Rodrigo, M. M. T. et al. (Eds.) (2021). Proceedings of the 29th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Design for Remote Embodied Learning: The Hidden Village-Online

Ariel FOGEL[a*], Michael SWART[a], Jennifer SCIANNA[b], Matthew BERLAND[b] & Mitchell J. NATHAN[a]

[a]*Educational Psychology, University of Wisconsin-Madison, USA*
[b]*Curriculum and Instruction, University of Wisconsin-Madison, USA*
* ajfogel@wisc.edu

Abstract: Digital embodied learning environments have increasingly become a subject of interest in the learning science community, offering insights into the nature of mathematical and scientific thinking. However, they have often relied on proprietary software and equipment to remain functional, creating obstacles for research and scaling up for widespread adoption in classrooms. In this paper, we highlight the affordances of redesigning existing embodied activities with ubiquitous technology in mind. Using *The Hidden Village* as a starting point, we demonstrate how the integration of open-source software and use of a standard, integrated webcam unlocks the ability of users to participate in embodied learning on any web-connected computer platform.

Keywords: Directed actions, embodied visual novel, action-cognition transduction, embodied cognition, remote learning

1. Introduction

Embodied learning leverages bodily interactions as central in the process of conceptual understanding and abstraction (Nathan, 2021). Advances in motion-based technologies have produced inexpensive options to incorporate embodiment into learning environments (Georgiou & Ioannou, 2019), providing students with novel opportunities to learn through integrated and high bodily-engaged tasks (Skulmowski & Rey, 2018).

In ongoing work, Nathan, Walkington, and colleagues (Kirankumar et al., 2021; Nathan & Swart, 2020; Nathan & Walkington, 2017; Swart et al., 2020; Walkington, Nathan, Wang, & Schenck, 2021) have been developing an embodied learning environment, *The Hidden Village* (THV). In THV, participants mimic movements of in-game avatars before responding to geometry conjectures. Experiments demonstrate that players who perform mathematically relevant gestures and movements show improved mathematical intuition, reasoning, and proof performance (Swart et al., 2020; Walkington et al., 2021). Despite promising early evidence of THV's impact on mathematics education, constraints tied to location and the availability and reliability of external hardware limit which types of gestures can be investigated and with whom. In this paper, we propose a design prototype to address the following research question: *How can the available platforms be expanded to support embodied experiences that reach students in varied educational settings, including for remote learning?*

2. Theory

Recent work has identified the broad role of bodily activity as a promising source of educational interventions (Skulmowski & Rey, 2018). Bodily activity has been shown to help with learning and intellectual performance in mathematics (Abrahamson et al., 2020; Goldstone et al., 2017; Walkington et al., 2019), reading, science, and other domains (Glenberg et al., 2004; Lindgren et al., 2016). Within this general form, scholars (e.g., Johnson-Glenberg et al., 2014) proposed attending to three aspects of embodiment: motoric engagement, gestural congruency, and perceived immersion.

Relatedly, a number of scholars have begun to identify an emerging set of design principles for building embodied learning interventions (Abrahamson et al., 2020; Lindgren & Johnson-Glenberg, 2013; Malinverni & Pares, 2014). One such finding is that people's gestures reflect their thinking. For example, when explaining the concept of parallel lines, a person may hold up their arms or hands next to each other. In other words, people generate semantically relevant body movements along with speech when engaged in cognitive tasks, such as mathematical reasoning and learning. *Gesture as Simulated Action* (GSA) is an empirically-based theoretical framework asserting that perceptual-motor processes are activated by spatial speech or thought, resulting in gestures when an activation level exceeds a certain threshold (Hostetter & Alibali, 2008; Hostetter & Alibali, 2019).

Recognizing the coupling between thinking and acting, Nathan (2017) asserted that people's performance of any action—including gestures—can induce cognitive states. This bi-directional relationship between thinking and acting is known as *Action Cognition Transduction* (ACT). ACT theorizes that eliciting movements can influence people's mathematical reasoning (Nathan, 2017; Thomas, 2013). Furthermore, *cognitively relevant* movements, or movements that map a concept to a set of body poses and movements and conform to gestural congruency (Johnson-Glenberg et al., 2014), will be beneficial for reasoning. Cognitively relevant movements in THV enact key mathematical relationships—for example, crossing their arms in front of their torso leads a player to create two sets of opposite angles.

Building on the theories of GSA and ACT, Nathan and Walkington (2017) developed the Grounded and Embodied Cognition framework, which proposed that directing players' bodily movements through *directed actions* complements learners' verbal expressions of mathematical reasoning. Additionally, recent work by Nathan and Swart (2020) suggests that people who produced dynamic depictive gestures, or gestures that explore the generalized properties of geometric objects and their simulated transformations, demonstrated superior performance in reasoning about geometric conjectures. Only cognitively relevant directed actions have been found to generate dynamic depictive gestures that complement mathematical reasoning (Walkington et al., 2021).

3. Design and Redesign of *The Hidden Village*

3.1 Original Game Design

The Hidden Village (THV) is an embodied learning environment structured as a *visual novel*, a video game sub-genre under interactive fiction. THV is designed to implement and capture evidence of learners' connections between movements, verbal reasoning, and mathematical thinking. As a research tool, THV allows us to investigate: (1) how mathematically-relevant directed actions influence students' proof practices, as measured by intuition, insight, and proof validity; (2) how pedagogical language, such as hints, connects players' actions to the mathematics and further influence the quality of students' proof practices; (3) the learning that occurs from enacting versus observing directed actions; and (4) how collaboration influences embodied mathematical reasoning (for students) and embodied lesson planning (for teachers).

3.1.1 Gameplay

In THV, players (top of Figure 1) proceed through a story composed of different challenges, with each following the same structure: players meet a villager in the story (bottom of Figure 1, a), then are asked to perform a set of directed actions displayed by an avatar (bottom of Figure 1, b), and finally are presented with a mathematical conjecture (bottom of Figure 1, c). To measure their mathematical intuition, players are prompted to make a True/False "snap" judgement about the mathematical conjecture. Players are then prompted to provide additional insights and rationales to justify their mathematical intuition through their multimodal responses, comprising speech and gestures. Following this exercise, players are formally assessed with a multiple-choice question related to the conjecture (bottom of Figure 1, d), after which they receive a token of knowledge and access to more of the Village Map (bottom of Figure 1, e) before moving to the next challenge.

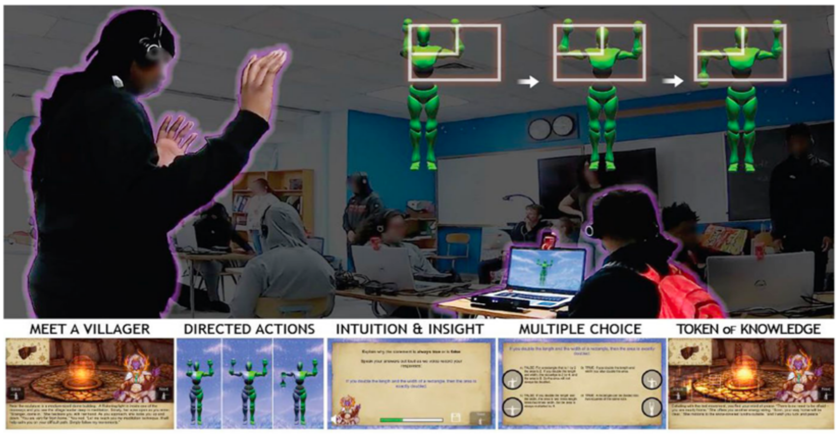

Figure 1. *Top*. Images of THV use in a high school classroom, with overlays in the top right showing the directed actions. *Bottom*. Sequence of gameplay of *The Hidden Village*. (a) A villager invites the player to perform a task. (b) The task has a series of 3 directed actions. (c) Player is then asked to judge the truth of a conjecture, (d) answer a multiple choice question, and (e) receive a token of knowledge and access to more of the village map.

3.1.2 Technology

The Hidden Village leverages players' movements for both navigation as well as core gameplay using infrared 3D-motion-capture through Microsoft Kinect™ hardware. The Kinect™ tracks the positions of key landmarks on a player's body (e.g., wrists, elbows, shoulders) and calculates the angle between landmarks to determine whether players have positioned their bodies in the correct poses (i.e., the directed actions).

3.1.3 Embodied Interface and Feedback

When prompted to perform a set of directed actions, the player is presented with the avatar performing the first pose of the sequence (Figure 1b) and is asked to match that pose with their own body. After a successful match is established using sensor technology, the avatar animates to the next pose, and so on. Markers on the avatar change from red to green when the player matches the pose. Once the player adequately recreates the avatar's pose (using adjustable tolerance levels), the avatar advances to the next pose in the sequence. Directed action sequences are typically repeated 3 to 5 times, a parameter set by the teacher or researcher.

3.2 Challenges

At the time of its inception, the Microsoft Kinect™ attracted the attention of many researchers interested in developing embodied learning environments as classroom research tools (Georgiou & Ioannou, 2019). However, classroom experiments have revealed several substantial constraints associated with using sensors like the Kinect.

3.2.1 Technology

Relying on dedicated peripheral hardware endangers the longevity of THV, as it requires continuous driver and coding updates to ensure compatibility for functional UX design. Between 2010 and 2020, rapid developments in extended reality (XR) technologies (i.e. augmented reality and virtual reality) caused several peripheral XR hardware projects to be discontinued, including the Microsoft Kinect, Oculus Rift, Google Glass, and Leap Motion controller. Continued use of these outdated technologies has required researchers to seek costly and specialized software development to advance their research agendas. THV's viability in research and classroom settings has likewise been hindered by Microsoft's discontinuation of the Kinect in 2015, which led to delays in development and inconsistent performance.

Predicting which XR hardware will continue to be developed and which will be discontinued remains an impossible task for research teams to manage.

Moreover, both the labor required to seamlessly set up THV in classrooms and the cost of equipment have proved costly. Each participant was provided with their own kit, comprised of a laptop, Kinect sensor, recording devices, and electrical stations. This meant that a dedicated operating surface and space of about 6 feet by 8 feet had to be reserved for each player engaged in the intervention. Positioning these kits strategically in an averaged-sized classroom for 12 students to play concurrently proved to be a significant logistical challenge for classroom management, fidelity of treatment, data collection and data management.

Finally, the onset of COVID-19 and the resulting school closures limited student access to school-based technology, moving schooling (ergo research) entirely online. Thus, it was imperative for the research team to develop a new version of THV that could be administered remotely and still deliver effective embodied curricula for school-aged children.

3.3 Proposed Solutions

3.3.1 Technology

To address these constraints, we present a proof-of-concept design for an in-browser version of THV, aptly called *The Hidden Village-Online* (THV-O). The engine of THV-O is built in Javascript using the open-source Monogatari visual novel framework (Ocampo, n.d.). In lieu of dedicated hardware to perform pose detection, this new version leverages computer webcams and MediaPipe (Lugaresi et al., 2019), an open-source computer vision algorithm developed by Google.

The interface introducing a directed action to players in the visual novel was also redesigned in THV-O. The right side of the screen displays the real-time detection of the player's body, while the left side of the screen presents the poses for prompting the directed action (Figure 3). The pose highlighted by a green border indicates which pose the player should currently match.

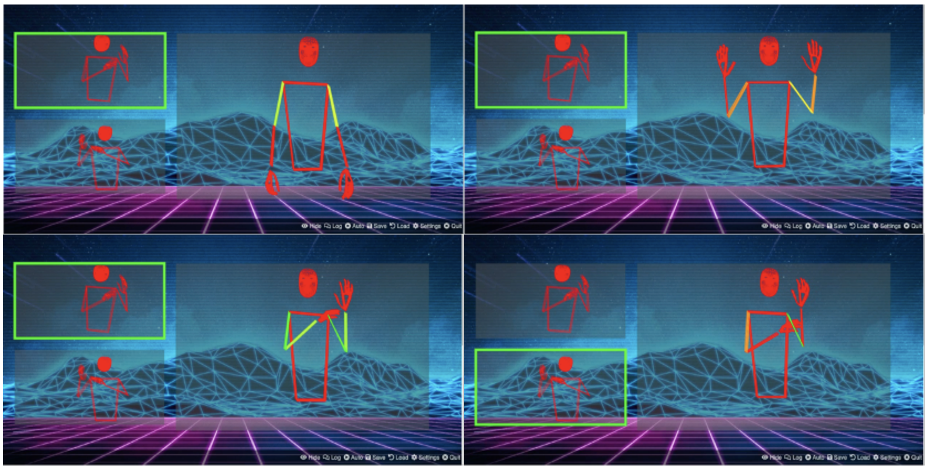

Figure 3. A sequence of the directed action instruction interface in four panels. Players' arms are green when their position matches the arm position of the figure inside the green border and red when their arms do not match (upper left). As the position of the arm gets closer to the desired pose, colors change on the red-green spectrum (upper right). Once the pose matches, the visual feedback on the player's body turns green (bottom left) and the game updates the next pose to match (bottom right).

4. Discussion

Moving THV online addresses a host of monetary, user experience, and logistical issues compared to its predecessor. First, it untethers THV from its dependence on dedicated peripheral hardware, eliminating substantial monetary expenditures and minimizing the amount of space required to administer research activities. Second, it obviates the need for troubleshooting hardware and using third-party software for group participation. Third, the ubiquity of computers equipped with internet browsers and webcams increases the accessibility and scaling potential of our interventions. In effect, THV-O is able to offer an embodied curriculum to any classroom that visits the digital learning environment online. Additionally, building THV-O with open-source software and using a programming language supported by all modern browsers lowers the projected costs for future development, thereby reducing the likelihood of THV-O obsolescence.

Another advantage to migrating THV to an online platform is the ability to change pose detection. These improvements are driven by a central claim of the Grounded and Embodied Cognition framework: performing cognitively relevant directed actions transduces players' cognitive processes in ways that can increase the likelihood they engage in the appropriate mathematical reasoning. Since the directed actions in THV are conceptualized as a series of dynamic movements, students' abilities to faithfully reproduce them can influence the potential mathematical insights generated.

Additionally, improvements in pose detection broaden the variety of actions that can be performed in the learning environment. While the Kinect is only capable of detecting gross motor movements of the limbs (e.g., elbows, shoulders, and wrists) and lacks the fidelity of detection that allows players to use their hands, MediaPipe's Holistic solution performs real-time detection of both a player's body pose as well as their hands and fingers. Hand shape detection expands the repertoire of cognitively relevant actions and ways of documenting players' spontaneous gestures, thus greatly increasing both the scope of in-game movements and range of real-time, multimodal data.

As with all research, this work has several limitations. First, detecting poses using webcam images rather than using hardware with multiple sensors means that poor lighting or low-resolution webcam images may impact the precision of the pose detection. We anticipate that improvements to computer vision algorithms and the increasing affordability of higher resolution webcams will address this limitation in future work. Second, the computer vision algorithm used by THV-O works best when detecting the pose of a single individual, limiting current research to the investigation of directed actions on mathematical reasoning of individual students. Future work will explore the use computer vision algorithms capable of detecting the poses of multiple individuals in the frame, enabling the use of collaborative gestures (Walkington et al., 2019) in directed actions. Finally, using computer vision algorithms in a browser is still a resource-intensive operation, meaning some machines may have difficulty running THV-O. Future versions of THV-O will seek to use computer vision algorithms that can be run on computers with low computational resources.

Despite these limitations, we believe the design of THV-O represents a significant step forward on the path towards integrating principles of embodied learning in mathematics education and education research. The improvements detailed in this paper unlock future research directions that, up until now, have been limited by technological constraints, while simultaneously lowering barriers for learners to access embodied learning environments. By scaling up user access, expanding the scope of the curriculum through end-user content creation, and improving the ease of use of ubiquitous resources, embodied learning technologies such as THV-O are poised to extend the reach of learning environment design and advance our understanding of the nature of learners' conceptual development.

Acknowledgements

We would like to thank everyone who prepared and revised previous versions of this document, especially Pauley Tedoff. Research reported here was supported by the Institute of Education Sciences and the U.S. Department of Education through IES Grant R305A160020, awarded to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the IES or the U.S. Department of Education. This work was also supported by generous funding from the James S. McDonnell Foundation.

References

Abrahamson, D., Nathan, M. J., Williams-Pierce, C., Walkington, C., Ottmar, E. R., Soto, H., & Alibali, M. W. (2020). The future of embodied design for mathematics teaching and learning. *Frontiers in Education*, *5*, 147. https://doi.org/10.3389/feduc.2020.00147

Georgiou, Y., & Ioannou, A. (2019). Embodied learning in a digital world: A systematic review of empirical research in K-12 education. In P. Díaz, A. Ioannou, K. K. Bhagat, & J. M. Spector (Eds.), *Learning in a Digital World* (pp. 155–177). Singapore: Springer Singapore.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*(3), 424–436. https://doi.org/10.1037/0022-0663.96.3.424

Goldstone, R. L., Marghetis, T., Weitnauer, E., Ottmar, E. R., & Landy, D. (2017). Adapting perception, action, and technology for mathematical reasoning. *Current Directions in Psychological Science*, *26*(5), 434–441. https://doi.org/10.1177/0963721417704888

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, *26*(3), 721–752.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology*, *106*(1), 86–104. https://doi.org/10.1037/a0034008

Kirankumar, V., Sung, H., Swart, M., Kim, D., Xia, F., Kwon, O. H., & Nathan, M. J. (2021). Embodied transmission of ideas: Collaborative construction of geometry content and mathematical thinking. *CSCL 2021*. Presented at the 14th International Conference on Computer Supported Collaborative Learning, Bochum, GER.

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*, *42*(8), 445–452. https://doi.org/10.3102/0013189X13511661

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., … Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. *CoRR*, *abs/1906.08172*. Retrieved from http://arxiv.org/abs/1906.08172

Malinverni, L., & Pares, N. (2014). Learning of abstract concepts through full-body interaction: A systematic review. *Journal of Educational Technology & Society*, *17*(4), 100–116.

Nathan, M. J. (2017). Chapter 8. One function of gesture is to make new ideas: The action-cognition transduction hypothesis. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Gesture Studies* (Vol. 7, pp. 175–196). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/gs.7.09nat

Nathan, M. J. (2021). *Foundations of embodied learning: A paradigm for education*. New York, NY: Routledge.

Nathan, M. J., & Swart, M. I. (2020). Materialist epistemology lends design wings: Educational design as an embodied process. *Educational Technology Research and Development*.

Nathan, M. J., & Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive Research: Principles and Implications*, *2*(1), 9. https://doi.org/10.1186/s41235-016-0040-5

Ocampo, D. I. (n.d.). Monogatari Visual Novel Framework (Version 2.0).

Skulmowski, A., & Rey, G. D. (2018). Embodied learning: Introducing a taxonomy based on bodily engagement and task integration. *Cognitive Research: Principles and Implications*, *3*(1), 6. https://doi.org/10.1186/s41235-018-0092-9

Swart, M., Schenck, K., Xia, F., Kwon, O. H., Nathan, M., Vinsonhaler, R., & Walkington, C. (2020). *Grounded and embodied mathematical cognition for intuition and proof playing a motion-capture video game*.

Thomas, L. E. (2013). Spatial working memory is necessary for actions to guide thought. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1974–1981. https://doi.org/10.1037/a0033089

Walkington, C., Chelule, G., Woods, D., & Nathan, M. J. (2019). Collaborative gesture as a case of extended mathematical cognition. *The Journal of Mathematical Behavior*, *55*, 100683.

Walkington, C., Nathan, M. J., Wang, M., & Schenck, K. (2021). The effect of cognitive relevance of directed actions on mathematical reasoning. *Manuscript Currently under Review*.