# Storytelling to Sensemaking: A Systematic Framework for Designing Auditory Description Display for Interactives

**Taliesin L. Smith**
University of Colorado Boulder
Boulder, USA
taliesin.smith@colorado.edu

**Emily B. Moore**
University of Colorado Boulder
Boulder, USA
emily.moore@colorado.edu

## ABSTRACT
Auditory description display is verbalized text typically used to describe live, recorded, or graphical displays to support access for people who are blind or visually impaired. Significant prior research has resulted in guidelines for auditory description for non-interactive or minimally interactive contexts. A lack of auditory description for complex interactive environments remains a tremendous barrier to access for people with visual impairments. In this work, we present a systematic design framework for designing auditory description within complex interactive environments. We illustrate how modular descriptions aligned with this framework can result in an interactive storytelling experience constructed through user interactions. This framework has been used in a set of published and widely used interactive science simulations, and in its generalized form could be applied to a variety of contexts.

### Author Keywords
Auditory description display; Description design; Non-visual access; Interactive information spaces.

### CCS Concepts
•**Human-centered computing** → **Accessibility design and evaluation methods; Accessibility systems and tools;** *Systems and tools for interaction design;*

## INTRODUCTION
Since the origin of language, humans have been storytellers. To paraphrase Gershon and Page (2001), "stories are efficient, easier to understand, and just more compelling" [8]. With the graphical user interface, the HCI community has made tremendous advances in the design and implementation of engaging and effective storytelling through interactive and highly visual experiences. Most modern websites, games, and learning resources emphasize visual display to tell a story, whether it be about a company's brand, a hero's journey, or a learning progression. For people with visual impairments, without associated descriptive content with these visual displays, there is no access to these stories. Auditory description display is verbalized text typically used to describe live, recorded, or graphical displays to support access for people who are blind or have significant visual impairment. However, advancements in auditory description display are not keeping pace with advancements in visual display.

Significant prior research has produced auditory description guidelines for non-interactive contexts. Guidelines, training materials, and widely used technologies support effective and efficient description of live performances [9], film and video [20], images and static graphics [1, 2], and charts[6, 42] . While work remains to expand these guidelines, increase overall uptake, and investigate advancements (such as automation [12, 7, 39]), this existing body of knowledge stops short of addressing the needs of many interactive contexts.

The complexity of most interactive digital resources poses a unique challenge for auditory description display. Typically, in description the designer (or describer) relies on the visual layout or an existing narrative to decide on wording to use, focusing on providing descriptions that are concise, accurate, objective, avoid potential spoilers, and are delivered in a way to minimize conflicts with other auditory display. In contrast, within interactive environments, the visual layout and narrative are often co-constructed between the user and the interactive environment over the course of the users' unique sequence of interactions. It is a highly complex task to design auditory description display capable of engaging users in a dynamic non-visual experience of a compelling story, while simultaneously remaining in alignment with all other display modalities.

Herein we present an approach to designing and delivering an interactive auditory description display capable of providing access to all relevant information. The outcome can support user agency in inquiry and sensemaking [19] by situating the user within a robust and compelling interactive story. As technology continues to evolve and display modalities co-mingle, this framework presents a foundation for efficient and effective design of description display for many existing interactive environments. This framework serves as a necessary bridge towards a comprehensive approach for description display spanning in complexity from static graphics to highly complex interactive environments. While our work focuses on educational interactives, the potential applications of this framework span digital interactive experiences broadly, including games and XR experiences. Additionally, our approach highlights


CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376460

the important role of storytelling [8] in multimodal displays, relevant to all designers.

Traditionally, auditory description display in digital resources is considered as an accessory to a visual display object [13]. For example, alternative text (or "alt text") is a text description provided for a static image that can be read aloud for a user with screen reader software. With alternative text, the graphical display of the image is the primary modality and description is an accessory to provide non-visual access to the image [1, 2]. Emphasizing a different approach, our work extends prior research on description for interactive scientific graphics [11] that elevates auditory description display to a primary - rather than accessory - modality. Within our design framework and implementations auditory description display functions in parallel [36] with graphical displays. Advancing this perspective, we present a systematic description design framework that can be used for designing modular descriptions for complex interactives.

### Prior Work in Auditory Description Display

Auditory description display has a rich history in the contexts of theater [9], movies and television [20], and images [1, 2], where people with visual impairment would not otherwise have access to the visual components of a performance or resource. This history included the development of guidelines and software for describers, telecommunications and digital infrastructure for public access to accessible materials, and advocacy to advance federal legislation regarding accessible media [20]. Out of these efforts emerged multiple guidelines for live description (spoken description provided in real time), recorded description (called video description, a description added to a recording of an event or performance) [32], and the previously mentioned alt text [1, 2]. These guidelines detail necessary qualities of description such as accuracy, consistency, emphasis on essential content, and avoidance of overlaps with other auditory displays. These resources primarily refer to media where the narrative or graphical structure is known in advance and remains unchanged by the user – requiring the delivery of a single stream of thoughtfully designed description. In our work, we focus on addressing the auditory description needs in interactive situations, in which user actions affect what happens.

### Advancements in Auditory Description Display

While using an interactive resource, the user develops a conceptual model of the resource – a mental model [18] of what is represented, and the relevant capabilities and constraints of the interface. In analysis of the development of mental models by users of digital resources relying primarily or entirely on auditory information [14, 4], users enacted strategies to first explore and develop an initial mental model (e.g., listening carefully while navigating content), prior to interacting (e.g., taking an action or executing a change) and then updating their mental model as the resource changes. Consequently, effective auditory description display for interactives needs to be able to "set the scene" for users, supporting users in constructing an initial mental model for the experience and then providing updated information as the scene, characters, and story evolves.

Keane's investigation of description for interactive scientific graphics, such as dynamic charts and graphs, supports simple interactive graphics containing two object types: the changing graphics and the interactive controls. The resulting guidelines for the description of interactive scientific graphics [11] recommend that description for these two object types should differ in content and delivery, making first steps towards a modular approach to auditory description display.

In this paper we present a significant new advancement in auditory description display, a comprehensive description design framework presented in a generalized form that has been successfully applied to a set of complex interactives. We first summarize the development of the description design framework and then introduce the description framework by defining each component and how each component is accessed by or delivered to the learner. We then describe how the description framework can be used as a design medium to systematically design a fully described, user-centered, interactive experience. In the section *Storytelling to Sensemaking*, we demonstrate how the descriptions for the *Resistance in a Wire* simulation engage a learner in an interactive described experience – an experience in which the learner is in control and through which they engage with and make sense of the simulation story – the story of the changing resistance. We conclude with reflections on the framework as a design aid, and possible uses and extensions to contexts beyond interactive simulations.

## DESCRIPTION DESIGN FRAMEWORK'S DEVELOPMENT

In 2015, the PhET Interactive Simulations project began supporting the implementation of auditory description display for a subset of the 150+ free and open-source interactive simulations comprising the PhET suite of simulations (or sims). As part of a master's research project in inclusive design [35], Smith (2016) explored iterative designs for accessible interactions and descriptions for the sim *Balloons and Static Electricity* [24] with 13 blind screen reader users. This investigation explored what descriptions, structures, and interactions were needed to make an auditory description display accessible and engaging for blind learners.

The descriptions and interactions implemented used an early version of PhET's Parallel DOM architecture (PDOM) [36]. The PDOM's accessibility API was designed and developed in conjunction with descriptions for the sim. Through iterative analysis, Smith et al identified three description categories (i.e., "static descriptions", "dynamic descriptions", and "interaction alerts") [37] and shared description design strategies found to be effective for interactive descriptions [38].

With the beginnings of a description design framework and software architecture in place, we began expanding our work in auditory description display across multiple sims utilizing PhET's iterative and inclusive design research approach [21, 22]. For each sim we start by drafting descriptions. With a working prototype, we have early design discussions with consultants who are blind who have both content knowledge and screen reader expertise. After refining the descriptions, we conduct an iterative series of interviews (in-person or remote) to investigate how the descriptions and interactions support access and engagement. In all interviews we ask participants to freely explore a sim for 5-10 minutes or more, while 'thinking aloud'. We then ask follow-up questions to better understand their use and interpretations. We analyze the recordings and interview notes to identify use patterns, sensemaking, successes, and challenges to inform design improvements.

Currently, descriptions in PhET sims are accessible solely through use of screen reader software. All data collection to support the design and evaluation of description (and correspondingly, the description framework) was collected from participants using screen reader software. Interview participants varied in age (19 to 61), geographic location, science knowledge, and expertise with assistive technology, representing a broad diversity within this population of technology users.

While creating descriptions for a dozen sims (including *John Travoltage* [27], *Friction* [25], *Resistance in a Wire* [28], and *Gravity Force Lab: Basics* [26]) and conducting more than 100 interviews with screen reader users, we continuously sought structures and patterns that could increase the efficiency of description design, improve the description-supporting software architecture, and organize the workflow. After the first few sims, more refined categories and patterns emerged, and we applied these proto-frameworks to each subsequent sim, reflecting on what held in a new context, what did not, and why, iteratively refining the framework until application to new sims resulted in no updates to the foundational structure (Figure 1). We present the generalized version of the resulting framework in the following section.

Notably, rather than providing learners with a static, linear sequence of events in the story, an interactive auditory description display designed with the description design framework provides relevant components of the story as the learner explores and interacts. In this way, discovery of the story characters, character relationships, and character evolution occurs through learner interaction - inviting learners to advance one of many possible narrative arcs through interaction and interpretation of their changes. Learners can use this form of storytelling to engage in sensemaking, a "dynamic process of building an explanation in order to resolve a gap or inconsistency in knowledge" [19], though presumably the approach can serve other purposes depending on the context, e.g., enjoyment within a game.

## DESCRIPTION DESIGN FRAMEWORK

The Description Design Framework (Figure 1) has two primary components (State Descriptions and Responsive Descriptions) and four subcomponents (Static and Dynamic States, and Object and Context Responses). Applied to a specific interactive system or information space [30], we believe these categories encapsulate the complete and necessary information and delivery methods (on-demand or automatic) that are needed to provide robust text-based access to interactive experiences in which the user initiates all changes.

### State Descriptions Provide Current State

The complete and accurate description of the current state of the information space is made available to the user by State Descriptions (Figure 1, left box). State Descriptions

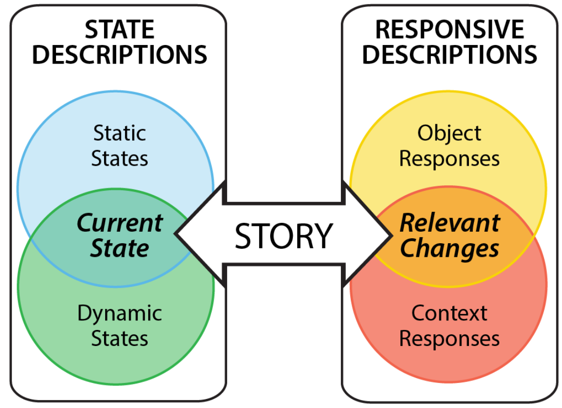


**Figure 1. Description Design Framework**

can be accessed on-demand at any time, providing the user with a consistent frame of reference to contextualize changes to the information space. State Descriptions consist of two description types: Static State descriptions and Dynamic State descriptions. Presented in two distinct, but interleaved groups, these descriptions, together, form a union that constitutes a complete description of the current state of the information space (Figure 1, left box, union of upper and lower circles).

#### *Static State Descriptions*

Within an interactive information space, some information remains constant. Static States (Figure 1, left box, upper circle) identify and/or describe entities within the information space that are always present and unchanging. Because of this, Static States themselves remain constant – always accurate and true within an otherwise changing information space.

#### *Dynamic State Descriptions*

Of course, in an interactive system some aspects of the information space must change when the user interacts with the information and/or objects within the space. Dynamic States (Figure 1, left box, lower circle) identify and describe entities that appear, disappear, or change state within the information space. Dynamic States change with the system – appearing, disappearing, or being altered – silently in the background.

### Responsive Descriptions Deliver Relevant Changes

State Descriptions alone – always complete and accurate – do not provide full access to an interactive system. A different mechanism is needed to convey relevant changes as they happen. Responsive Descriptions (Figure 1, right box) are provided at the start of and throughout an interaction with an interactive object. Responsive Descriptions consist of Object Responses and Context Responses that are triggered and delivered automatically when encountering and interacting with an interactive object. Delivered sequentially and together, Responsive Descriptions announce (automatically) all relevant changes (Figure 1, right box, union of upper and lower circles) to the information space, e.g., changes to the interactive object and any surrounding contextual changes.

### *Object Responses*

Object Responses (Figure 1, right box, upper circle) are provided automatically upon encountering an interactive object to prepare the user for interaction. Object Responses can convey multiple attributes, including: the interactive object's current value (or state), identity, and function (or role). Object Responses continue to be provided in conjunction with Context Responses as interaction with the object proceeds.

### *Context Responses*

In an interactive system, changes made to one object may impact objects or representations throughout the system. Context Responses (Figure 1, right box, lower circle) capture changes happening outside the object the user is actively manipulating. Like Object Responses, Context Responses are delivered automatically upon any state changes to the surrounding context. When associated with a changed object's value, the Context Response is delivered after the Object Response. Some interactions do not have an associated value in which case the Context Response is delivered immediately, upon the change or changes that occurred to the system's overall state.

## THE FRAMEWORK AS A DESIGN MEDIUM

In this section, we demonstrate how we use the framework as a design medium to design the descriptions for the PhET sim, *Resistance in a Wire*. Two screenshots of the visual display of *Resistance in a Wire* are shown in Figure 2 – one before a change has occurred (top) and one after changes have been made (bottom). In this sim, the resistance equation, $R$ equals $\rho$ (Greek letter 'rho') times $L$ over $A$, is displayed above an illustration of a piece of wire. Learners can change the resistivity ($\rho$), length ($L$), and area ($A$) of the wire by interacting with the large slider controls situated to the right of the equation and wire. As the value of a variable is increased (or decreased), the size of the corresponding letter in the equation increases (or decreases) as well as the size of letter $R$ (resistance). Simultaneously, the value for resistance (above the sliders) and the wire representation change, with increasing (or decreasing) resistivity (impurities in the wire), wire length, or wire area. Using the slider controls learners freely explore how changes to resistivity, length, and area affect resistance, thereby discovering, through interactive investigation, the learning goals of the sim.

### State Descriptions Encourage and Frame Interaction

A learner is able to read or skim through the State Descriptions (Figure 3 first and second columns) in part or in full to glean from them what they need in the moment. Because the State Descriptions are accessed on-demand at the learner's own pace they are the medium that affords the designer the most flexibility in terms of providing framing and details. Static States do not change, but during any change to the information space, Dynamic States are updating ensuring an always accurate variation of the State Description is available to be re-explored and examined.

#### *Applying Static State Descriptions*

From the learner's perspective, details in the first column of Figure 3 succinctly answer questions such as:

- What is this about, and what is here?

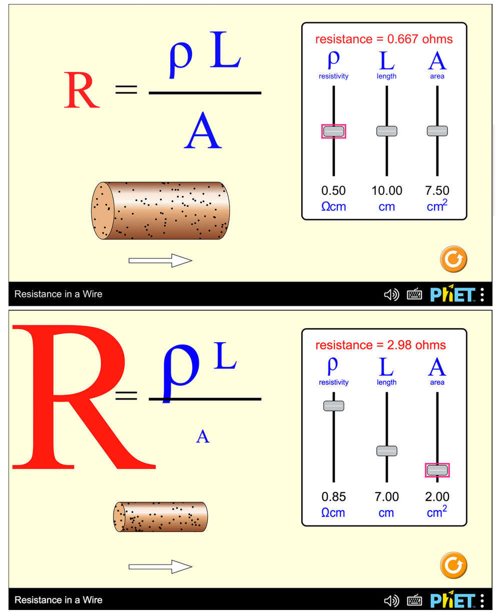


**Figure 2. The PhET Simulation *Resistance in a Wire*, on startup (top), and after interaction with each slider (bottom). Keyboard focus (pink highlight) shown on a slider in each screenshot. Images copyright PhET Interactive Simulations.**

- What is important or interesting, and needing my attention?
- What can I interact with or change?
- What should I do first?

From the designer's perspective, descriptions that answer these framing-type questions are basic design features - features that have no need to change within an interactive system, like:

1. Identifying names such as the system name, and names for primary objects, regions or groups;
2. A brief summary introducing the information space by identifying primary objects and their locations within regions;
3. General guidance for interaction such as interaction hints or help text for interactive objects.

For the *Resistance in a Wire* sim, notice in the first column of Figure 3 the bold text (headings) identify the sim's main characters (i.e., "Resistance Equation" and "The Wire") and important regions of the scene (i.e., "Play Area", "Control Area", and "Sim Resources"). The first two paragraphs describe the scene, and just before the "Play Area" there is an interaction hint naming the three sliders, thus indicating a potential path of interaction for the learner. An additional hint, flagged as optional, indicates where to find help. In our work on interactives, we utilize Static States to provide consistency across sims (e.g., common region names) and necessary framing details to implicitly scaffold [31, 16] learners into productive explorations.

#### *Applying Dynamic State Descriptions*

Constantly updated and accurate Dynamic States (Figure 3, second column) accessed together with Static States (lines

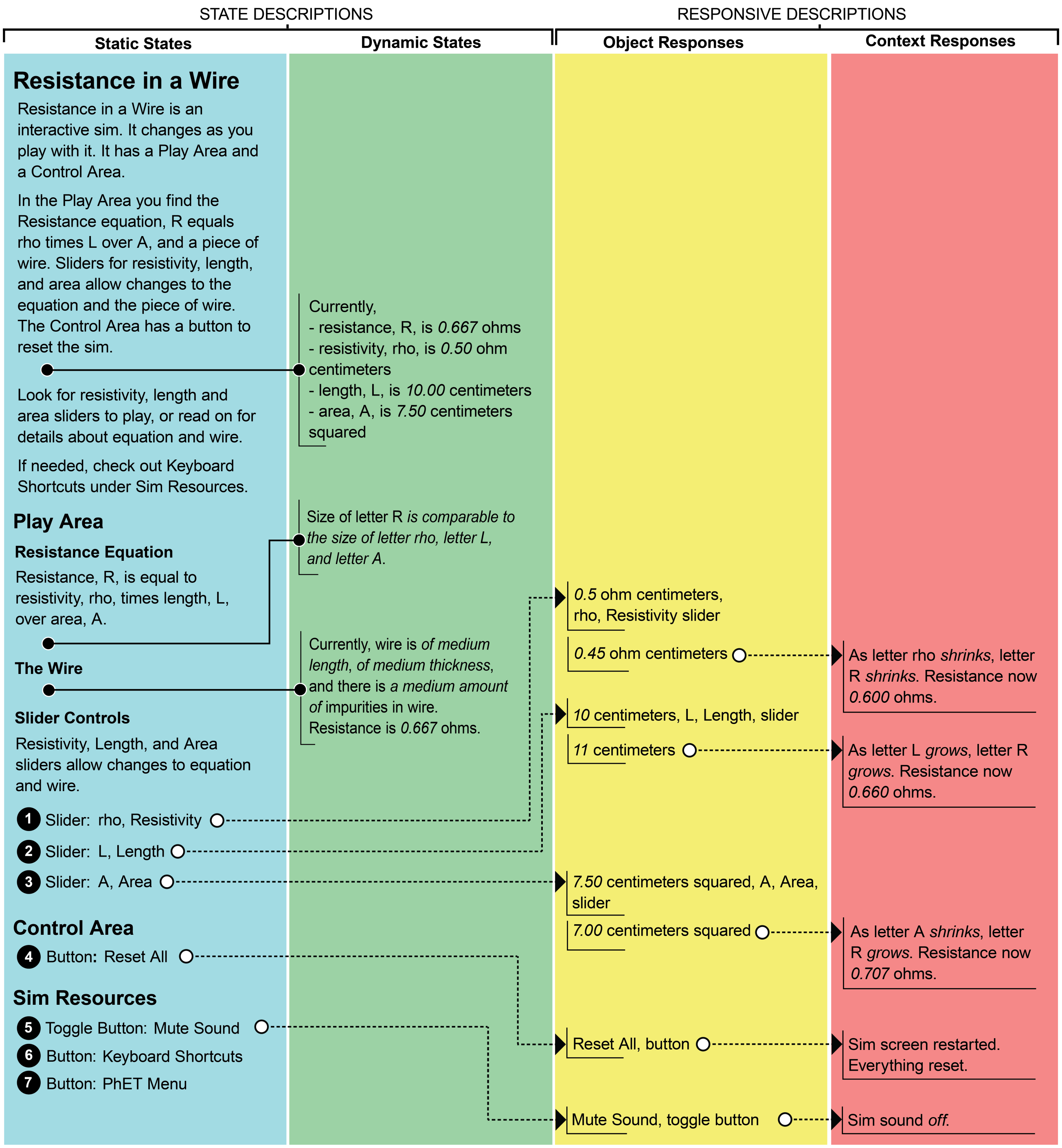


**Figure 3. State and Responsive Descriptions for *Resistance in a Wire* Simulation**

between first and second columns ending with filled circles mark locations of Dynamic States) enable users to answer more open-ended questions, such as:

- I just changed [object 1], did anything else change?
- What effect did changing [object 1] have on [object 2]?
- Something different happened when I changed [object 3]; what is the big picture of the whole system now?

There are three Dynamic States in this sim that contain details to answer such questions. The first one serves as a summary, providing consolidated access to important information – exact values for resistance, resistivity, length, and area. Two additional Dynamic States provide specific details about the sim's main characters – the "Resistance Equation" and "The Wire". Neither character is directly interactive, though changes that occur simultaneously to both represent the main concepts conveyed in this sim. Dynamic States, thus, are used to express the evolving characteristics of the characters. Placing them under discrete headings (Figure 3, bold text in first column) ensures these changing details have a prominent place within the overall interactive story being told through the learner's use of the sim.

Not all dynamic objects warrant names and descriptions like the equation and wire do here. Consideration of the goals, complexity, and overall experience being designed determines when individual characters need to be created, or when state changes can be effectively conveyed as part of a dynamic summary. Note this sim employs both approaches, a dynamic summary near the beginning, as well as Dynamic States highlighted under discrete headings.

Even though Dynamic States and their parameters (Figure 3, second column, italicized text) make up much less content than Static States, they are how we provide details needed for open-ended questions and to support and encourage deeper investigations of different states. Learners use them to verify and compare details, not available during interaction, allowing them to gain new insights about the changes they made - ideally sparking new questions.

### Responsive Descriptions Sustain Interaction

The learner's perspective changes during a choice to interact. They are no longer seeking information from the current state of the system (i.e., State Descriptions), and instead are making changes to it. They now need information regarding the objects they are interacting with, the changes they are making, and the impact their changes are having on the information system. Responsive Descriptions are the medium used to indicate how the story evolves as the learner interacts and begins a conversation with it, a dialogical inquiry [10, 30]. Object Responses and Context Responses are applied to directly and automatically respond to the learner's interactive explorations.

#### *Applying Object Responses*

As the user begins to interact, they need information about where they are (e.g., what object has their physical or focused attention), from where they are starting, and an indication of what changes occur to this object as they interact with it. Object Responses immediately answer questions such as:

- What is this [object] and how do I interact with it?
- What is this [object's] current value or state?
- How will this [object] change as I interact with it?

Object Responses (Figure 3, third column), provide the specific object details that are needed at the start of an interaction and as interaction continues. For consistency, initial Object Responses may be composed in part or whole with State Descriptions, and include information about the function of the object. For example, an interactive object is named in the Static States (see Figure 3, first column), and that name becomes part of that object's initial Object Response as the user begins to interact with it (see Figure 3, connecting arrows from first column to third column).

When a learner, using screen reader software, moves their keyboard focus to the resistivity slider (an interactive object) the sim delivers an initial Object Response – for example: "0.50 ohm centimeters, $\rho$, Resistivity, slider." The Object Response provides: the current value of the object, the name of the object, and the function of the object (i.e., slider). As the learner interacts with the slider, they automatically receive new Object Responses containing the new current value for resistivity, for example: "0.55 ohm centimeters," "0.60 ohm centimeters," for slow careful exploration, or when a more rapid interaction results in larger changes, the latest current value is read out, "0.85 ohm centimeters" skipping the resistivity values that were passed over quickly, and staying synchronized with the speed of the learner's exploration. Object Responses are applied to the description design to ensure object-details are available during interaction.

#### *Applying Context Responses*

In an interactive system, a change to one object typically results in changes to one or more additional objects. Conveying the new value of the object being changed with Object Responses does not fully describe the progression of the story. An additional response that contains more contextual information is needed to describe multiple simultaneous changes when they occur. Context Responses (Figure 3, fourth column) provide additional information about changes happening to other areas of the information space, beyond the object being interacted with by the learner. Object Responses and Context Responses, together, fully inform the learner of all *relevant changes* as they interact and the changes occur. Context Responses allow users to make connections between the changes they are making with one object and how the information space is changing in response (see Figure 3, connecting arrows from third column to fourth column).

Prior to interaction, users often have investigative questions or ideas about what will happen during or after an interaction, such as:

- What happens when I change [object 1]?
- If I change a [different object], will the same thing happen, or will something different happen?
- How does [object 2] change if I change [object 1]?

- I think if I change [object 1], [object 2] will change in certain ways. Is this true?

As the user interacts, Context Responses provide information allowing the user to actively investigate deeper contextual questions. Delivered in response to their interaction choices, Context Responses make re-reading the changed State Descriptions optional to understand changes taking place in the story.

In the *Resistance and a Wire* sim, during a change to any of the three slider controls (resistivity, length, or area) the learner needs to be made aware of the relevant changes happening to the relative size of the letters in the equation and the resulting changed value of the resistance in the wire. When the resistivity slider is changed, the full Responsive Description delivered is, for example, first an Object Response, “0.55 ohm centimeters”, followed immediately by a Context Response, “As letter $\rho$ grows letter $R$ grows. Resistance now 0.733 ohms.” Further increasing the resistivity slider delivers the Object Response, “0.60 ohm centimeters”, followed immediately by the Context Response, “As letter $\rho$ grows, letter $R$ grows. Resistance now 0.800 ohms.” A big increase to resistivity, a jump to “1.00 ohm centimeters” results in a slightly different Context Response: “As letter $\rho$ grows, letter $R$ grows a lot. Resistance now 1.33 ohms.” Through interacting with each slider, the learner can discover that resistivity and length share a linear relationship with resistance, in contrast to area which has an inverse relationship with resistance. Importantly, with Context Responses the learner does not have to stop their interactive exploration and return to the State Descriptions to determine the resulting change to the size of the letters in the equation or to measure the value for resistance. Context Responses do not have the space to deliver all details. They are applied to sustain interaction by delivering contextual changes needed to discover the essential concepts of the sim’s story.

By using State Descriptions to appropriately frame and encourage interactions and Responsive Descriptions to deliver relevant and timely responses about what is happening, the learner is able to fully engage in an interactive described experience that unfolds as a seamless story and makes sense at every interaction and with every examination of the system’s state.

## NOTE ON ACCESSIBLE IMPLEMENTATION

In other work, we describe in detail the Parallel DOM (PDOM) architecture [36] we built that creates the accessibility layer for PhET sims. Importantly, the PDOM, is visually hidden and runs in parallel with other modalities (e.g., visual display and non-speech sound display). As we design descriptions we also determine and design the semantic structures that house the descriptions in the PDOM. While these structures are not technically descriptions, without them the descriptions in the sims would be inaccessible. Because PhET sims are rendered in a browser the document and interaction semantics defined by the HTML and WAI-ARIA specifications [5, 3] are what we use to create the necessary semantic “hooks” [17] to make the auditory description display of the sims interactive and fully navigable when using screen reader software and alternative input methods (e.g., keyboard). This web-based ontology embeds a meaningful and navigable hierarchy around State Descriptions and provides a robust delivery system for Responsive Descriptions. When applying the description design framework, designers (and/or developers) need to also consider the ontology of the system [29].

## STORYTELLING TO SENSEMAKING

From over 100 interviews with screen reader users, we have identified a common interaction pattern across sims, and a related pattern variant that is less commonly observed. These interaction patterns are utilized by learners to engage with the storytelling descriptions, which in many cases leads to a rich sensemaking [19] experience that supports science learning [41]. To illustrate the most common interaction pattern, we provide examples from two interviews with the multi-modal sim, *Resistance in a Wire*. We also use these examples to highlight contrasting learning experiences: one learner transitions from a storytelling to sensemaking experience, while the other encounters challenges that impact the transition to sensemaking.

Figure 4 shows the interaction patterns of two learners, Lynne and Rachelle (pseudonyms), using *Resistance in a Wire* for the first time. Note that, in addition to description, this multi-modal sim displays non-speech sounds, in the form of a short tone representing the current value of resistance, triggered each time resistance is changed. The result is that a tone plays just prior to hearing each Object Response. A higher pitch indicates a smaller amount of resistance and lower pitch indicates a larger amount of resistance. This data was collected from a study of the auditory description and sound design for this sim. The study included eight adult learners with a visual impairment using the sim with screen reader software (6 female, 2 male; age range 24 - 59 years). The interviews consisted of up to 10 minutes of free ‘think aloud’ exploration of the sim, followed by questions regarding their experience with the sim’s auditory displays (description and sound), and three standardized surveys evaluating user experience of sound, description, and usability. Data segments shown in Figure 4 capture the first six minutes of each learner’s free exploration of the sim, without interruption or questions from the interviewer. Their continued sim use, not included in Figure 4, were prompted by researcher questions – detailed analysis of the full data set from these interviews is the focus of Tomlinson, et al (2020) [41].

### Common Interaction Patterns

From investigation of users’ interaction patterns and their verbalizations as they interact, the most common interaction pattern proceeds as follows; we use Lynne’s sim use (Figure 4, top) to illustrate. First, the learner starts by reading through the State Descriptions (Figure 4, hashed green segments). Lynne reads the State Descriptions (including the introductory summary and Play Area content) for one minute and 22 seconds (00:00 - 01:22) with her screen reader speech rate set to fast. This initial exploration of the State Descriptions provides her with an overview of the sim, and its characters. While reading through the State Descriptions, the learner constructs an initial mental model of the sim. A participant in this study

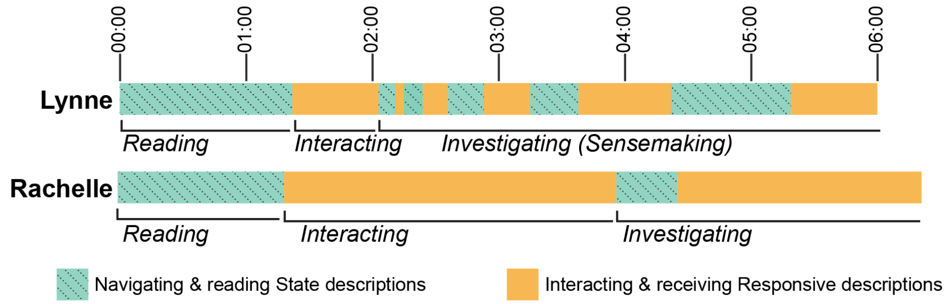


**Figure 4. Interaction patterns of two learners using the PhET Simulation *Resistance in a Wire***

.

(not shown), summarized their mental model after reading through the State Descriptions in the following way, "[...]It's basically explaining that the amount of electricity that's gonna go through this wire is going to be based on the amount of resistance in it, and the length in it, and how thick it is. And it's giving you that you can change the amount of resistivity in the wire."

Next, the learner transitions to interacting with the sim's interactive objects (in this case, the slider controls) and starts hearing the Responsive Descriptions (Figure 4, hashed green segments) , which are delivered automatically on interaction. Lynne initially interacts with all three sliders (01:22 - 02:03), hearing complete Responsive Descriptions providing her with the changes happening to resistivity, length, area, and resistance. After interacting with the sim's interactive objects and hearing the sim's Responsive Descriptions, the learner has updated their initial mental model and gained some understanding of what the interactive objects can do and how their changes can impact the details of the story. Ideally, the updated mental model leaves the learner with useful questions that need further investigation through sensemaking.

Now the learner transitions from investigating the story to engaging in sensemaking, a dynamic process of building an understanding of the sim's scientific content [19]. During sensemaking, the learner constructs an understanding of the relationships between objects and how the sim (and story) evolves as new scenarios are created or encountered.

Up to this point, Lynne's exploration has taken approximately two minutes and she is about to transition to sensemaking. At (02:03) Lynne uses screen reader shortcut keys to quickly jump to the State Descriptions and reads the updated description of the Wire. She then jumps to the sliders and quickly maximizes length. Without waiting to hear the complete Context Response, she jumps again to the State Descriptions to check the current state of the wire and the new value for resistance. She continues identifying and investigating specific scenarios by interacting with the sliders (resistivity, length, and area) and then reading the updated State Descriptions describing the wire. On her last read through of the State Descriptions (04:26 - 05:23), she reads through the State Descriptions of everything - the summary, the equation (including the comparative size of all the letters), and the wire.

In six minutes of transitioning back and forth between State and Responsive Descriptions, Lynne is making sense of the story of resistance, and says aloud to the interviewer, "so these pitches are all equivalent to what resistance is doing." Lynne continues to actively explore the sim, listening to Responsive Descriptions to verify that her understanding is correct as she summarizes how resistance changes to the interviewer, making some verbal mistakes but demonstrating clear understanding of how resistance changes, "as length grows, resistance grows which is why the pitch gets lower" and then later, "so higher pitches mean shrinking resistance."

A variant (not shown) of this interaction pattern we have observed starts with learners opting to interact with a sim's interactive objects prior to reading through the State Descriptions. In this case, learners get a sense of what is available to them for interaction first (from the Responsive Descriptions), and next, read through the State Descriptions to complete their development of a mental model of the sim's overall story.

Notably, each users' interaction pattern is unique – descriptions with the use of the design framework do not constrain learners to prescribed pathways of interaction. What is consistent is that once users have a sense of the story and begin investigating the science content in the sim (i.e., engaging in sensemaking), their interaction involves transitions between State and Responsive Descriptions. After exploratory and semi-structured interviews, learners engaging in these weaving interaction patterns describe their experience as a positive and supportive one. When asked about the descriptions, Lynne said, "I really enjoyed looking at the wire [...] I liked the descriptions when I would do all the extreme things. It would say, 'Wire is extremely long, extremely thin' [...] It struck me as very passionate [...] It made it really easy to picture, too."

### Common Challenges

With any interactive experience, some users will experience challenges. Common challenges encountered by learners in our work include 1) challenges related to familiarity with screen reader software and 2) challenges related to familiarity with science content. Screen reader technologies are sophisticated tools, often requiring specific training and practice to develop the skill to use fluidly. Novice screen reader users, or those not utilizing their preferred screen reader, can become distracted or inhibited during sim use (for example, encountering a forgotten keyboard shortcut), ultimately complicating the transition to sensemaking. Additionally, within our context of interactive science simulations, some users may have less interest in or familiarity with science learning experiences, also inhibiting the transition to sensemaking, potentially from lack of experience or personal interest in investigating the science content more deeply.

A second sim use, by Rachelle (Figure 4, bottom) provides an example of a learner encountering challenges. Rachelle indicated at multiple points a lack of confidence in her science learning capabilities, and though she effectively explored the basics of the simulation content – the simulation story – she did not make the transition into sensemaking.

Rachelle's exploration is much more conservative in nature. She uses the simplest navigation methods with the screen reader – the Down Arrow key and the Tab key – and she seems, overall, less comfortable with the science content. That said, Rachelle exhibits the same general interaction pattern as Lynne, starting with reading the State Descriptions, and then transitioning back and forth between interacting and reading (though much less than Lynne) as needed to find answers to her questions. She was able to make sense of basics of the simulation's story and its characters, but did not yet transition to the sensemaking that occurs with deeper investigation.

Over the course of our work in auditory description display for interactive science simulations, the description framework has been refined to increase support and flexibility across the interaction patterns we have observed. As users are free to follow unique interaction pathways, they need the flexibility to switch between State Descriptions and Responsive Descriptions, or change their overall approach to using the simulation, fluidly and spontaneously. With facility using the screen reader software and a learning frame of mind, the description resulting from the framework supports a transition from basic access to a rich sensemaking experience.

## DISCUSSION

### Contributions of the Framework

Creating auditory description display for interactives is a challenging task for design and implementation. In this work, we have introduced and illustrated a systematic framework for the design of auditory description display for interactive resources, including description design and delivery attributes. The framework serves as a useful organizational design tool for this emerging information display modality in interactive environments.

Within our own context of interactive science simulations, we have found this framework to be a highly effective design tool. The framework creates a consistent design process. The building out of descriptions starts with State Descriptions, followed by Responsive Descriptions – with iteration to ensure coherence and consistency, and to optimize the design of both. For example, Dynamic State descriptions are ideal for secondary details that might not fit within an Object or Context Response which need to be short in order to be timely. Additionally, the components of the framework create a consistent design language that has enhanced communication between designers and developers on the team.

In addition to the framework, we have created other interactive tools to support the description design process. The Accessibility View (or "A11y View") was the first tool we created, and now utilize heavily. It displays the visual experience and the described experience side-by-side making it possible to verify, test, and debug descriptions for both designers and developers without requiring screen reader software (see [23] for the *Resistance in a Wire* A11y View). We are also exploring a dynamic phrase builder that could be useful for description designers (see [34] for an early prototype).

Notably, this work also highlights an example deployed at scale, the *Resistance in a Wire* simulation, of a complex auditory description display effectively co-existing with the visual display. Rather than offering an alternative resource, we are instead able to utilize an inclusive approach that affords opportunities for those with and without visual impairments to engage together with the same interactive system. In the case of *Resistance in a Wire*, it is conceivable that a non-visual experience could be created that would afford achievement of the same learning goals through an alternative activity, such as an accessible document containing data tables. But this would not afford access to an equivalently playful and collaborative experience across visual and non-visual displays [33].

### Advancing Auditory Description Display

Our framework builds on the work of Keane (2014) [11] by addressing multiple challenges encountered in complex interactives. The framework accounts for objects that serve simultaneously as controls and readouts, a common feature in complex interactives. Additionally, the framework addresses the combinatorics of a multi-object system by creating a sophisticated system of modular descriptions, structured to provide access to both the current state of the interactive system and relevant changes made (and caused) by the user's interactions.

We anticipate the approach of providing parallel visual and non-visual displays will lead to new uses of description content. Once an interactive system is instrumented with auditory description display, components of the description content could be made available for additional purposes. For example, subcomponents of the State Descriptions could be used within the visual display – e.g., being able to toggle on and off a visual display of the names of important objects in the information space. Alternatively, interactive systems could be instrumented with an option to toggle on or off Context Responses delivered automatically (e.g., using the Web Speech API), to support users who do not have visual impairments and so can view the objects in the visual display, but could use support in interpreting the changes that occur as they interact with objects in an additional modality.

### Limitations of the Framework

The description design framework presented here has been successfully applied across multiple unique interactive web-based science simulations, but has not yet been applied in other contexts or with non web-based technologies. We invite others to make use of this design framework, and share their experiences and refinements. Additionally, this design framework accounts for interactive experiences in which the user is initiating all of the changes to the system. This approach may need to be augmented for interactive systems in which multiple users, or the system itself, initiates changes to the system.

## THE FUTURE OF AUDITORY DESCRIPTION DISPLAY

As the interface between humans and computers becomes more complex and varied, with increasingly data-rich presentations and ubiquitous computing experiences, we anticipate the role of description display will evolve, grow, and merge with other categories of auditory display. We foresee a future where technologies include rich auditory display experiences with layered and customizable description, sonification, and voice user interfaces [15] – allowing for robust conversational, situated, and transformative experiences. This work presents a step forward in the development of structures and guidelines for description display for complex interactives. Continued efforts at the forefront of auditory description display include the following topics across design delivery, customization, and professionalization.

### Design

While applying the design framework across multiple simulations, we frequently encounter interaction patterns common in interactive tools though significantly more complex than typical webpage or web form interactions. We have found it useful to develop interaction (and associated description) patterns. Once a description pattern has been designed and validated with users in one context (e.g., the silder in *Resistance in a Wire*), it can be applied or adapted to a similar or related context, increasing the efficiency of the description process and the consistency of the resulting described experience for users. While we have developed some interaction patterns, there still remain many interaction patterns common to complex interactives that have yet to be described and validated with users. Within our suite of simulations, we anticipate creating at least a dozen new design patterns, addressing non-standard – though common – drag and drop interactions, diverse types of button interactions (e.g., on/off switches, contextualized push buttons, and step buttons), complex menu interactions such as carousel-type selection, and others. Once complete, these design patterns will add to a growing corpus of described design patterns to be used as-is or adapted for future scenarios. Such a corpus would contribute to increasingly efficient description design and implementation.

Historically, auditory description display has been considered as an accessibility attribute to complement visual display. Description, however, has unique affordances that can complement or be complemented by other modalities – such as sound effects, sonification, natural language interfaces, visual display, and haptics. Research into the effective layering of modalities is needed to advance comprehensive and accessible interfaces.

### Delivery

Currently, resources with auditory description display typically deliver the descriptions using a live or pre-recorded audio track (e.g., theater, television, movies) or the description content is embedded/encoded using HTML and ARIA and accessed using assistive technologies such as screen reader software. Within complex web-based interactive technologies, the available delivery pathways require a combination of standard and custom solutions (see [36] for a summary) resulting in points of fragility. Additionally, existing solutions do not address the needs of non-web-based contexts, or mixed high- and low-tech interactive experiences.

### Customization

While typically considered for users with significant visual impairments, many users in a broad range of user contexts could potentially benefit from access to auditory description display. Prior work with young adults with intellectual and developmental disabilities [40] identified subcomponents of auditory description display as a potential solution to support these learners in using science simulations. More research is needed to better understand the breadth of use cases and users that would benefit from auditory description displays. We are investigating opportunities to customize a system that contains a fully described experience to deliver an abbreviated version, or a simplified version of the description design that could meet pre-existing and emerging access needs.

### Professionalization

Worldwide, only a handful of describers are tackling complex interactives, with no existing pipeline to train or expand this community of professionals. For web-based applications, efficient design of effective descriptions can be accomplished by describers with knowledge or significant interest in linguistics combined with expertise in HTML and ARIA. To contribute to the advancement of auditory description display design and to support the education of those interested in becoming describers, our group is launching (Fall 2020) a free online short course on auditory description display for interactive learning tools.

## CONCLUSIONS

Here we presented a systematic framework that can be used for the design of auditory description display for complex interactives. The framework includes detailed design and delivery attributes for an interactive system of descriptions, including State and Responsive Descriptions. We defined its components, illustrated how we use it, and demonstrated the capability of the resulting description display in supporting users in engaging with an interactive storytelling experience. This work presents a significant step forward in realizing the possibilities for designing enjoyable and engaging auditory description display that is capable of providing access and agency for users with disabilities in interactive digital environments, specifically users with limited to no vision.

## ACKNOWLEDGEMENTS

We thank the many participants who contributed their time and perspectives during interviews. We also thank Clayton Lewis, Jonathan Hung, and the PhET team for many conversations that contributed to the design of descriptions, and thus, the evolution of the description design framework. This material is based on work supported by the National Science Foundation under DRL-1503439, DRL-1621363, and DRL-1814220.

## REFERENCES


[1] Diagram Center. 2012. POET Training Tool, A Benetech Inititiative. (2012). https://poet.diagramcenter.org/index.html

[2] Diagram Center. 2015. Image Description Guidelines. (2015). http://diagramcenter.org/table-of-contents-2.html

[3] Joanmarie Diggs, Shane McCarron, Michael Cooper, Richard Schwerdtfeger, and James Craig. 2017. Accessible Rich Internet Applications (WAI-ARIA) 1.1. (Dec. 2017). https://www.w3.org/TR/wai-aria-1.1/

[4] M. Fakrudeen, M. Ali, S. Yousef, and A. H. Hussein. 2013. Analysing the Mental Modal of Blind Users in Mobile Touch Screen Devices for Usability. In Proceedings of the World Congress on Engineering, Vol. 2. IAENG, Springer, London, U.K., unknown.

[5] Steve Faulkner, Arron Eicholz, Travis Leithead, Alex Danilo, and Sangwhan Moon. 2017. HTML 5.2. (Dec. 2017). https://www.w3.org/TR/html52/

[6] Leo Ferres, Gitte Lindgaard, Livia Sumegi, and Bruce Tsuji. 2013. Evaluating a Tool for Improving Accessibility to Charts and Graphs. ACM Transactions on Computer-Human Interaction 20, 5 (Nov. 2013), 1–32. DOI:http://dx.doi.org/10.1145/2533682.2533683

[7] Donal Fitzpatrick, A. Jonathan R. Godfrey, and Volker Sorge. 2017. Producing Accessible Statistics Diagrams in R. In W4A '17 Proceedings of the 14th Web for All Conference on the The Future of Accessible Work. ACM Press, Perth, Australia, 1–4. DOI: http://dx.doi.org/10.1145/3058555.3058564

[8] Nahum Gershon and Ward Page. 2001. What storytelling can do for information visualization. *Commun. ACM* 44, 8 (Aug. 2001), 31–37. DOI: http://dx.doi.org/10.1145/381641.381653

[9] Andrew Holland. 2009. Audio Description in the Theatre and the Visual Arts: Images into Words. In Audiovisual Translation, Jorge Dìaz Cintas and Gunilla Anderman (Eds.). Palgrave Macmillan UK, London, 170–185. DOI: http://dx.doi.org/10.1057/9780230234581_13

[10] T. de Jong. 2006. COMPUTER SIMULATIONS: Technological Advances in Inquiry Learning. Science 312, 5773 (April 2006), 532–533. DOI: http://dx.doi.org/10.1126/science.1127750

[11] Kyle Keane. 2014. Interactive Scientific Graphics Recommended Practices for Verbal Description. Research. Wolfram Research, Inc., Champaign, IL. 53 pages. http://diagramcenter.org/accessible-dynamic-scientific-graphics.html

[12] Edward Kim and Kathleen F. McCoy. 2018. Multimodal Deep Learning using Images and Text for Information Graphic Classification. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18. ACM Press, Galway, Ireland, 143–148. DOI: http://dx.doi.org/10.1145/3234695.3236357

[13] Andrew Kirkpatrick, Joshua O Connor, Alistair Campbell, and Michael Cooper. 2018. Web Content Accessibility Guidelines (WCAG) 2.1. (June 2018). https://www.w3.org/TR/WCAG21/

[14] Sri H. Kurniawan, Alistair G. Sutcliffe, Paul L. Blenkhorn, and Jae-Eun Shin. 2003. Investigating the usability of a screen reader and mental models of blind users in the Windows environment. International Journal of Rehabilitation Research 26, 2 (2003), 145–147. http://journals.lww.com/intjrehabilres/Abstract/2003/06000/Investigating_the_usability_of_a_screen_reader_and.11.aspx

[15] Oussama Metatla, Alison Oldfield, Taimur Ahmed, Antonis Vafeas, and Sunny Miglani. 2019. Voice User Interfaces in Schools: Co-designing for Inclusion with Visually-Impaired and Sighted Pupils. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19. ACM Press, Glasgow, Scotland Uk, 1–15. DOI: http://dx.doi.org/10.1145/3290605.3300608

[16] Emily B. Moore, Taliesin L. Smith, and Emily Randall. 2016. Exploring the Relationship Between Implicit Scaffolding and Inclusive Design in Interactive Science Simulations. In Universal Access in Human-Computer Interaction. Users and Context Diversity, Margherita Antona and Constantine Stephanidis (Eds.). Lecture Notes in Computer Science, Vol. 9739. Springer International Publishing, Cham, 112–123. http://link.springer.com/10.1007/978-3-319-40238-3_12

[17] Elizabeth D. Mynatt and W. Keith Edwards. 1995. Audio GUIs: interacting with graphical applications in an auditory world. In Conference companion on Human factors in computing systems - CHI '95. ACM Press, Denver, Colorado, United States, 85–86. DOI: http://dx.doi.org/10.1145/223355.223441

[18] Donald A. Norman. 1987. Some observations on mental models. In Readings in human-computer interaction: a multidisciplinary approach, Ronald M. Baecker and William A. S. Buxton (Eds.). M. Kaufmann, Los Altos, Calif, 241–244.

[19] Tor Ole B. Odden and Rosemary S. Russ. 2018. Defining sensemaking: Bringing clarity to a fragmented theoretical construct. Science Education 103, 1 (Jan. 2018), 187–205. DOI: http://dx.doi.org/10.1002/sce.21452

[20] Jaclyn Packer, Katie Vizenor, and Joshua A. Miele. 2015. An Overview of Video Description: History, Benefits, and Guidelines. Journal of Visual Impairment & Blindness 109, 2 (March 2015), 83–93. DOI: http://dx.doi.org/10.1177/0145482X1510900204

[21] PhET Interactive Simulations. 2008. PhET Simulation Design Process. (2008). https://phet.colorado.edu/publications/phet_design_process.pdf

[22] PhET Interactive Simulations. 2015. Research and Design, Accessibility. (2015). https://phet.colorado.edu/en/accessibility/research

[23] PhET Interactive Simulations. 2019a. A11y View: Resistance in a Wire. (Dec. 2019). https://phet-dev.colorado.edu/html/resistance-in-a-wire/1.7.0-dev.4/phet/resistance-in-a-wire_a11y_view.html

[24] PhET Interactive Simulations. 2019b. Balloons and Static Electricity. (Oct. 2019). http://phet.colorado.edu/en/simulation/balloons-and-static-electricity Version 1.4.14.

[25] PhET Interactive Simulations. 2019c. Friction. (Oct. 2019). https://phet.colorado.edu/en/simulation/friction Version 1.5.10.

[26] PhET Interactive Simulations. 2019d. Gravity Force Lab: Basics. (Oct. 2019). https://phet.colorado.edu/en/simulation/gravity-force-lab-basics Version 1.0.0.

[27] PhET Interactive Simulations. 2019e. John Travoltage. (Oct. 2019). https://phet.colorado.edu/en/simulation/john-travoltage Version 1.5.12.

[28] PhET Interactive Simulations. 2019f. Resistance in a Wire. (Oct. 2019). http://phet.colorado.edu/en/simulation/resistance-in-a-wire Version 1.6.9.

[29] William Pike and Mark Gahegan. 2007. Beyond ontologies: Toward situated representations of scientific knowledge. International Journal of Human-Computer Studies 65, 7 (July 2007), 674–688. DOI: http://dx.doi.org/10.1016/j.ijhcs.2007.03.002

[30] William A. Pike, John Stasko, Remco Chang, and Theresa A. O'Connell. 2009. The Science of Interaction. Information Visualization 8, 4 (Jan. 2009), 263–274. DOI:http://dx.doi.org/10.1057/ivs.2009.22

[31] Noah S. Podolefsky, Emily B. Moore, and Katherine K. Perkins. 2013. Implicit scaffolding in interactive simulations: Design strategies to support multiple educational goals. arXiv preprint arXiv:1306.6544 (2013). http://arxiv.org/abs/1306.6544

[32] DCMP (Described and Captioned Media Program). 1994. Captioning Key. (1994). https://www.captioningkey.org/quality_captioning.html

[33] Jaime Sànchez, Nelson Baloian, Tiago Hassler, and Ulrich Hoppe. 2003. AudioBattleship: blind learners collaboration through sound. In CHI '03 extended abstracts on Human factors in computing systems - CHI '03. ACM Press, Ft. Lauderdale, Florida, USA, 798. DOI:http://dx.doi.org/10.1145/765891.765998

[34] Taliesin Smith, Michael Kauzmann, and PhET Interactive Simulations. 2019. Dynamic Phrase Manipulator. (June 2019). https://htmlpreview.github.io/?https://github.com/phetsims/a11y-research/blob/master/js/phrase-builder/phrase-builder.html

[35] Taliesin L. Smith. 2016. Access, Action, & Agency: Inclusive Design for the Non-visual Use of a Highly Interactive Simulation. Master Research Project. OCAD University, Toronto, Ontario. http://openresearch.ocadu.ca/id/eprint/713

[36] Taliesin L. Smith, Jesse Greenberg, Sam Reid, and Emily B. Moore. 2018. Parallel DOM Architecture for Accessible Interactive Simulations. In Proceedings of the Internet of Accessible Things on - W4A '18. ACM Press, Lyon, France, 1–8. DOI: http://dx.doi.org/10.1145/3192714.3192817

[37] Taliesin L. Smith, Clayton Lewis, and Emily B. Moore. 2016. Demonstration: Screen Reader Support for a Complex Interactive Science Simulation. In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility. ACM New York, NY, USA, Reno, Nevada, USA, 319–320. DOI:http://dx.doi.org/10.1145/2982142.2982154

[38] Taliesin L. Smith, Clayton Lewis, and Emily B. Moore. 2017. Description Strategies to Make an Interactive Science Simulation Accessible. Journal on Technology & Persons with Disabilities 5, 22 (2017), 225–238. http://scholarworks.calstate.edu/handle/10211.3/190214

[39] Volker Sorge, Mark Lee, and Sandy Wilkinson. 2015. End-to-end solution for accessible chemical diagrams. In Proceedings of the 12th Web for All Conference. ACM Press, Florence, Italy, 1–10. DOI: http://dx.doi.org/10.1145/2745555.2746667

[40] Brianna J. Tomlinson, Prakriti Kaini, Bruce N. Walker, Jared M. Batterman, and Emily B. Moore. 2018. Supporting Simulation Use for Students with Intellectual and Developmental Disabilities. Journal on Technology & Persons with Disabilities Volume 6, 2018 (2018), 202 – 218. http://hdl.handle.net/10211.3/202996

[41] Brianna J. Tomlinson, Bruce N. Walker, and Emily B. Moore. 2020. Auditory Display in Interactive Science Simulations: Description and Sonifcation Support Interaction and Enhance Opportunities for Learning. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. ACM, Honolulu, HI, USA. DOI:http://dx.doi.org/10.1145/3313831.3376886

[42] Hong Zou and Jutta Treviranus. 2015. ChartMaster: A Tool for Interacting with Stock Market Charts using a Screen Reader. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility. ACM New York, NY, USA, Lisbon, Portugal, 107–115. DOI: http://dx.doi.org/10.1145/2700648.2809862