# A Comparison of Augmented Reality and Browser Versions of a Citizen Science Game

Kutub Gandhi
Northeastern University
gandhi.ku@northeastern.edu

Josh Aaron Miller
Northeastern University
miller.josh@northeastern.edu

Sofia Eleni Spatharioti
Northeastern University
spatharioti.s@northeastern.edu

Archana Apte
Northeastern University
archana2013apte@gmail.com

Borna Fatehi
Independent Scholar
borna.f@gmail.com

Sara Wylie
Northeastern University
s.wylie@northeastern.edu

Seth Cooper
Northeastern University
se.cooper@northeastern.edu

## ABSTRACT

Science is increasingly being achieved through citizen science (CitSci) — research conducted in part by non-professional scientists. In order to increase recruitment and retention, CitSci is sometimes gamified. Yet, research suggests that being a game does not alone make CitSci more enticing; rather, retention is dependent on the player's experience. One recent trend in gaming is Augmented Reality (AR), as demonstrated by games like *Pokémon Go*. In this work, we examine the potential benefits of applying AR to CitSci gaming. We conducted a user study where we invited participants to play two citizen science adaptations (browser and AR) of the traditional game of Memory. We identified four major themes through thematic analysis. We conclude that AR is a leisurely toy: while the browser interface is better suited to rapid, accessible, productive play, AR is better suited to reflective, one-time, unstructured experiences. Our findings reveal potential towards future work on exploring AR in contexts such as museums and classrooms.

## CCS CONCEPTS

• **Human-centered computing**;

## KEYWORDS

augmented reality, citizen science, games

**ACM Reference Format:**

Kutub Gandhi, Josh Aaron Miller, Sofia Eleni Spatharioti, Archana Apte, Borna Fatehi, Sara Wylie, and Seth Cooper. 2021. A Comparison of Augmented Reality and Browser Versions of a Citizen Science Game. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021 (FDG'21), August 3–6, 2021, Montreal, QC, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3472538.3472557



*FDG'21, August 3–6, 2021, Montreal, QC, Canada*



ACM ISBN 978-1-4503-8422-3/21/08…$15.00
https://doi.org/10.1145/3472538.3472557

## 1 INTRODUCTION

In the last two decades, scientific fields like biology and conservation research have seen a huge rise in support from citizen science (CitSci) — research conducted in part by non-professional scientists [12, 21]. Through CitSci projects like iNaturalist,[1] a wide population of people can contribute to scientific advancement. While CitSci encompasses a variety of projects, one of the most common tasks of CitSci is image classification [16, 33, 36] because of its low cost and complexity [35].

However, CitSci routinely struggles with recruitment and retention. Drop-outs are a common occurrence [9] and projects typically rely on "power users" — dedicated contributors — who account for the vast majority of contributions in a given project [31].

Out of a desire to improve engagement and provide a more enjoyable experience to participants, several CitSci projects have organized themselves as games in order to improve retention through gamification [3, 15]. Recent research on CitSci games suggests that the retention problem remains unsolved, however, as retention is partially dependent on the player's experience (a combination of usability and playability [30]) [6].

One recent trend in gaming is alternative play modalities, such as Augmented Reality (AR) and Virtual Reality (VR) [2, 27]. AR technology (overlaying virtual graphics onto camera displays in real-time) is increasingly used in mobile games; for example, the worldwide gaming phenomenon *Pokémon Go* [26] demonstrated that AR games have mainstream appeal. Other tools such as iScape's garden planner[2] and Warby Parker's eyeglasses testing app[3] show that AR is useful for the average user, despite the technology's infancy.

AR also has powerful benefits for serious and educational games [20]. Games such as ExCora [22] show that AR can increase exploration of and appreciation for the natural world. Furthermore, there is evidence that AR games can be more engaging for children than similar physical games [17], and AR augmentations (such as guides and gamification elements) can improve the play experience of physical games [37].

[1] https://www.inaturalist.org/
[2] https://www.iscapeit.com/
[3] https://www.warbyparker.com/prescription-check-app

Because of the great potential of AR, an AR CitSci tabletop game toolkit called Tile-o-Scope AR (TOSAR) is being developed in order to explore if and how AR can improve engagement in image classification CitSci games. TOSAR is a flexible toolkit that uses an AR app on a mobile device to display images on physical tiles with AR markers. A variety of games can be played by making matches between tiles that have images in the same category (e.g., birds or mammals). Some images have known categories (i.e., ground truth); by matching ground truth to non-ground truth data, the players provide information for categorizing the images.

A previous exploratory design study on TOSAR revealed that groups of players from different backgrounds (i.e., game design, sociology, or environmental health expertise) have different gaming preferences [33]. Yet, despite the social benefits of physical, co-located play, TOSAR's AR technology may potentially distract participants from their contribution efforts via overlapping labor and focusing on the game instead of the citizen science content [33].

Therefore, extending this preliminary research, we examine how AR CitSci compares to non-AR CitSci on areas of engagement and effectiveness. To make this comparison, we developed a non-AR version of one of the games (Memory) that can be played alongside TOSAR. We refer to the non-AR version as the "browser" version; this version was developed as a separate game mode on Tile-o-Scope Grid [34]. Both Tile-o-Scope Grid and TOSAR are part of a wider image classification platform called Cartoscope,[4] and have been used for a variety of image sets ranging from flooding and infrastructure damage to animal classification to algal blooms. In the present study, we compared AR and browser versions of the game Memory using an iNaturalist data set on animals native to Boston.

As this work was exploratory, our investigation was driven by the following research questions:

RQ1. How do these modalities (AR and browser) differ in engagement?
RQ2. Do these modalities fill different play niches, or do players engage similarly with both?
RQ3. How do these modalities differentially interact with the CitSci experience?

This work provides one major contribution: the qualitative comparison of AR and non-AR (browser-based) modalities as they affect the CitSci gaming experience for single-player games. We provide this contribution in the form of four themes generated through thematic analysis: (1) The browser game is more familiar, fast, and accessible; (2) Physicality slows AR play; (3) AR affords and demands spatial processing; and (4) The AR novelty has toy qualities.

We thus conclude that AR is a leisurely toy, defined as suited to unstructured, slower and learning oriented play rather than rapid, accessible, productive play. In the browser game, players took roughly a quarter of the time to make each match as compared to the AR game. Furthermore, participants described the AR experience as more reflective and thoughtful with respect to CitSci. We therefore recommend future work explore AR in appropriate contexts such as museums and classrooms rather than as a replacement to traditional non-AR CitSci gaming. Future work should also further investigate co-located multiplayer AR in contrast with other multiplayer modalities as a method of critical engagement with CitSci.

[4] http://cartosco.pe/

## 2 BACKGROUND

Similar to AR, several recent applications have explored integrating physical components. BacPack [23] uses tangible tokens in a multi-touch tabletop museum exhibit to increase collaboration and discussion. Similarly, the SynFlo exhibit [28] uses Sifteo cubes to provoke peer collaboration in a tangible, playful museum activity.

AR games have been compared to non-AR games specifically in educational contexts. In two studies that compared player experiences, players preferred the AR version [11, 18]. In another study AR was described as being more costly to implement [8].

Each of the prior works described so far examined a specific AR or physical application with predefined rules. ARToolKit [14] on the other hand is a platform for prototyping in AR. However, this platform requires significant effort for the end-user to create a playable activity.

Unlike the previous applications mentioned which have high costs (SynFlo), come with a loss in portability (BacPack), or require programming (ARToolKit), TOSAR is a low-cost, portable platform for combining multiple image labeling games in a single platform [33].

AR has been recently applied to CitSci as well, such as for environmental monitoring [1], annotating images of mountains [10], and flood risk assessment [13]. AR may also have value in improving disaster management and visualizing environments [32]. A review of emerging technologies in CitSci describes the purpose of AR as "Creat[ing] an immersive experience to augment or replace real world environments." [24]. From this perspective, several researchers see the value of AR in CitSci as the ability to view real environments with augmented user interfaces for guiding the citizen scientist through a task. In this study, however, we focus on the gaming potential of AR as a novel technology for providing playful experiences.

## 3 METHODS

### 3.1 Games

We compared an AR and browser version of a simple Memory game.[5] In Memory, players flip over pairs of tiles and collect the tiles as points if they match. If they do not match, players flip the tiles back over while trying to remember their positions. Skilled players who remember the positions of tiles can collect points with fewer flips than a player who selects tiles at random.

In order to use Memory as a CitSci game, we made some minor variations from the standard Memory game. Namely, players are not looking to match duplicates of the same image, but are instead looking to match images of the same category. The image set used in this study was a collection of 24 animal images from the Boston area (12 species, 2 images per species) provided by iNaturalist. The categories for this image set were the individual species, resulting in 12 matches (both images from the same species). We refer to the two versions of the game implemented as MemoryBrowser and MemoryAR.

[5] https://en.wikipedia.org/wiki/Concentration_(card_game)

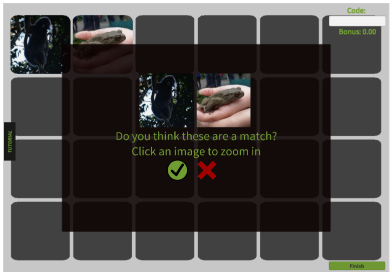


**Figure 1: Screenshot of MemoryBrowser. The player has just selected two tiles to reveal a monkey and a toad, and the game is prompting them to select whether the animals match.**


**Figure 2: Screenshot of MemoryAR. The player has physically flipped two tiles to reveal images of birds. The player can press the camera button to mark the two images as a match.**

*3.1.1 Image Labeling.* Recall that some images in each set have "ground truth" — known, verified labels — although players do not know which images these are. In this way, matching a known image to an unknown image provides some confidence on the category of the unknown image. When a player matches two unknown images, the game prompts them to provide a category label for both images. This results in four possible outcomes for a match:

- If the player matches **two known images of the same category**, they are told the match is **correct** and rewarded points.
- If the player matches **two known images of different categories**, they are told the match is **incorrect**.
- If the player matches **a known and unknown image**, they are rewarded points for a **new, label-assumed** match.
- If the player matches **two unknown images**, they are prompted to provide a category label, and then they are rewarded points for a **new, label-selected** match.

Audiovisual feedback is given to participants depending on the outcome of their match.

Although ground truth images would be typically seeded randomly into image sets, for the purpose of this study the image set was constructed with a specific amount of ground truth images.

In MemoryAR, there was exactly one ground truth and one unknown image per category,[6] though the unknown images were known to the researchers for the purpose of measuring participant accuracy. In MemoryBrowser, due to a technical limitation, all images had ground truth. Thus, only correct and incorrect matches were possible in the browser version. Although some participants noticed that the AR version prompted them to choose a category for new, label-selected matches, their experience did not appear strongly affected by this difference; therefore, we assume the difference in implementation is negligible for the purpose of the study.

*3.1.2 Game Details.* MemoryBrowser is implemented as a separate game mode in the browser game Tile-o-Scope Grid [34]. The player is presented with 24 tiles in a $6 \times 4$ grid. Clicking any tile reveals the image hidden in the tile, and clicking another tile prompts the player on whether it matches the first. If the player agrees, the game provides feedback based on the match result as described above. If a match is correct, the tiles are removed from play. An image of MemoryBrowser is shown in Figure 1.

MemoryAR is played with Tile-o-Scope AR (TOSAR). It uses a set of 24 paper tiles with AR tags on one side. These paper tiles are shuffled, placed face-down, and used just as they would be in the standard Memory game by flipping them over. Players use the TOSAR Android application to scan the AR tags, which replaces them (on the screen of the Android device) with the iNaturalist images. Once the two images are in view, the player can press a button to capture the match, which then provides feedback based on the match result as described above. One notable feature of TOSAR is that the app allows selection of one to four images to make a match (despite matches in Memory being exactly two images). This is because, as a toolkit, TOSAR is intended for other games beyond Memory. Though, as a result, a small portion of matches in this study were of one, three, or four images; these data were included for the purpose of gameplay metrics analysis as shown in Table 1. An image of MemoryAR is shown in Figure 2.

## 3.2 Pilot

In order to ensure a valid methodology, we conducted a pilot study with 22 volunteers known to the researchers. The pilot study was performed similarly to the study procedure as described below, but with different image sets, UI elements, and interview questions. From these pilot data, we determined that the choice of image set affected the difficulty of the game, and thus engagement; participants found image sets with fewer categories easy and boring, while image sets with blurrier or complex images were difficult to identify and frustrating to play with. We determined that the iNaturalist image set used in the primary study provided a good balance of difficulty.

Moreover, we determined from pilot results that we would be unable to accurately measure the time participants would spend with the game. Players often claimed that they had finished playing because of time constraints[7] or because they believed they had

[6]Yet, this creates an edge case where — if a player matches two unknown images of different categories — this can result in an unwinnable game, i.e., with only non-matching ground truth remaining. Unwinnable games due to unmatchable tiles are also present in the app normally. During pilot testing, this edge case created confusion and feelings of playing the game incorrectly; therefore, during the study participants were informed of this possibility. We intend to look into more natural solutions to this issue for versions of MemoryAR outside of this study.

[7]We did not impose time constraints during the pilot, however many participants had obligations after the study or generally did not want to spend an excess of time in the study.

a good enough understanding of the game to answer interview questions. In combination with the fact that our study parameters differ from the naturalistic settings in which players would find and play these games, we decided that for the primary study we would not record time played as a measure of engagement; instead, we would prompt the participant to move on after a designated amount of time in order to keep session lengths consistent.

## 3.3 Participants

Participants were recruited through invitations sent to online communication channels accessible to the researchers, including: personal contacts, social media pages, CitSci communities, gaming communities, and university groups. Interested people were directed to an informed consent information sheet and a scheduling form. Recruitment resulted in a total of 24 participants (13 male, 10 female, 1 non-binary). In regards to age, 19 participants were aged 18-24, 3 participants were age 25-34, 1 participant was aged 45-54, and 1 participant was aged 65-74. Participants were offered a $15 (USD) Amazon gift card. The study was approved by the Institutional Review Board of the researchers' institution.

## 3.4 Procedure

During initial recruitment, participants scheduled a video conference session with one of the researchers. At the start of this video session, consent was confirmed with the participant. Then the researcher provided background information on CitSci, TOSAR, and Memory. Participants were randomly assigned (using block randomization) to either MemoryBrowser or MemoryAR to play first; they were told that they could play as long as they wanted; however, to restrict the scope of the study, the researcher warned them about time if they played longer than 15 minutes and prompted them to move on if they played longer than 20 minutes. Players were given the rules of Memory, but these rules were not enforced by the researcher in order to allow participants to modify their play experience. After playing, participants were asked about their play experience through a semi-structured interview. Topics included general enjoyment, game feel, replayability, and novelty. Finally, players filled out three subscales from the Intrinsic Motivation Inventory (IMI) survey [25, 29]: Interest / Enjoyment, Effort / Importance, and Value / Usefulness for CitSci.

This procedure was then repeated with the other game (MemoryAR or MemoryBrowser) for within-subjects comparison. During play, gameplay actions were automatically logged. Finally, participants were offered remuneration and debriefed.

## 3.5 Analysis

Study sessions were recorded and analyzed via reflexive thematic analysis [4, 5]. One author coded the data using semantic coding and an essentialist framework to capture the participants' experiences with each game. Two authors then discussed the codes to generate themes. The analysis was driven by the research questions listed earlier.

A "Big Q" Qualitative approach was taken towards analysis with a focus on exploration and theme generation rather than numerical analysis [19]: no inferential statistical tests were performed on the IMI subscales or gameplay data; however, descriptive statistics are provided in the Results section (see Table 1), including positivist characterizations of participant responses for the sake of contextualizing the Qualitative analyses.

# 4 RESULTS

## 4.1 Thematic Analysis

When asked directly, 17 of the 24 (71%) participants reported that they enjoyed MemoryAR over MemoryBrowser for a variety of reasons. Through reflexive thematic analysis, four themes were generated to describe the participants' experiences. These themes can be summarized as **AR is a leisurely toy**. Whereas the browser experience was familiar, fast, and accessible, the physicality and spatiality of AR hindered completion of the task. Yet, these same properties added novelty and toy qualities to the AR experience, making it more enjoyable or easier to play with for some participants. We conclude that AR may therefore be useful for communal locations that focus more on learning than rapid data analysis (e.g. museums and classrooms), while non-AR interfaces remain more suited to accessible, single-player CitSci games.

We describe the themes here below.

*4.1.1 The browser game is more familiar, fast, and accessible.* Many people already have familiarity with browser games, and this background knowledge transfers to understanding how to interact with MemoryBrowser (e.g., P18, P19, P20). Participants referred to MemoryBrowser as a *"simple, intuitive matching game"* (P7) or a *"simple, monotone [monotonous] game"* (P2). Conversely, few participants were familiar with the play experience of AR: only one participant mentioned extensive experience with AR games and AR functionality; every other player had limited experience through games like *Pokémon Go* [26] or no experience at all. This unfamiliarity contributed to the novelty of AR, but also to its difficulty in getting started. In addition to AR's physicality, this created a significant learning curve compared to the browser modality.

> *"You had to get used to [MemoryAR], right, you had to get used to the AR part of the game, flipping the tiles over, taking your phone there."* (P7)

> *"I just felt like the AR was too, like I got it, but it was an unnecessary curve. This one [MemoryBrowser] for a lot of the same thing like, yeah... [non-AR] is an easier path.*[8] (P8)

Many participants also enjoyed the faster play speed afforded by the browser version's ease-of-use. Notably, participants were able to make matches in roughly a quarter of the time, on average, in MemoryBrowser compared to MemoryAR (see Table 1).

> *"As I was playing [MemoryAR], I was thinking, there could be a faster way of doing this."* (P16)

> *"Ugh it took so much longer though, right? Like just physically turning over those papers and just my setting and my environment was... ugh."* (P21)

[8]For clarity, we paraphrase and rewrite the participants sentiment here: "I felt that the AR version was too complex. I understood it, but it had an unnecessary learning curve. The browser version had a lot of the same things and had an easier path to understanding.""

Lastly, unlike AR, the browser game doesn't require a physical play space or other game components. This makes it more accessible because of its availability in more contexts, and participants often noted this. The ability to quickly begin a game (requiring much less set-up) made the browser version a more appealing modality for casual play.

> *"If there was a PC version of the game on the phone where all you had to do was click on the screen and the tiles would flip on their own without having to carry this physical paper with QR codes, putting them on a table, and going to scan them... [that would be more convenient]"* (P7)

*4.1.2 Physicality slows AR play.* Approximately half of the participants (11 of 24) commented specifically on enjoying the tactility of AR play while the other half (10 of 24) expressed negative feelings about the physicality. However, in all cases the physicality slowed their play. In the best case, this led to more reflective, critical processing, but many participants reported tedium, frustration, and potentially physical discomfort (P2, P21).

The first physical barrier is setting up the game: printing out the tiles and laying out the grid (e.g., P7, P10, P17). When it came time to actually play the game, participants commented on how, in the browser version, the entire board was within one screen, but in AR, they could not compare two tiles on opposite sides of the grid simultaneously.

To resolve this dilemma, players took a variety of creative approaches. Some rearranged the board to move similar tiles together. Others stood up or lifted their phone higher so that the AR camera could see the entire grid. Most commonly, though, players temporarily moved the two tiles they wanted together in front of them, then moved them back after the match.

The AR physicality caused several major frustrations. Many players were frustrated with the slowness of flipping tiles, while many others struggled with holding their phone and tiles simultaneously. A few players even had to crane their necks to see the game pieces and camera.

> *"It took a second to even scan the images."* (P7)

> *"I'm definitely going to have to do some face and neck yoga after this."* (P21)

> *"We only have two hands... it required some dextrousness to move the tiles."* (P19))

> *"[MemoryAR] seemed a lot of trouble to go through to match two-dimensional images. It felt like I was doing the Rube Goldberg version of a robot check [captcha]."* (P11)

Despite AR's low throughput, there are benefits to slow play. Citizen science projects have a variety of goals including education, raising awareness, and provoking conversation. While AR could be a good fit for those goals, the low throughput and physical frustrations imply that it may be a poor modality for purely productive play.

*4.1.3 AR affords and demands spatial processing.* Because of the physicality and spatiality of AR, MemoryAR demanded more spatial processing. For some players, this added cognitive load: trying to handle a phone and tiles simultaneously and remember where everything was while playing a game and performing CitSci tasks was entirely overwhelming.

> *"I have to focus on two things instead of getting the accuracy down."* (P22)

On the other hand, this spatiality was also an affordance, and some participants took advantage of this to reduce cognitive load. They would move tiles to an easy-to-remember location, take note of small physical imperfections in the tiles, or use their fingers to mark certain tiles in their memory. In this way, because AR existed in physical space, they could take advantage of that to assist their play.

> *"The tiles felt more real to me."* (P21)

> *"Having the physical thing in front of me made me better at it."* (P17)

> *"It was intuitive to use [the physical movement of the tiles as a manual zoom] to get a closer look."* (P18)

Finally, while players were given the rules of Memory, these rules were not enforced: players took advantage of the control they had over the game space to play in ways that they enjoyed. As argued in prior work [33], this is a feature distinct to AR since the browser version can use only pre-programmed game rules. In MemoryAR, players modified their game in several ways ranging from quality-of-life changes such as breaking the grid of the game (P12), to minor rule variations such as keeping one tile up and looking for its match, to complete overhauls such as flipping every tile over and sorting the tiles into categories. Players spoke positively about the feeling of control they had over their play experience (e.g., P3, P21) that the generalized AR platform afforded. To recreate a similar experience on a browser platform would require extensive work; in this way, AR is reminiscent of tabletop games that can be modified at will to suit the player's desired experience.

> *"[MemoryAR] enabled me to cheat."* (P23)

The player did not cheat, but was referencing a desire to flip all the tiles over.

> *"I enjoyed that I could move the tiles around to keep track of the pairs."* (P24)

> *"I can move them around on my desk to physically sort them."* (P13)

Notice how several participants suggested a desire to simply sort the tiles. This too is an affordance of the AR platform, and the customization capabilities of the TOSAR toolkit works in favor of participants contributing to the CitSci task in whichever way they would like. By placing the game in a physical space and not asserting a given rule set, AR allows for varied play.

*4.1.4 The AR novelty has toy qualities.* AR was, unsurprisingly, a novel experience for many participants.

> *"[MemoryBrowser is similar to] games that I have played in my childhood on computers like this, so there was a lot of familiarity involved... that was the [benefit] of it, the weakness of it I felt was that, so there's a reason I don't play these games any longer, you kind of grow out*

> *of them. Sure it felt familiar, but is it something that I want to do? ... It felt like this was me putting a lot of effort, and really, what are the rewards?... I liked the AR because it was new, it was a twist on the game."* (P19)

This novelty engaged them and, for some, encouraged reflective thinking about CitSci.

> *"I realized I was paying a lot more attention in the AR version."* (P19)

> *"[AR is] more real, you're more involved."* (P23)

> *"It seemed like, oh, [MemoryAR] is more of a science-related thing, like I got more of that impression."* (P12)

Yet, like many toys, after the initial novelty wore off, some players had a change in opinion.

> *"I think [MemoryAR could become] tedious... I think there's a threshold... at a certain point the PC game just becomes more convenient... at a certain point [my enjoyment in the physicality] would start to dip."* (P7)

> *"I wouldn't enjoy the physical aspect in the longer term."* (P16)

With respect to the quality of the toy, several participants commented on the AR technology. Some were frustrated by its *"glitchy [strong negative emphasis]"* (P1) shortcomings, but for others, it worked better than expected.

> *"[I was] pleasantly surprised how fast everything scanned, like the pictures picked up... there wasn't even a delay."* (P12)

Although participants recognized the limitations and frustrations of playing AR as a traditional game at home, they spoke positively of AR's potential use in communal, learning oriented spaces, such as museums or classrooms.

> *"I would be very interested in this as a learning activity."* (P18)

> *"I could see this as something a teacher would do for their students."* (P24)

> *"I believe this activity would have many benefits for both students and senior citizens both to develop memory and concentration skills as well as to maintain memory in older people. I believe it could be useful to help students learn to be good observers of details and to notice characteristics. I do [feel] it would enhance children's interest in Science."*[9] (P23)

> *"AR brought out my inner child... I can imagine children at a museum playing this."* (P21)

AR's position as a toy may not make it the best modality for long-term citizen science retention in the same contexts as non-AR games[10], though it should be noted that this study was not longitudinal. Regardless, the positive engagement created in the short term is particularly interesting, especially in contexts where the drawbacks of AR can be ameliorated.

| | Browser | AR |
|---|---|---|
| IMI: Interest/Enjoyment | 4.86 (0.93) | 5.64 (0.71) |
| IMI: Effort/Importance | 5.20 (0.90) | 5.40 (0.90) |
| IMI: Value/Usefulness | 5.14 (0.71) | 5.29 (0.79) |
| Label accuracy | 60.0% (10.5%) | 56.9% (16.9%) |
| Time between matches (s) | 11.6 (2.2) | 42.1 (17.7) |

**Table 1: Descriptive statistics of the IMI subscales and gameplay data (n=24). IMI results are reported as median (with median average deviation) of each averaged subscale (a 1-7 Likert scale). Label accuracy and Time between matches are also reported as median (with median average deviation). Accuracy was measured as the number of matches where the images selected were of the same known category divided by the total number of matches (per participant). Time between matches was calculated as the average time between matches (per participant). While AR IMI scores were slightly higher in all categories, AR suffered from a notable increase in match time without any increase in accuracy.**

## 4.2 Survey and Match Analysis

Table 1 provides the IMI median results as well as the participants' median accuracy and time between matches for each version of the game. MemoryAR had higher scores in all three IMI subscales, especially Interest / Enjoyment, however the speed at which matches could be made in MemoryAR was far slower. Participants claimed that this slowness prevented "spamming" (thoughtless task attempts) and led to accuracy:

> *"[MemoryBrowser] was more trial and error than [MemoryAR]."* (P23)

> *"I think that... even though the laptop version might be faster and generate more data in the long run, I think it might be better for like, users to do the AR version because I think... engagement in the game would like, make the results more accurate or allow the participants to want to think more about it, [rather] than like, passively playing the laptop game... Over time, participants in the laptop game would become more passive."* (P9)

However this observation was not reflected by the data, as accuracy scores between MemoryAR and MemoryBrowser were similar. A limitation in our study was that, as mentioned previously, there was a difference in the ground truth of the image sets used in the two different interfaces. This may have had some impact on the matches, although players do not know which images were ground truth beforehand[11].

# 5 DISCUSSION

In this study, we compared two citizen science adaptations of the traditional game of Memory in order to explore the potential value

[9] This quote was provided by the participant as written feedback after the study.

[10] It is plausible that the increased engagement would lead to replayability, though it is similarly plausible that the erosion of novelty and frustrations with physical tiles would lead to a lack of replayability.

[11] Since the game does provide feedback for matches made with ground truth or without, particularly savvy players could have used this to deduce the nature of the dataset, creating a difference in how the game was played between MemoryAR and MemoryBrowser. Since this is hypothetical, and no player commented on using this sort of tactic, we consider the gameplay differences negligible.

of AR in citizen science gaming. From our findings, we report that AR is a leisurely toy. This has four implications, addressed in the four themes we generated through thematic analysis: (1) The browser version is more familiar, fast, and accessible; (2) Physicality slows AR play; (3) AR affords and demands spatial processing; and (4) The AR novelty has toy qualities.

AR's physicality not only slowed play, it delayed the game setup, added cognitive demand for spatial processing, and could potentially create physical discomfort[12]; however, for some participants, the affordances of slow, spatial play afforded reflection and mental organization which ultimately made playing easier and more enjoyable. On the other hand, the browser interface was more familiar, more accessible, and faster, leading to more throughput on the CitSci task.

AR had a novelty effect, giving it toy-like qualities as an activity. That said, novelty effects eventually wear off, and some participants reported their feelings toward the modality changing even within the single hour of the study.

Returning to our research questions, how do these two games differ in engagement? Do these modalities fill different play niches? And how do these games interact with the CitSci experience and task goals? While 17 of the 24 participants claimed they liked MemoryAR over MemoryBrowser within the context of this study, the nuances within their answers lead us to conclude that these modalities do, in fact, serve different niches of play. The browser interface is better suited to rapid, accessible, productive play, while AR is better suited to reflective, one-time, unstructured experiences. It is not the case of one modality being simply better or more enjoyable than another; rather, the modality affects the kind of experience of play.

Players of the browser interface found themselves *"spamming"* (P22) or playing mindlessly (P2), while AR encouraged intentionality by requiring players to physically engage with the task (P2, P19).

Because of their different niches, future work may investigate how people engage with these modalities in other contexts, such as museums and classrooms: contexts where players approach the game in a more natural fashion and play without implicit or explicit time constraints. Other variations may also affect engagement, including: using a different image set, playing a different game, and varying the proportion of ground truth images in the image set. Moreover, TOSAR was developed for multiplayer play and in-person community building [33]. Indeed, Memory is traditionally a multiplayer game; however, we examined only single-player experiences in this study due to restrictions of the global COVID-19 pandemic. As suggested by previous work [33], there are many key factors to co-located multiplayer behavior worth studying with respect to how they modulate the citizen scientist experience, and more work remains to determine how to best facilitate this type of engagement.

As research into gamification continues, researchers should be cognizant about its negative effects. A cultural focus on productivity at the expense of everything else creates an increasing blur between work and enjoyment [7]. Pushes for gamification out of a desire for increased productivity run the risk of corrupting the goodwill and enjoyment inherent in citizen science by exacerbating participant burnout. While an extended discussion of this topic is out of the scope of this paper, we urge those implementing gamification in production citizen science projects to include digital wellness features, such as encouraged breaks for participants.

[12] Though it should be noted that physical discomfort is not unique to AR, there are ergonomic concerns with stationary screens and mouse based-interfaces.

## 6 CONCLUSION

In this study, we compared two citizen science adaptations (browser-based and AR) of the traditional game of Memory in order to explore the potential value of AR in citizen science gaming. Through thematic analysis we generated four themes and conclude that AR is a leisurely toy: while the browser interface is better suited to rapid, accessible, productive play, AR is better suited to reflective, one-time, unstructured experiences. The browser game produced faster throughput of image labels on the CitSci task, but participants described the AR experience as more memorable and thoughtful with respect to CitSci. We recommend future work explore co-located multiplayer AR as a method of critical engagement, AR's engagement amongst children, and AR within contexts such as museums, classrooms, or citizen science events.

## ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grant no. 1816426 and a Northeastern University TIER1 grant.

## REFERENCES


[1] B Albers, N De Lange, and S Xu. 2017. Augmented citizen science for environmental monitoring and education. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* 42, Part 2 (2017), W7.

[2] Abrar Omar Alkhamisi, Saudi Arabia, Muhammad Mostafa Monowar, et al. 2013. Rise of augmented reality: Current and future application areas. *International journal of internet and distributed systems* 1, 04 (2013), 25.

[3] Anne Bowser, Derek Hansen, Yurong He, Carol Boston, Matthew Reid, Logan Gunnell, and Jennifer Preece. 2013. Using gamification to inspire new citizen science volunteers. In *Proceedings of the first international conference on gameful design, research, and applications*. 18–25.

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[5] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exxercise and Health* 11, 4 (2019), 589–597.

[6] Carlos Díaz, Marisa Ponti, Pinja Haikka, Rajiv Basaiawmoit, and Jacob Sherson. 2020. More than data gatherers: exploring player experience in a citizen science game. *Quality and User Experience* 5, 1 (2020).

[7] Anne Dippel and Sonia Fizek. 2018. Laborious Playgrounds: Citizen science games as new modes of work/play in the digital age. In *The playful citizen: civic engagement in a mediatized culture*. Amsterdam University Press.

[8] Alejandro Echeverría, Matías Améstica, Francisca Gil, Miguel Nussbaum, Enrique Barrios, and Sandra Leclerc. 2012. Exploring different technological platforms for supporting co-located collaborative games in the classroom. *Computers in Human Behavior* 28, 4 (2012), 1170–1177.

[9] Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L Cox. 2014. Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2985–2994.

[10] Roman Fedorov, Darian Frajberg, and Piero Fraternali. 2016. A framework for outdoor mobile augmented reality and its application to mountain peak detection. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 281–301.

[11] David Furió, Santiago GonzáLez-Gancedo, M-Carmen Juan, Ignacio Seguí, and Noemí Rando. 2013. Evaluation of learning outcomes using an educational iPhone game vs. traditional game. *Computers & Education* 64 (2013), 1–23.

[12] Trisha Gura. 2013. Citizen science: amateur experts. *Nature* 496, 7444 (2013), 259–261.

[13] Paul Haynes, Sigrid Hehl-Lange, and Eckart Lange. 2018. Mobile augmented reality for flood visualisation. *Environmental modelling & software* 109 (2018), 380–389.

[14] Eva Hornecker and Thomas Psik. 2005. Using ARToolKit Markers to Build Tangible Prototypes and Simulate Other Technologies. In *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction.* 30–42.

[15] Ioanna Iacovides, Charlene Jennett, Cassandra Cornish-Trestrail, and Anna L Cox. 2013. Do games attract or sustain engagement in citizen science? A study of volunteer motivations. In *CHI'13 extended abstracts on human factors in computing systems.* 1101–1106.

[16] M. Jiménez, M. Torres Torres, R. John, and I. Triguero. 2020. Galaxy Image Classification Based on Citizen Science Data: A Comparative Study. *IEEE Access* 8 (2020), 47232–47246.

[17] Carmen M Juan, Giacomo Toffetti, Francisco Abad, and Juan Cano. 2010. Tangible cubes used as the user interface in an augmented reality game for edutainment. In *2010 10th IEEE international conference on advanced learning technologies.* 599–603.

[18] Carmen M. Juan, Giacomo Toffetti, Francisco Abad, and Juan Cano. 2010. Tangible Cubes Used as the User Interface in an Augmented Reality Game for Edutainment. In *Proceedings of the 2010 10th IEEE International Conference on Advanced Learning Technologies.* 599–603.

[19] Louise H Kidder and Michelle Fine. 1987. Qualitative and quantitative methods: When stories converge. *New directions for program evaluation* 1987, 35 (1987), 57–75.

[20] George Koutromanos, Alivisos Sofos, and Lucy Avraamidou. 2015. The use of augmented reality games in education: a review of the literature. *Educational Media International* 52, 4 (2015), 253–271.

[21] Christopher Kullenberg and Dick Kasperowski. 2016. What is citizen science?–A scientometric meta-analysis. *PloS one* 11, 1 (2016), e0147152.

[22] María Teresa Linaza, Aitor Gutierrez, and Ander García. 2013. Pervasive augmented reality games to experience tourism destinations. In *Information and Communication Technologies in Tourism 2014.* Springer, 497–509.

[23] Anna Loparev, Lauren Westendorf, Margaret Flemings, Jennifer Cho, Romie Littrell, Anja Scholze, and Orit Shaer. 2017. BacPack: Exploring the Role of Tangibles in a Museum Exhibit for Bio-Design. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction.* 111–120.

[24] Suvodeep Mazumdar, Luigi Ceccaroni, Jaume Piera, Franz Hölker, A Berre, Robert Arlinghaus, and Anne Bowser. 2018. Citizen science technologies and new opportunities for participation. UCL Press.

[25] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1 (1989), 48–58.

[26] Niantic and Nintendo. 2016. *Pokémon Go.* Game [iOS, Android]. Nintendo, Kyoto, Japan.

[27] Trond Nilsen, Steven Linton, and Julian Looser. 2004. Motivations for augmented reality gaming. *Proceedings of FUSE* 4 (2004), 86–93.

[28] Johanna Okerlund, Evan Segreto, Casey Grote, Lauren Westendorf, Anja Scholze, Romie Littrell, and Orit Shaer. 2016. SynFlo: A Tangible Museum Exhibit for Exploring Bio-Design. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction.* 141–149.

[29] Richard M Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology* 43, 3 (1982).

[30] JL González Sánchez, N Padilla Zea, and Francisco L Gutiérrez. 2009. Playability: How to identify the player experience in a video game. In *IFIP Conference on Human-Computer Interaction.* Springer, 356–359.

[31] Henry Sauermann and Chiara Franzoni. 2015. Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences* 112, 3 (Jan. 2015), 679–684. https://doi.org/10.1073/pnas.1408907112

[32] Yusuf Sermet and Ibrahim Demir. 2020. Virtual and augmented reality applications for environmental science education and training. *New Perspectives on Virtual and Augmented Reality: Finding New Ways to Teach in a Transformed Learning Environment* (2020), 261–275.

[33] Sofia Eleni Spatharioti, Borna Fatehi, Melanie Smith, Avery Rosenbloom, Josh Aaron Miller, Magy Seif El-Nasr, Sara Wylie, and Seth Cooper. 2020. Tile-o-Scope AR: An Augmented Reality Tabletop Image Labeling Game Toolkit. In *International Conference on the Foundations of Digital Games.* 1–4.

[34] Sofia Eleni Spatharioti, Sara Wylie, and Seth Cooper. 2019. Using Q-Learning for Sequencing Level Difficulties in a Citizen Science Matching Game. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts.* 679–686. https://doi.org/10.1145/3341215.3356299

[35] Luis Von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 319–326.

[36] Marco Willi, Ross T Pitman, Anabelle W Cardoso, Christina Locke, Alexandra Swanson, Amy Boyer, Marten Veldthuis, and Lucy Fortson. 2019. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution* 10, 1 (2019), 80–91.

[37] Lesley Xie, Alissa N Antle, and Nima Motamedi. 2008. Are tangibles more fun? Comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In *Proceedings of the 2nd international conference on Tangible and embedded interaction.* 191–198.