemed to find the difficulty part of the fun, even as it served its purpose of reducing the "repeat allure." We propose that after decades of trying to make user interfaces more seamless and more efficient, interface designers may have something of a blind spot for how strategic friction can be charming, and should embrace creative approaches for reducing repeat allure when it is demanded by collective usability.

Figure 14. A visitor selecting which available seed to plant in the Grasslands biome

We did not formally assess the collective access of the final exhibit because the earlier problems with access were completely resolved — even when used in 15 minute sessions with school groups, no access problems were observed. Visitors could and did have the opportunity to try all of the different interactives within minutes. As a whole, the exhibit has been a great success, winning multiple national awards like the 2016 Jackson Hole Science Media Award and the 2017 American Alliance of Museums MUSE Gold Award.

7. Conclusions and future work

This paper makes the case that shared XR LE design requires attention to the "holistic" configuration of people and objects in time and space (Wurdel, 2009). It defines the concept of collective usability, which positions a group of users, rather than an individual user, as the unit of analysis as they make use of an interactive experience where the human-computer and human-human interactions combine to form a complex system. It further provides an initial proof-of-concept that ABMs, which can capture the temporal and spatial relationships in collective use scenarios, can be used as collective usability models to explore the design space for shared XR experiences. Using ABMs to explore the usability design space is another novel contribution of this work. Unlike the abstruse formal models used for multi-user system validation, ABMs are easily constructed to be "deliberately approximate" (John & Kieras, 1996), extending the reach of this approach beyond experts.

This is just a proof-of-concept, so many directions for future work remain. Collective access is just one of many collective usability metrics that designers may wish to explore. For example, collective awareness, a measure of the extent to which users have detected a stimuli, may be highly relevant to collaborative XR LEs because shared visual grounding is known to be critical to supporting collaboration (Gergle, Rosé, & Kraut, 2007). Architecture researchers have long known how sight lines can impact how people behave in shared spaces, and have developed ways to model and simulate sight lines that could be borrowed from to construct models of collective awareness (Kaynar, 2005).

The highly simplified visitor movements in our collective usability model were adequate for our investigation, but other problem domains (e.g., studying if an exhibition layout poses challenges for maintaining a 6-foot social distance) may need more sophisticated models. Fortunately, there is ample literature on how visitors move in museums, from how different social groups distance themselves (vom Lehn, Heath, & Hindmarsh, 2001) to Bitgood's value model of visitor circulation (Bitgood, 2006), to documentation of how different kinds of visitors move differently — children, especially, are more prone to prioritizing exploration over exploitation (Carlisle, 1985). Creating and validating a movement model that can reproduce the spatial paths of learners may be time intensive, but once created, it can be reused and shared.

We argue that collective usability models should be thought of as another design tool, like design games (Brandt & Messeter, 2004), or design cards (Hornecker, 2010), that help designers attend to an aspect of the design space that might be otherwise overlooked. Our proof-of-concept work demonstrates how these collective usability design tools can be used. An initial collective usability model is created, either using empirical data from a prototype tested with representative users, or using theory-grounded models of user behavior. Designers use the model to generate unanticipated questions about the design space, questions that can be systematically explored by altering the design parameters embedded in the model. As the trends in outcomes become clear, design guidelines that speak to the collective usability experience can be generated using bottom-up propagation of effects and top-down propagation of design constraints. Collective usability modeling is much more logistically and financially feasible for exploring the design space for XR experiences than repeated prototyping, and can help generate design questions (like reducing interactive stickiness and repeat allure) that designers (who are not accustomed to "detuning" interactives) may not think to explore otherwise.

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant nos. 1123832, 1623094, and 1822864.

References

Abar, S., Theodoropoulos, G. K., Lemarinier, P., & O'Hare, G. M. P. (2017). Agent based modelling and simulation tools: A Review of the state-of-art software. *Computer Science Review, 24*, 13-33. doi:10.1016/j.cosrev.2017.03.001

Bitgood, S. (2006). An Analysis of visitor circulation: Movement patterns and the general value principle. *Curator: The Museum Journal, 49(*4), 463–475. doi:10.1111/j.2151-6952.2006.tb00237.x

Bernardo, M., & Corradini, F. (Eds.). (2004). *Lecture notes in computer science: Vol. 3185. Formal methods for the design of real-time systems*. Springer. doi:10.1007/b110123

Brandt, E., & Messeter, J. (2004). Facilitating collaboration through design games. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices* (pp. 121–131). doi:10.1145/1011870.1011885

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.

Carlisle, R. W. (1985). What do school children do at a science center? *Curator, 28*(1). doi:10.1111/j.2151-6952.1985.tb01687.x

Enyedy, N., Danish, J. A., & DeLiema, D. (2015). Constructing liminal blends in a collaborative augmented-reality learning environment. *International Journal of Computer-Supported Collaborative Learning, 10*(1), 7–34. doi:10.1007/s11412-015-9207-1

Garrido, J. L., & Gea, M. (2001). Modelling dynamic group behaviours. In Johnson, C. (Ed.), *Lecture Notes in Computer Science: Vol. 2220. Interactive Systems: Design, Specification, and Verification* (pp.128–143). doi:10.1007/3-540-45522-1_8

Gergle, D., Rosé, C. P., & Kraut, R. E. (2007). Modeling the impact of shared visual information on collaborative reference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1543–1552). doi:10.1145/1240624.1240858

Gilbert, N. (2020). *Agent-based models* (2nd ed.). Thousand Oaks, CA: Sage.

Goldstone, R. L. (2006). The complex systems see-change in education. *Journal of the Learning Sciences*, *15*(1), 35–43. doi:10.1207/s15327809jls1501_5

Grotzer, T. (2012). *Learning causality in a complex world: Understanding of consequence*. New York, NY: Rowman & Littlefield Publishers, Inc.

Harrison, M. D., Kray, C., & Campos, J. C. (2008). Exploring an option space to engineer a ubiquitous computing system. *Electronic Notes in Theoretical Computer Science, 208*, 41–55. doi:10.1016/j.entcs.2008.03.106

Herrmann, T., Hoffmann, M., Kunau, G., & Loser, K. U. (2004). A Modelling method for the development of groupware applications as socio-technical systems. *Behaviour and Information Technology*, *23*(2), 119–135. doi:10.1080/01449290310001644840

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences*, *16*(3), 307–331. doi:10.1080/10508400701413401

Hornecker, E. (2010). Creative idea exploration within the structure of a guiding framework: The Card brainstorming game. In *Proceedings of the 4th International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 101–108). doi:10.1145/1709886.1709905

Ivory, M. Y., & Hearst, M. A. (2001). The State of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys, 33*(4), 470–516. doi:10.1145/503112.503114

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences*, *15*(1), 11–34. doi:10.1207/s15327809jls1501_4

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques. *ACM Transactions on Computer-Human Interaction (TOCHI), 3*(4), 320–351. doi:10.1145/235833.236054

Kaynar, I. (2005). Visibility, movement paths and preferences in open plan museums: An Observational and descriptive study of the Ann Arbor Hands-on Museum. In *Proceedings of the 5th International Space Syntax Symposium*, (pp. 1–15). Retrieved from http://kaynar-rohloff.com/papers/kaynar_SpaceSyntax05.pdf

Langner, N., & Kray, C. (2014). Assessing the impact of dynamic public signage on mass evacuation. In *Proceedings of the 3rd International Symposium on Pervasive Displays* (pp. 136–141). doi:10.1145/2611009.2611033

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers and Education, 95*, 174–187. doi:10.1016/j.compedu.2016.01.001

Lyons, L. (2009). Designing opportunistic user interfaces to support a collaborative museum exhibit. In *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning* (pp. 375–384). doi:10.3115/1600053.1600110

Malinverni, L., Ackermann, E., & Pares, N. (2016). Experience as an object to think with. In *Proceedings of the 10th International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 332–339). doi:10.1145/2839462.2839477

Mangina, E., Carbo, J., & Molina, J. (Eds.). (2010). *Agent-based ubiquitous computing*. Amsterdam, The Netherlands: Atlantis Press.

Massink, M., Latella, D., Beek, M. H., Harrison, M. D., & Loreti, M. (2008). A Fluid flow approach to usability analysis of multi-user systems. In P. Forbrig & F. Paternò (Eds.), *Lecture Notes in Computer Science: Vol. 5247. Engineering Interactive System* (pp. 166–180). doi:10.1007/978-3-540-85992-5_15

Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., & Saarikko, P. (2008). It's mine, don't touch!: Interactions at a large multi-touch display in a city centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1285–1294). doi:10.1145/1357054.1357255

Poltrock, S., & Handel, M. (2010). Models of collaboration as the foundation for collaboration technologies. *Journal of Management Information Systems, 27*(1), 97–122. doi:10.2753/MIS0742-1222270105

Reeves, S., Benford, S., O'Malley, C., & Fraser, M. (2005). Designing the spectator experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 741–750). doi:10.1145/1054972.1055074

Roberts, J., & Lyons, L. (2017). The Value of learning talk: Applying a novel dialogue scoring method to inform interaction design in an open-ended, embodied museum exhibit. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 343–376. doi:10.1007/s11412-017-9262-x

Salvador, T., Scholtz J., & Larson J. (1996). The Denver model for groupware design. *SIGCHI Bulletin, 28*(1), 52–58.

vom Lehn, D., Heath, C., & Hindmarsh, J. (2001). Conduct and collaboration in museums and galleries. *Symbolic Interaction*, *24*(2), 189–216.

Wilensky, U. (1999). *NetLogo*. [Computer software]. Retrieved from http://ccl.northwestern.edu/netlogo/

Wurdel, M. (2009). Towards an holistic understanding of tasks, objects and location in collaborative environments. In M. Kurosu (Ed.), *Lecture Notes in Computer Science*: Vol. *5619. Human Centered Design* (pp. 357–366). doi:10.1007/978-3-642-02806-9_41