Computers in Human Behavior 52 (2015) 492–506

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Points, stories, worlds, and diegesis: Comparing player experiences in two citizen science games

Nathan R. Prestopnik [a], Jian Tang [b,*]

[a] Department of Computer Science, Ithaca College, 401C Williams Hall, Ithaca, NY 14850, USA
[b] School of Information, Central University of Finance and Economics, No. 39 Xueyuan South Road, Haidian District, Beijing 100081, China

ARTICLE INFO

Article history:
Received 5 March 2015
Revised 27 May 2015
Accepted 28 May 2015

Keywords:
Gamification
Crowdsourcing
Citizen science
Diegesis
Stories
Narrative

ABSTRACT

We conducted an experiment to examine how people perceive differences between points-based and story-based gamification approaches. We were interested in how these differences impact peoples' play experiences and perceptions of working on a citizen science task.

Our findings show that the story-based game, Forgotten Island, was strongly preferred over the points-based game, Happy Match. Participants indicated that this was because of "diegesis" in Forgotten Island – in other words, a focus on story-motivated activities and rewards made the citizen science task more enjoyable and gave participants various reasons to continue play.

This study suggests that story-based games can be a powerful tool for attracting participants to citizen science tasks. In particular, compared to point-based games, story-based games may be more useful for attracting and engaging participants who are ambivalent about scientific inquiry. This paper also discusses some of the challenges and possibilities for both points-based and story-based gamification.

1. Introduction

In recent years, the term "crowdsourcing" has emerged to describe information systems that distribute work or tasks amongst large groups of people. Existing crowdsourcing systems address a wide variety of commercial, educational, and scientific tasks. In this present study we direct our interest toward "citizen science" systems, information systems that support crowdsourced involvement of non-scientist members of the general public in scientific inquiry (Cohn, 2008; Wiggins & Crowston, 2011).

One important challenge of instantiating a citizen science system is the need to recruit and retain participants, i.e. to attract a crowd. Yet citizen science tasks can sometimes be mundane or repetitive, and they may also be complex or require specialized participant training and knowledge. Project participants are human beings, not simply organic CPUs, so making challenging scientific tasks interesting, worthwhile, and achievable is critical for any successful citizen science system (Franzoni & Sauermann, 2014; von Ahn, 2006).

There are many approaches to attracting a crowd of volunteers to participate in citizen science project, including community building, competitions, reward systems, and more. In this research, we are interested in so-called "games with a purpose" (von Ahn, 2006), also sometimes referred to as "gamification." This is the notion of turning non-play activities into games (Deterding, Dixon, Khaled, & Nacke, 2011; Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011; von Ahn, 2006; von Ahn & Dabbish, 2008). In the citizen science context, games with a purpose merge scientific tasks with engaging game elements. Yet entertainment game players value games for the fun, interesting, and rewarding experiences they provide (Schell, 2008). With this in mind, we are most interested in how games with a purpose can attract volunteers who have limited enthusiasm for helping scientists or working on the underlying science task. That is, we are interested in people who would not normally think of themselves as "citizen scientists." We are interested in this because well-designed citizen science games with a purpose have the potential to attract large crowds of helpful volunteers, even in circumstances where the science task is difficult or uninteresting (Flatla, Gutwin, Nacke, Bateman, & Mandryk, 2011).

The current enthusiasm for gamification and games with a purpose sometimes overlooks a fundamental issue: what *is* a "well-designed" game with a purpose? There are many different motivational techniques that can be employed when designing games (Chen & Chen, 2013), and many different philosophies about what a game really is (Rogers, 2010; Salen & Zimmerman, 2004; Schell, 2008). Games can be heavily mechanic-oriented, or focus

* Corresponding author at: School of Information, Central University of Finance and Economics, No. 39 Xueyuan South Road, Haidian District, Beijing, 100081, China.
E-mail addresses: nprestopnik@ithaca.edu (N.R. Prestopnik), jiantangruc@gmail.com (J. Tang).

http://dx.doi.org/10.1016/j.chb.2015.05.051
0747-5632/

more on stories or characters. Games can be framed around high scores and points, or around imagined, fictional experiences. Some researchers evangelize about the great potential and benefits of games with a purpose (e.g., McGonigal, 2011; von Ahn, 2006), while others argue that current gamification techniques are flawed, even immoral, mechanisms for tricking participants into doing work (e.g., Bogost, 2011).

Many – perhaps most – games with a purpose in the citizen science domain rely upon techniques such as points, scores, badges, and achievements to motivate and engage players (Bogost, 2011). In contrast, commercially available entertainment games, especially those targeted toward enthusiast rather than casual gamers, frequently stress in-depth storytelling, immersive game worlds, and rich visual and aural experiences.

Questions result: given the popularity of stories in commercial entertainment video games, why do so few citizen science games take advantage of story-oriented design elements? Moreover, how might games designed around story features also engage participants in purposeful activities? Can such games attract volunteers who would not normally involve themselves in a citizen science project?

In this present research, we ran a controlled study of player experiences in two citizen science games with the goal of addressing some of these questions. The first game, *Happy Match*, is a points-driven quiz game, similar to many currently published citizen science games. The second, *Forgotten Island*, is a story-driven, exploration-oriented game that differs from the mainstream of citizen science gaming in many respects.

Happy Match and Forgotten Island are part of an ongoing citizen science research project called *Citizen* Sort,[1] and were built to study player experiences in the context of citizen science games with a purpose. Our literature review, experimental results, analysis, and discussion unpack some of the most important themes and concepts that came out of our exploration of points vs. story-oriented gaming, purposeful game design, task design, and the player experience.

2. Literature review

2.1. Points and gamification

The popular but controversial term "gamification" is relevant to our work, not the least because points-based games are, in many ways, gamified tasks first, and play experiences second. However, this term is laden with rhetorical baggage, with some even likening gamification to "exploitationware" (Bogost, 2011). This criticism is directed toward games where players undertake tasks in exchange for a score, a badge, experience points (XP), or similar; but if a player is earning points for undertaking some activity, he or she must be playing a game, right?

The critics of score-driven activities argue that points, badges, and the like are not gameplay. Rather, these are metrics by which really meaningful interactions – the *experiences* that truly engage and delight players – are measured and progress is recorded. To remove the meaningful aspects of play while retaining the measurement system is to produce something that is not really a game at all (Bogost, 2011; Deterding, 2012; Deterding, Dixon, et al., 2011; Deterding, Sicart, et al., 2011; Salen & Zimmerman, 2004). Bogost (2011), in particular, levels harsh criticism at such games: "Like having a website or a social media strategy, 'gamification' allows organizations to tick the games box without fuss. Just add badges! Just add leaderboards!"

Yet various researchers have fielded a variety of highly successful, very engaging citizen science games that do adopt point, badge, achievement, and leaderboard approaches. These include games for language acquisition and translation (DuoLingo[2]), simulated protein string folding (Fold.It[3]), simulated genetic sequencing (Phylo[4]), analyzing historical records (Old Weather[5]), and mapping neural pathways in the brain (EyeWire[6]) among others. Participation levels vary, but these examples have all attracted considerable numbers of voluntary players. Clearly, something about these experiences – perhaps a social experience or the activity itself – provides real meaning, at least for some.

Since our objective is to probe the differences between points-based science activities that feel "gamified" and story-based science activities that may or may not fit this controversial label, we favor von Ahn's (2006) broader term "games with a purpose" (and a variation, "purposeful games"). These accommodate many variations on merging games and tasks, including both points-based "gamification" and other approaches. von Ahn's (2006) term permits us to think of points and stories as related but distinct mechanisms for convincing players to become participants of citizen science projects.

2.2. Stories and diegesis

The term "diegesis" is an important way of thinking about stories within games. It refers to the notion of the "story world" vs. the "real world" (De Freitas & Oliver, 2006; Galloway, 2006; Stam, Burgoyne, & Flitterman-Lewis, 1992). Diegesis is most easily understood through an example: the label on a treasure chest found by a game player.

Deeply etched into cracked wood by a rough hand and a dull knife, the misspelled word "Tresur" is suggestive of the former owner of the chest – perhaps a vicious and unlettered pirate, perhaps a highwayman or bandit. This is a diegetic label. The handwriting, the texture of the letters, the knife scrapes, and the misspelling all elaborate upon the game world and story.

A non-diegetic alternative might be cleanly printed white text, Helvetica font, hovering in space over the chest, rotating to always face the player's POV. This label is part of the game's GUI, not the game story. It is functional and useful, but the diegetic label better preserves a player's sense of immersion in the experience.

The notion of diegesis is an interesting way to frame points-based vs. story-based games with a purpose. Points, ranks, and badges measure real things like player accuracy, time spent playing, or milestone accomplishments, but they only matter to players insomuch as they quantify things the player values *outside the game*. A player who altruistically desires to help biologists taxonomically classify insect species would be interested in earning points based upon the quality of his or her classifications. The points are a valued measure of something external to the game: how helpful a player's contributions are to a wider scientific discipline. When coupled with a leaderboard or community, the points grow in value, fostering competition, cooperation, and prestige. Yet they are non-diegetic; they do nothing to expand upon the world of the game.

Are all citizen science participants altruistic enough to value points, scores, and badges for the externalities they represent? Probably not. Most citizen science projects find that a core group of users provide most contributions, while a long tail of participants will contribute much less frequently (Franzoni & Sauermann, 2014). These "long tail" (Anderson, 2008) players seem unlikely to value scores or points that quantify their produc-

[1] http://www.citizensort.org

[2] http://www.duolingo.com/.
[3] http://fold.it/.
[4] http://phylo.cs.mcgill.ca/.
[5] http://old.oldweather.org/.
[6] https://eyewire.org/signup.

tivity in a task that wasn't interesting to them in the first place. A player who doesn't care much about the task won't care much about points; if they don't care about points, they also won't care about the game.

Individuals like this are more likely to be motivated by play experiences that have endogenous value (Schell, 2008) and are enjoyable in and of themselves (Berkovsky, Coombe, Freyne, Bhandari, & Baghaei, 2010; Garris, Ahlers, & Driskell, 2002; Ryan & Deci, 2000). This is where the notion of diegesis begins to show its value as a tool for gamification. For example, a biology task might be recast as a wildlife adventure. Player productivity could be rewarded with better exploration equipment such as shovels, picks, climbing ropes, maps, tents, or camping gear.

In this scenario the points have become diegetic. Players are still rewarded for performance, preserving the opportunity to attract altruistic, task-focused participants to the game. Yet the rewards also enhance the game world and story. The rewards are a foundation upon which more meaningful gameplay can be built. The game is no longer just about classification. Now it is about other things – adventure, excitement, exploration, the great outdoors – that may be of great interest to players who would otherwise pass the science task by.

2.3. Motivating games: fantasy

What makes a motivating purposeful game? Researchers have identified a variety of design heuristics that are important, and story, fantasy, and characters are a recurring theme (e.g., Costikyan, 2013; Garris et al., 2002; Malone, 1980; Malone & Lepper, 1987; Prensky, 2005). This literature deals primarily with learning or entertainment games, yet it likely applies to scientific games with a purpose as well.

As a heuristic, "fantasy" refers to the way a game, "evokes mental images of physical objects or social situations that are not actually present," (Malone & Lepper, 1987). Fantasy is implemented into a play experience primarily through the game world and story, allowing players to safely experience events, risks, and rewards that are not possible in the real world (Qin, Rau, & Gavriel, 2009). Malone and Lepper (1987) argue that fantasy is one of the most important features of video games.

Rieber (1996) splits the notion of fantasy into two types: exogenous and endogenous. Exogenous fantasy describes a "layering" of whimsical material on top of purposeful content. For example, in a game featuring exogenous fantasy, a player might add 2 + 2 correctly in order to launch a rocket. Here, the purposeful content (simple addition) leads to fantasy feedback (the rocket launch). However, the two are not linked diegetically, as they would be if, for example, the player performed calculations needed to actually launch rocket (e.g. a calculation for mass or fuel).

Endogenous fantasy is a more diegetic approach, where the task is thematically and/or narratively linked to the game world. Garris et al. (2002) describe an endogenous fantasy game where physics students pilot a spaceship through Earth's atmosphere to learn key principles.

Though the notion of using fantasy to motivate players is not a new idea, few designers depart from proven score-based strategies to use endogenous fantasy, including stories, in more diegetically-oriented, story-driven citizen science games. One reason may be the difficult and resource-intensive creative challenges that come with developing a story and writing a script, envisioning a game world, producing concept and final artwork, designing characters, creating a compelling sound design, composing a musical score, and connecting these fantasy elements to already complex functionality (Prestopnik & Crowston, 2012).

2.4. Player experiences

"The game enables the experience, but *it is not the experience*," (Schell, 2008, p. 10). Player experiences encompass the subjective feelings that occur when interacting with a game. Game designers implement specific features to engender certain kinds of experiences, but ultimately the player experience is both personal and individual, a relationship between player and game. Player experiences are sometimes referred to as user experiences, play experiences, or game experiences (Nacke, Grimshaw, & Lindley, 2010; Schild, LaViola, & Masuch, 2012). In this research, we are interested in the player experience because we want to understand how potential participants *perceive* points-based or story-based citizen science games.

The concept of player experience is complex, covering multiple dimensions of interaction, including positive affect, negative affect, flow, sensory immersion, tension, challenge, and competence (Gajadhar, Kort, & IJsselsteijn, 2008; IJsselsteijn, de Kort, Poels, Jurgelionis, & Bellotti, 2007; Schouten, Pfab, Cremers, van Dijk, & Neerincx, 2014). Some of these dimensions are positive or neutral while others are negative. Positive dimensions are mechanisms for enhancing player enjoyment, while negative dimensions are threats to enjoyment that can occur during play.

Positive player experience dimensions include flow, positive affect, competence, immersion, and challenge. Flow describes the optimal experience, one in which the player is fully absorbed in an activity, even to the point of losing track of time, through an ideal balance of their own abilities and the difficulty of a task (Csikszentmihalyi, 1991; Webster, Trevino, & Ryan, 1994). Positive affect is the feeling of fun and enjoyment that a player might have when playing a game. Competence refers to players' feelings of successfully applying their skills and performing well, and immersion refers to the impact of the multi-sensory properties of a game on the user. Finally, challenge indicates the player's experience of how difficult or easy the game is. Challenge is related to flow, in that it is important to confront the player with an appropriate level of challenge if they are to achieve a flow state (Csikszentmihalyi, 1991).

Negative dimensions of a player experience include negative affect and tension. Negative affect is the boredom and distraction that players might feel when they are not challenged enough or interested enough in a game. Tension is a negative experience relating to the difficulty of a game or the competition a player feels.

Player experience studies have relied on different research methods, including observations, interviews, focus groups, self-reported data (Brockmyer et al., 2009), or physiological data (Schild et al., 2012), and have attempted to study several or all subscales of game experience. For instance, Cox, Cairns, Shah, and Carroll (2012) studied how the speed of player interactions and time pressure affected game immersion. These prior studies have shown that the various features designed into games can have significant impacts on the player experience, and can further lead to the subsequent intentions of players to play the game again. However, little of this prior scholarship has looked specifically at citizen science games.

2.5. Research questions

It is an open question whether diegetic, story-driven games with a purpose are truly a viable alternative to more common points-based gamification in the citizen science context. While the design differences between these approaches are understood, the different perceptions that players have of them – the actual player experiences they produce as games with a purpose – is unknown.

Accordingly, we adopt a set of open-ended research questions to guide our study of points-based vs. story-based citizen science games:

RQ1: *What do players perceive to be the key differences between points-based and story-based citizen science games?*
RQ2: *How do perceived differences in the design of points-based vs. story-based citizen science games shape the player experience?*
RQ3: *How do perceived differences in the design of points-based vs. story-based citizen science games impact players' perceptions of working on the embedded science tasks?*

In section two we characterized the design difference between story-based and points-based games in the abstract, and we identified various concepts that contribute to an overall player experience. In section three, we make these concepts concrete by describing two specific games that were designed and implemented for this study. We also elaborate upon our assumptions as to the kinds of experience that these different games may engender. Following section three, we present the methods and results of a controlled, experimental study that addresses our research questions and probes player experience in our two citizen science games.

3. Two games with a purpose: Happy Match & Forgotten Island

3.1. Science task

We developed two purposeful games for study, *Happy Match* and *Forgotten Island*. These games were framed around a crowdsourced "citizen science" activity: the taxonomic classification of plant, animal, and insect photos. *Happy Match* is a points-based game, while *Forgotten Island* takes a story-driven, diegetic approach.

In the life sciences, researchers routinely collect photographs of living things. When captured with digital cameras or cell phones, these photographs are often tagged with time and location data, and the resulting imagery and metadata can help to address important research questions, for example on wildlife populations, migration patterns, or the impact of urban sprawl on local ecosystems. Yet time and location tagged photos are only valuable when the subject of the photograph is known and expressed in scientific terms, i.e., by scientific species name. This information is rarely recorded when the photograph is captured in the field.

In *Happy Match* and *Forgotten Island*, players examine expert-provided photographs of living things and answer questions about their various features. The versions of the games used in this study were instantiated with pictures and questions about moths. The questions are organized around biological "characters" and "states" (attributes and values). For example, a question might ask about the character "wing pattern," and possible states might include, "banded," "spot," "stripe," or "speckled." The character questions and state answer choices are tailored to enable classification to species.

Though *Happy Match* and *Forgotten Island* are different games with different gamification approaches, both are designed around the same taxonomic classification task. The task is structured identically and presented in a highly similar (though not quite identical) fashion in each game.

Happy Match provides players with example photos that help guide players to make good classification decisions (See Fig. 1). The example photos are shown as thumbnails associated with the answer choices for each question. Players have the ability to zoom in on both the example photos and the photos they are classifying. Answer choices and the questions themselves also include help text to explain what players should look for in order to classify a photograph correctly.

In *Forgotten Island*, players classify the same data set as *Happy Match* using the same selection of character and state questions (See Fig. 2). The same help text, example photographs, and zoom features are available, just as in *Happy Match*.

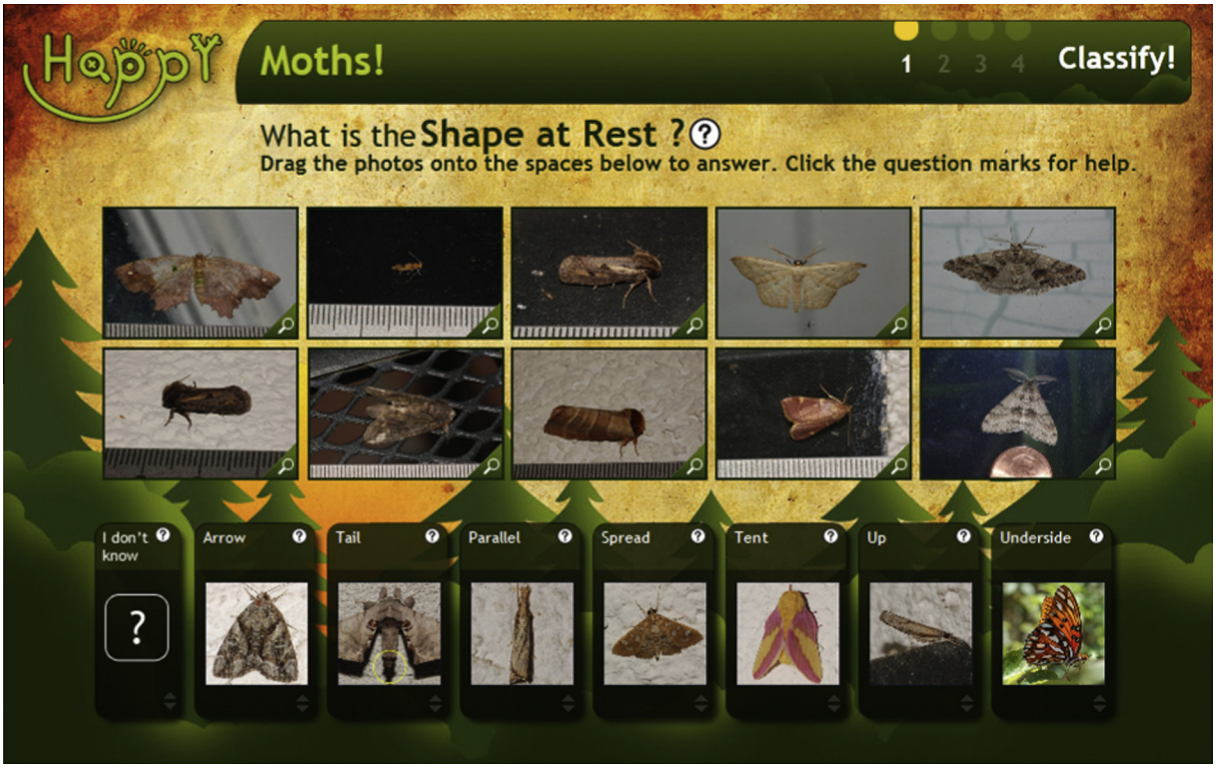

Fig. 1. The classification interface for *Happy Match*.

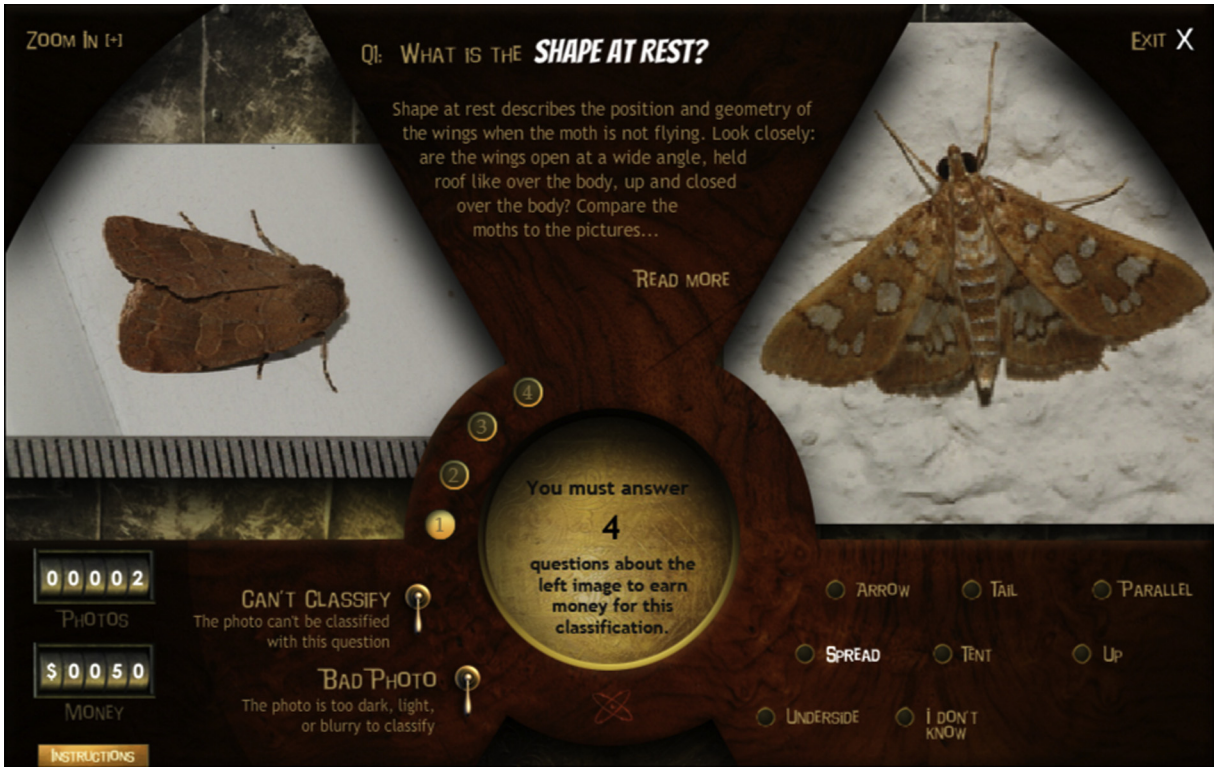

Fig. 2. The classification interface for *Forgotten Island*.

Happy Match players classify ten photos, progressing through rounds (one character question per round). In each round, players answer the same question for each of the ten photos before progressing to the next round/question. After all questions have been answered for all photos, players are shown a score based upon their performance. Performance is measured by comparing some (not all) photographs in the game to a gold standard set of classifications provided by expert naturalists.

A major difference between *Happy Match* and *Forgotten Island* is that *Forgotten Island* players classify just one photo at a time, rather

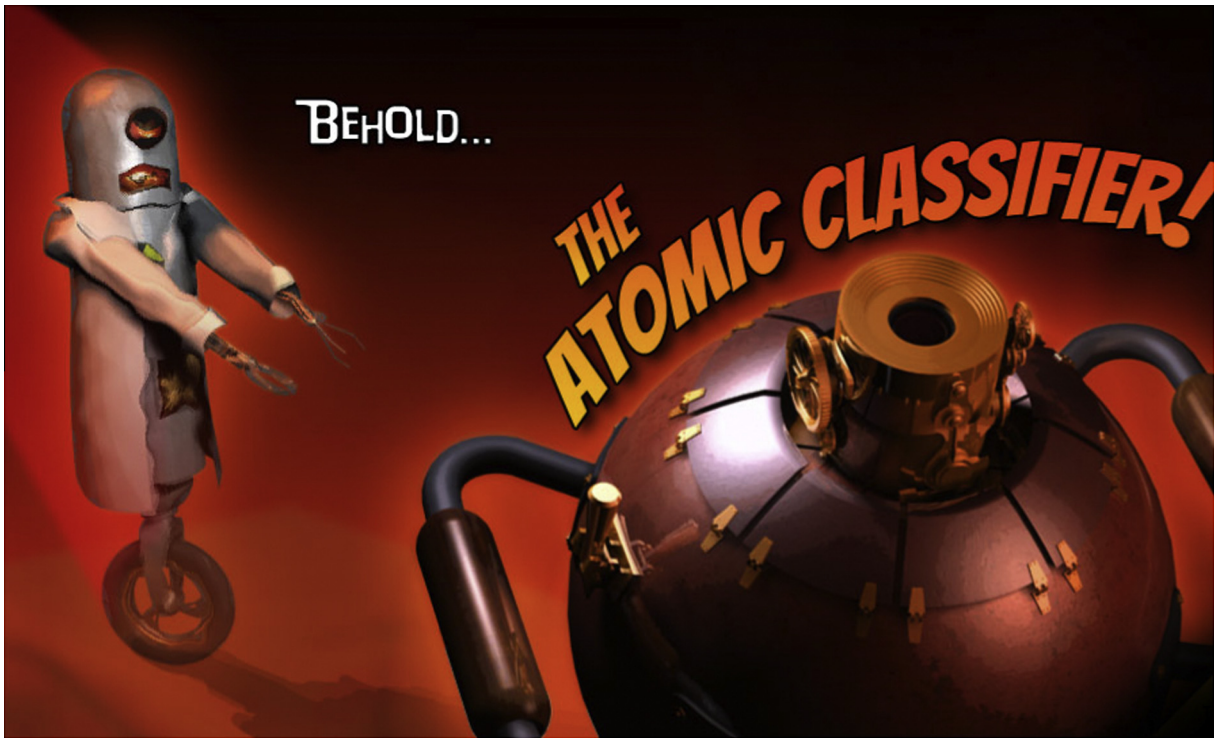

Fig. 3. The Atomic Classifier is a device given to the player early in the story. The machine-like look of the classifier and its interface (Fig. 2) helps establish diegetic connections to the game story and world.

than batches of ten images. This is because, in *Forgotten Island*, the classification task is connected diegetically to the game world through the use of endogenous fantasy.

Early in *Forgotten Island's* narrative, an archive of specimen photos is destroyed by the game's main antagonist, scattering photos of living things across the game world. Players are tasked with recovering these photos and rebuilding the archive by taxonomically classifying them. *Forgotten Island* players therefore have autonomy over how many photos in a row they wish to classify. Moreover, the *Forgotten Island* classification interface is called the "Atomic Classifier" (See Fig. 3) and is themed around a fictional device that is given to the player by another character early in the game. The use of endogenous fantasy in *Forgotten Island* is intended to motivate the classification activity with story rather than by connecting classification tasks to game points. Furthermore, instead of points, *Forgotten Island* players are diegetically rewarded with in-game currency that can be spent on puzzle items and equipment.

3.2. Happy Match gameplay

Happy Match represents a points-based approach to gamification, where the science task is placed in the foreground of the game, and players earn a score for their classification performance. As such, *Happy Match* gameplay is roughly the same thing as the science task itself. Players classify photographs for points by dragging them to the correct answer choices, earning a final score based on their performance. In essence, *Happy Match* is a quiz game.

One addition to the central quiz mechanic is a bonus round. The game is seeded with two photographs that have been pre-classified by experts on each of the character-state features of interest. The gold standard photographs are used to generate the player's score and also verify the quality of the data they are generating (we assume that a track record of good performance on gold standard photos can be used as an indicator of performance on unclassified photos). If a player answers all questions correctly about the

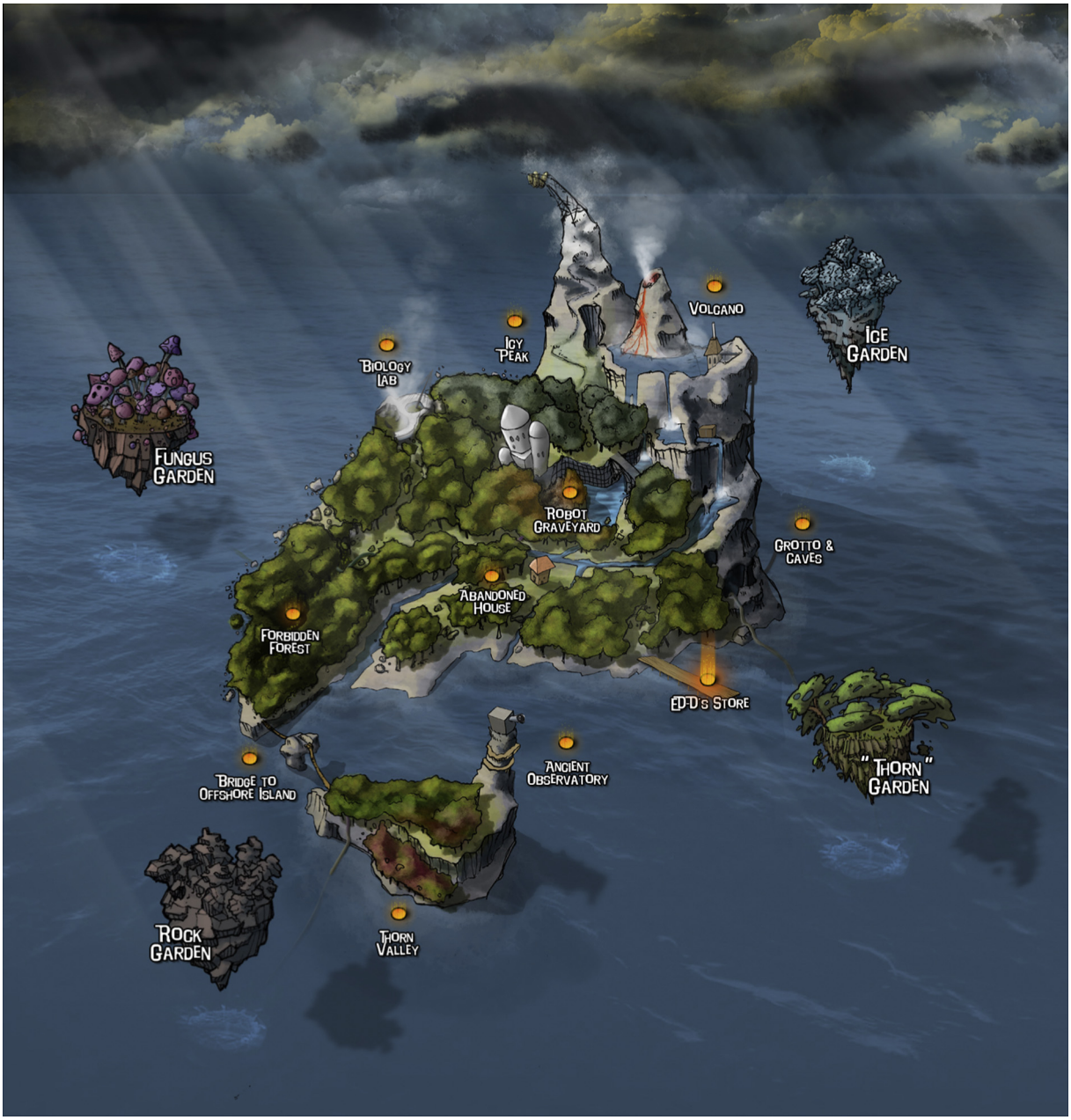

Fig. 4. The *Forgotten Island* game world.

seeded photos, he or she is given a bonus round and asked to identify to species in a similar drag and drop interface. A correct answer in the bonus round achieves an even higher game score. This overall score and a breakdown of correct answers is revealed to the player on a final score page. The data collected by players who earn high scores could contribute to the classification of moth species.

The points awarded in *Happy Match* are a measure of player performance – the ability to distinguish characterizing features on the moths and (potentially) to identify to species. Notably, the points have limited impact on gameplay. Only the bonus round is effected by player performance. Otherwise, the game simply presents a final score and the player is given the option to play again if they wish. With its focus on good performance in the embedded science activity, *Happy Match* is designed to appeal to players who have altruistic or science enthusiast motivations for playing (Garris et al., 2002).

3.3. Forgotten Island gameplay

Forgotten Island represents a story-based, endogenous fantasy approach to gamification (Malone, 1980, 1982; Malone & Lepper, 1987; Rieber, 1996), where the science activity is embedded into a much larger point-and-click adventure game (See Fig. 4). *Forgotten Island* was designed for players who have limited (or no) altruistic or science enthusiast motivations for playing, and who want the game to provide other motivations to play (Garris et al., 2002). In other words, *Forgotten Island* was designed under the explicit assumption that taxonomic classification of moths, far from being inherently interesting and educational, may well be tedious and boring for many players.

With this in mind, the classification activity described above is embedded into a much larger and more complex game that includes a unique, visually and aurally stimulating world to explore, a mystery narrative that paces the game and provides twists, payoffs, and goals for the player to pursue, humorous characters to interact with, and a variety of puzzles and other activities to engage in as they play. These features are intended to attract otherwise uninterested players and "seduce" (Jafarinaimi, 2012) them into repeatedly doing small bits of science in exchange for several hours of enjoyable interactive entertainment.

Most of *Forgotten Island* is organized around item-combination puzzles similar to those found in classic point-and-click adventure series like *Monkey Island* (1990–2011) and *King's Quest* (1984–1998). Players unlock new areas or story events by combining items together in logical but clever ways. For example, in *Forgotten Island*, players are only able to access the "Forbidden Forest" location once they have successfully found and combined three items together: a light bulb, battery, and copper wire. The resulting lamp can be used to clear away the darkness of the forest.

In traditional point-and-click adventures, items are typically found in the environment. In *Forgotten Island*, some items can be found, while others are purchased using in-game money. Money is earned using the Atomic Classifier (i.e. undertaking the science task), a device provided to the player early in the story. In this way, *Forgotten Island* reframes the notion of "points" as a diegetic element of the story world. New puzzles and goals are also assigned to the player as the story unfolds, necessitating more exploration, new items, game money, and regular classification interactions (See Fig. 5). Poor classifications as measured against the game's gold standard data set are also noted with a warning from the game's antagonist, along with a breakdown of the player's correct and incorrect decisions. A diegetically motivated training task early in the game provides guidance about how these mechanics will work.

The interlocking dependencies of *Forgotten Island* (exploration is needed to advance the story, items are needed in order to explore, money is needed to acquire items, and classifications are needed to earn money) mean that the player is rarely forced to focus on the science task for long durations. Rather, the game oscillates between various kinds of play, demanding engagement with the science task only at irregular intervals. However, successfully finishing *Forgotten Island's* roughly 6-h play time will require players to fully classify approximately 160 photographs (the equivalent of about 16 full games of *Happy Match*).

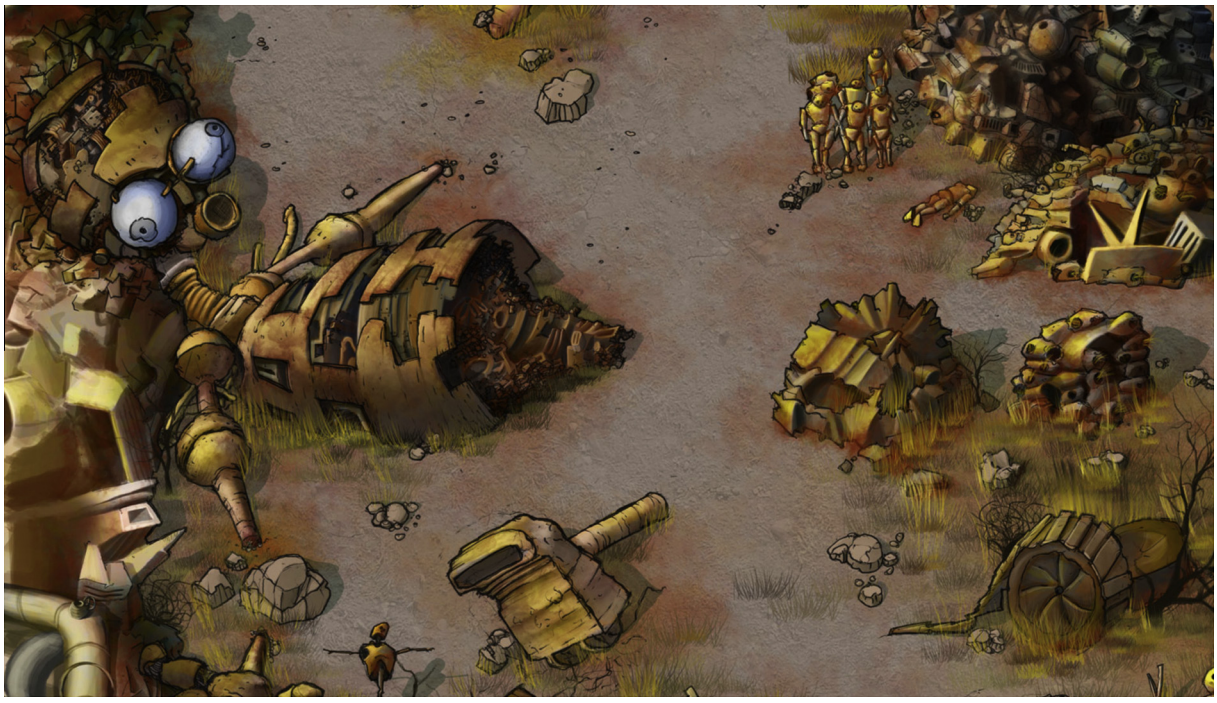

Fig. 5. Part of the "Robot Graveyard," one of the exploration locations of *Forgotten Island*.

4. Method

To compare *Happy Match* and *Forgotten Island*, we recruited participants to play both games and respond to a series of questionnaires about their background, play experiences in each game, and opinions about the two games in comparison.

4.1. Participants

We recruited twenty-seven participants from computer science courses at an undergraduate, liberal arts institution in upstate New York. Given our interest in studying how video games can attract "gamers" to citizen science projects, we drew upon computer science students as participants because of their likely interest in video games. Many computer science students are passionate about games, pursuing this interest both as gamers and as student game developers (Overmars, 2004). Note that the CS curriculum that we recruited from includes a game design minor.

22 of our participants were male, and 5 participants were female. Participant ages ranged from 18 to 22, with a mean and median age of 19 years old. 25 out of 27 participants were native English speakers. According to the G*Power calculation, for a one tailed paired *t*-test 27 participants is adequate to ensure a power of 0.80 at the significance level of 0.05 with a medium effect size (0.50) (Faul, Erdfelder, Buchner, & Lang, 2009).

4.2. Environment and equipment

The study took place in a quiet office with minimal distractions and was overseen at all times by one of the researchers. Participants attended the study session individually, and the same computer configuration was used for all participants: a Microsoft Windows 7 laptop with mouse running the Firefox web browser.

4.3. Procedure

Participants played the two games and responded to four questionnaires over the course of the experiment, which took about one hour to complete (mean of 56 min; median of 57 min). At the start of the study, the researcher provided participants with an informed consent document to be signed. This document provided basic detail about the study and the research objectives.

Following this, participants were asked to create a *Citizen Sort* account that would allow them to play both games, and then were given time to respond to the first questionnaire on demographics. Following the demographic questionnaire, participants played the two games. To control for ordering effects, the game order was randomized. 14 participants played *Happy Match* first, and 13 played *Forgotten Island* first.

Since the games have different gameplay approaches and mechanics, the amount of time spent on each game varied. For *Happy Match*, participants were asked to play one complete game, classifying ten photographs over four rounds (1 question per round). Participants stopped playing once they received a score for that game. This took most players about 10–15 min of play.

In *Forgotten Island* participants were asked to play until stopped. The researcher ended their play session at a predetermined point in the game, reached by most players in about 20 min of play. The design of the game ensured that participants were sufficiently exposed to both the story and the science task when they reached the determined stopping point. Since *Forgotten Island* gives players autonomy over their classification activity, the number of photographs classified before reaching the stopping point varied from participant to participant, but all participants fully classified at least five photographs.

Following the first play session, participants were asked to complete a player experience questionnaire, then play the second game, and then complete another player experience questionnaire for the second play session. Finally, participants were asked to respond to a comparison questionnaire that asked about their reactions to the games after having played both.

4.4. Survey instruments

The demographic and comparison questionnaires were developed by the study designers. The demographic instrument asked for basic demographic information and also about the participant's interest in science, nature, gaming, and storytelling media. In addition, the demographic instrument asked participants to rate themselves according to the "Bartle Test" (Bartle, 1996), which organizes players into four categories: killers, socializers, achievers, and explorers. These categories are derived from players desire to "act upon" or "interact with" either the "game world" or "other players." For example, "killers" most desire to act upon other players while explorers most desire to interact with the world itself.

The comparison questionnaire asked participants to think about and compare the two games they had just played. In addition, this questionnaire also asked about participants' intentions to play again and the various strengths and weaknesses of the two games. The comparison questionnaire asked some quantitative or yes/no questions, and several open-ended qualitative response questions.

The player experience instrument (completed twice, once for each game) was adapted from (IJsselsteijn, de Kort, Poels, Jurgelionis, & Bellotti, 2007) with only minor modification. Social items were removed since *Happy Match* and *Forgotten Island* are both single player games. Additionally, the number of overlapping questions for each construct was slightly reduced for time. The player experience instrument used Likert-style questions and was organized around the constructs describes in Section 2.6: positive affect, negative affect, flow, sensory immersion, tension, challenge, and competence. This instrument was used with the permission of its developers.

5. Results

5.1. Descriptive statistics (participant demographics)

Our demographic analysis showed that participants spent an average of 5.17 h per week on single-player video games over the month prior to participating in the study, supporting our supposition that a participant pool of computer science students would include a number of "gamers". Four participants didn't play any single-player video games, whereas six participants played 10 or more hours per week. 14 of 27 participants agreed or strongly agreed with the statement "I consider myself a gamer", and 4 disagreed or strongly disagreed.

17 participants indicated they were interested in nature-related activities (e.g., hiking, bird watching, insect collecting, or reading about nature), and 17 participants showed interest in science activities (e.g., going to science museums, reading about scientific discoveries, or watching television programs about science). However, only 6 participants answered "yes" when asked if they actively participated in science activities. 20 didn't participate in any science activities, and one person didn't reveal this piece of information.

When asked about their consumption of story-oriented entertainment media, 18 participants were interested in reading works of fiction, and 24 were interested in fictional TV programs or films.

In summary, the descriptive survey revealed that study participants could roughly be characterized as individuals with some

general interest in science, but no strong inclinations toward becoming actively involved in public science projects. On the other hand, participants were likely to play video games on a regular basis, though the nature of their interest in games varied.

5.2. Construct reliability of multi-item constructs

This within-subjects experiment collected 27 sets of data points. The reliability of the game experience questionnaire items was assessed using Cronbach's alpha calculated in SPSS 20.0. In general, reliability coefficients above 0.70 are considered "adequate", and values around 0.80 are "very good" (Kline, 2011).

Cronbach's alpha values of all the constructs were calculated. The results show that all constructs except "negative affect" have adequate or good Cronbach's Alpha scores, ranging from 0.72 to 0.90. The construct "negative affect" has a score of 0.69, which is very close to 0.70. Therefore, we consider the items within each construct consistently measure the same direction.

5.3. Results for RQ1

We primarily used qualitative data from the comparative questionnaire completed after both play sessions to address our first research question:

RQ1: *What do players perceive to be the key differences between points-based and story-based games with a purpose?*

In the comparative survey participants were asked (1) about their game preference, (2) the reasons for this preference, (3) what they perceived to be the biggest differences between the games, (4) which game they would use to classify either large or small numbers of photos and why, (5) their overall opinion of using video games to do a classification task, and (6) if they had any additional comments or opinions about the games.

Note that these questions do not ask specifically about points-based or story-based games. We were interested in collecting open feedback from players and inductively identifying key differences in this data. From our informal analysis, several themes strongly emerged.

5.3.1. Difference #1: story

A large number of players identified story, narrative, and the game world as very real differences between *Forgotten Island* and *Happy Match*. Since exploring this difference was a fundamental organizing principle for the two games, these results are unsurprising. However, we noted with interest that study participants did overwhelmingly identify the narrative and game world as key motivators to play and do the science task in *Forgotten Island*.

Table 1 shows a selection of participant statements about the value of story in *Forgotten Island* vs. *Happy Match*.

We also asked participants specifically what they perceived as important differences between *Forgotten Island* and *Happy Match*. Table 2 shows a selection of answers to this question where story was identified as an important difference between the games.

5.3.2. Difference #2: diegesis

In section two we developed a line of reasoning about developing play experiences based upon fantasy (Malone & Lepper, 1987) and using diegesis – elements of the story and world – to incentivize and engage players. We noted with interest that even though players were unaware of the term diegesis, many recognized this general idea as an organizing design principle and thought it worked well. Table 3 shows a selection of participant statements about how the story, world, and diegetic, task-based economy of *Forgotten Island* contributed to their desire to play and undertake the science task.

Table 1
Participant statements about story and the game world.

1	The game had a plot, this established a goal system
2	I think the added game elements, like having an "enemy" and being able to explore a world really made Forgotten Island stand out
3	After stopping [Forgotten Island], it left me wanting to find out more about the story and play more of the examination mini game
4	Coupled with compelling narrative and exploration, I felt engaged and had fun discovering each area of [Forgotten Island]
5	It felt like an actual game; it had a story; there was a more immediate incentive for getting classifications right... All around, [Happy Match] felt like a class activity while [Forgotten Island] felt like a game
6	I had a lot of fun playing Forgotten Island because there was an actual storyline. I enjoy adventure RPG games and it was fun to have to complete tasks and missions
7	Forgotten Island was the one that actually made me want to keep playing; I was invested in the story
8	It provided a story that did not center around science, but incorporated it in a unique and fun way
9	Forgotten Island allowed me to interact with the world and other characters whereas Happy Match was just simply a matching game
10	I liked the added component of adventure and story in [Forgotten Island]. The second game let you classify but was paired with a world to explore and a story to follow and become immersed in

Table 2
Perceived biggest differences between the two games.

1	No plot vs. an entertaining plot
2	Narrative, exploration, and player choice
3	It was a lot more fun when there is an actual storyline
4	The story included in [Forgotten Island]. The story makes it seem more like you are taking part in the game
5	The adventure based one had more creative and plot-based elements

Table 3
Statements about diegetic incentives.

1	[Forgotten Island] gave me both an incentive and end goal. When I correctly classified the moths I got to earn money and I got to progress in the story line which I found to be very interesting
2	The island game used the sorting as a feature of the game instead of having it be the game itself, which in my opinion is key because sorting pictures of moths is boring
3	I felt like there was a greater incentive to classify the moths in order to progress the story
4	Matching moths for credit in a mini game was a great idea in Forgotten Island, it was subtle. I felt like I was playing a game and not helping with research
5	It wasn't just about sorting moths, I liked the adventure aspect of it and how the moth identification was tied in... Forgotten Island made it more about the missions rather than the moths

5.3.3. Difference 3: sensory stimulation

Sensory stimulation is another motivating design feature that has been identified by game researchers (Malone & Lepper, 1987). This emerged as a third key difference identified by our participants. Table 4 shows a selection of statements about the role of sensory stimulation in *Forgotten Island* vs. *Happy Match*.

5.3.4. Difference 4: agency

Several participants provided feedback about how their perception of control, freedom, and agency was stronger in *Forgotten Island* than in *Happy Match*. These statements frequently referenced the fact that *Forgotten Island* offers many more activities to undertake than does *Happy Match*. This may be a result of both the diegetic approach to incentives in *Forgotten Island* and the

Table 4
Statements about sensory stimulation.

1	[Forgotten Island had a] . . .more soothing and relaxing atmosphere, more incentive to play. Very calming experience
2	The imaginative art style [in Forgotten Island] caught my attention immediately
3	I would rather play Forgotten Island because of its visuals, animation and comic book like interaction
4	The use of audio [in Forgotten Island] helped grab my attention more, and matched the theme for the areas of the game
5	The story and audio made all the difference
6	Forgotten Island was much closer to something that one might actually find on something like the Xbox Live Arcade store

Table 5
Statements about control and agency.

1	The second game [Forgotten Island] was an entire virtual world that I become invested in as I directed my avatar which way to go and what tasks to accomplish
2	There is narrative context around the classification [in Forgotten Island]. Therefore it gives the player control whether or not to unravel the narrative or play the classification mini game
3	There are things to do in Forgotten Island other than answer questions. I could take breaks from the classification and go do other things
4	The matching part was spread throughout and was not the main component of [Forgotten Island] so I didn't feel like I was being forced to learn. Having it built-in to a bigger overall game is a lot of fun and kept me extremely interested
5	[Forgotten Island had] Narrative, exploration, and player choice

Table 6
Statements about task efficiency.

1	I can imagine that classifying a large number of photos while using Forgotten Island could get very long
2	I felt that it was a lot easier to classify photos in Happy Match because that is what the game focused on
3	I'd much rather classify even one photo in a game built around the idea of classifying photos than in a game where the classification is just a small part of the game
4	I can be more efficient in Happy Match
5	I wouldn't have to wait through the story, assuming the only goal was to classify the photos

use of story as a structuring framework for gameplay. Table 5 shows a selection of statements about player control and agency in the two games.

5.3.5. Difference 5: efficiency

A final emergent theme was the relative efficiency of our two design approaches. Many participants recognized that *Happy Match* is a more efficient design if accomplishing the science task is of paramount importance. Table 6 shows a selection of participant statements to this effect.

In summary, qualitative participant responses revealed several key differences between points-based and story-based purposeful game design. Many participants felt that diegesis helped make tasks more interesting, and that the story-based game afforded more opportunities for sensory stimulation. Participants felt a stronger sense of control and agency in the story-based game because of the variety of activities they could accomplish during play, but also recognized that the points-based approach could be more efficient if task completion was important.

In the next section we continue to explore these themes by going more deeply into the player experiences that manifested when playing *Forgotten Island* and *Happy Match*.

5.4. Results for RQ2

We used quantitative data from the Game Experience Questionnaire (IJsselsteijn, de Kort, Poels, Jurgelionis, & Bellotti, 2007) to address our second research question:

RQ2: *How do perceived differences in the design of points-based vs. story-based games with a purpose shape the player experience?*

5.4.1. Paired-sample t-tests on player experience constructs

We averaged items within each construct and conducted paired-sample *t*-tests for each. Seven subscales of game experience (challenge, competence, flow, immersion, positive affect, negative affect, and tension) were measured and compared.

Table 7 shows the results from paired-sample *t*-tests. The results indicated that participants experienced a significantly lower level of challenge when playing *Forgotten Island* than *Happy Match* ($t = -4.58$, $p < .000$). Despite this, *Forgotten Island* performed better than *Happy Match* in most other categories, suggesting that while *Forgotten Island* may be perceived as easier than *Happy Match*, this is positive overall (perhaps because the easier level of challenge is more appropriate for achieving a flow state). It is also interesting, in that participants were asked to complete the same science task in both games, but felt that the story-based game, with many extra activities to accomplish, was still easier than the points-based game.

In *Forgotten Island*, participants felt more competent ($t = 3.85$, $p < .001$) and experienced higher levels of flow ($t = 3.93$, $p < .001$), immersion ($t = 6.40$, $p < .000$), and positive affect ($t = 4.97$, $p < .000$). Supporting our findings on flow, one player even suggested that, "I lost track of time in *Forgotten Island*."

For the negative dimensions of player experience, tension ($t = -1.66$, $p = .110$) and negative affect ($t = -0.61$, $p = .545$), there was no significant difference between the two games.

5.4.2. Cohen's d on player experience constructs

To further examine the player experiences in the two games, we calculated Cohen's *d* (Cohen, 1988) ($d = 0.20$, small; $d = 0.50$, medium; $d = 0.80$, large) to identify the effect size of mean differences.

The results show that the player experience constructs that had significant mean differences between *Forgotten Island* and *Happy*

Table 7
Results of paired sample *t*-tests (FI = *Forgotten Island*; HM = *Happy Match*).

	Mean difference	*SD*	*t*	*df*	Effect size Cohen's *d*	Sig. (2-tailed)
Challenge FI–challenge HM	−0.64	0.73	−4.58	26	0.88	0.000
Competence FI–competence HM	0.84	1.14	3.85	26	0.89	0.001
Flow FI–flow HM	0.53	0.70	3.93	26	0.54	0.001
Immersion FI–immersion HM	0.84	0.68	6.40	26	0.89	0.000
Neg. affect FI–neg. affect HM	−0.09	0.78	−0.61	26	0.13	0.545
Pos. affect FI–pos. affect HM	0.71	0.74	4.97	26	0.87	0.000
Tension FI–tension HM	−0.23	0.73	−1.66	26	0.39	0.110

Match also had large or medium effect sizes. The constructs of challenge, competence, immersion, and positive affect had large effect sizes, while the flow construct had a medium effect size. Although the difference in the tension construct between *Forgotten Island* and *Happy Match* was not significant, the effect size was between small and medium.

5.4.3. Motivation t-tests and intention to play again

We examined participants' intrinsic motivations for playing the two games. The motivation scales were adapted from the interest/enjoyment construct in Deci and Ryan's (2013) intrinsic motivation inventory (IMI) scales. The paired-sample *t*-tests show that participants felt that playing *Forgotten Island* was more interesting than playing *Happy Match* ($t = 4.86$, $p < .000$).

After playing each game, participants were asked about their intentions to play the game in the future. Though it is very difficult to project future behaviors from stated intentions, a paired *t*-test analysis showed that stated participant intentions to play *Forgotten Island* again were much higher than their intention to play *Happy Match* ($t = 3.99$, $p < .000$), with a medium effect size ($d = 0.75$).

In the post-game comparative questionnaire, the question about intentions to play again was reiterated, but reframed as a comparison of the two games. 17 participants expressed a willingness to continue playing *Forgotten Island* after leaving the experiment, while 10 indicated they would not continue to play the games. In fact, one study participant did play *Forgotten Island* all the way to its conclusion at home and after the study had concluded. For those who expressed an interest in continuing to play, all of them indicated they would choose to play *Forgotten Island*. The findings from the post-game comparative questionnaire are consistent with participants' responses to intention to play after each game session.

5.4.4. Overall game preference

In the post-game comparative questionnaire, we asked participants which game they preferred. 25 out of 27 participants preferred *Forgotten Island* to *Happy Match*, one participant didn't show any preference, and one participant preferred *Happy Match*. The degree of preference shows that participants favored *Forgotten Island* much more than *Happy Match* ($M = 4.04$, $SD = 1.29$). This is further supported by the stated intentions to play again presented in Section 5.4.3 and our qualitative data on the differences between games, presented in Section 5.3.

5.5. Results for RQ3

We used a combination of quantitative and qualitative data from all four surveys to address our third and final research question:

> RQ3: *How do perceived differences in the design of points-based vs. story-based games with a purpose impact players' perceptions of working on the embedded science tasks?*

5.5.1. Task structure

In the post-game comparative questionnaire, we asked about participants' game preferences with respect to the classification task presentation and the scope of the task they were asked to do. When asked about classifying just a *few* photos, 6 participants expressed that they would prefer to use *Happy Match*, whereas 20 would prefer to play *Forgotten Island*. One participant did not show any preference. When asked about classifying *many* photos, 12 participants expressed that they would prefer to use *Happy Match*. 15 participants expressed that they would prefer to play *Forgotten Island*. This suggests further participant recognition that there are efficiency advantages in points-based design approaches.

Recall from Section 5.3, however, that players expressed a greater sense of agency and control in *Forgotten Island*, and suggested that they felt more inclined to undertake the classification task when they had alternative activities to choose from as they played. Participants also indicated that the story, rewards system, and sensory stimulation all contributed to making the science task in *Forgotten Island* feel less like work and more like play. This may help to explain why more than half of our participants remained interested in classifying with *Forgotten Island* vs. *Happy Match*, no matter how many photos were to be classified.

5.5.2. Game impact on overall interest in nature and science

In our first (demographic) questionnaire we asked participants to indicate their overall interest in nature and science activities using two separate Likert-style questions. These questions were repeated in the game experience questionnaire completed after each play session. We conducted paired-samples *t*-tests to examine whether participant interest in science or nature changed after playing either game.

Our results were mostly not significant, but did show two significant findings (see Table 8): (1) participant interest in nature activities was significantly reduced after playing *Happy Match* ($t = -2.60$, $p < .05$) and (2) player interest in nature activities was significantly higher after playing *Forgotten Island* vs. *Happy Match* ($t = 2.84$, $p < .01$).

These results suggest that there may actually be risks associated with deploying points-based, enthusiast-oriented games in hopes of attracting participants with no special interest in the science activity. In our study, participants were turned off by our more enthusiast-oriented game, even to the point of a negative effect on their interest in nature activities. Though our story-oriented game did not have a significant positive effect on participant interest in nature, it also produced no negative effects.

5.6. Comparisons to the live Citizen Sort system

Though in this research we primarily focused on player perceptions of two different gamification approaches using an experimental approach, we also provide data from the live Citizen Sort system in order to contextualize our findings and address concerns about the quality of data collected using these two games.

Table 8
Results of paired sample *t*-tests (FI = *Forgotten Island*; HM = *Happy Match*); non-significant results excluded.

	Paired differences					*t*	*df*	Sig. (2-tailed)
	Mean	Std. deviation	Std. error mean	95% Confidence interval of the difference				
				Lower	Upper			
Pair 1: Interest in nature after playing HM–initial interest in nature	−.370	.742	.143	−.664	−.077	−2.595	26	.015
Pair 2: Interest in nature after playing FI–interest in nature after playing HM	.296	.542	.104	.082	.511	2.842	26	.009

Table 9
Participant classification accuracy vs. live system classification accuracy (FI = *Forgotten Island*; HM = *Happy Match*).

	Happy Match		Forgotten Island	
Participants	Study	Live	Study	Live
Mean	0.754	0.789	0.801	0.779
Median	0.763	0.800	0.800	0.806

5.6.1. Citizen Sort overview

Citizen Sort is available online and can be played for free with an account. Currently, the system provides different versions of *Happy Match* and *Forgotten Island*, including moths, sharks, and rays. So far it hosts nearly 5000 players, of which 1546 have played the moth version of *Happy Match* and 1035 have played the moth version of *Forgotten Island* (the versions used in our experimental study).

Over 500 more participants have chosen to play *Happy Match* vs. *Forgotten Island*, implying a reversed preference for the games from our controlled study. *Citizen Sort* has been promoted significantly in naturalist media (e.g. scistarter.org, Scientific American Online, etc.) but very little in gaming-oriented media. It is very possible that the live system therefore attracts a different player demographic than the "gamers" we targeted for this paper, and the preference for *Happy Match* suggests that this game is better at attracting players with an interest in nature and science than *Forgotten Island*.

5.6.2. Data quality

A key concern for scientists is that crowdsourced data be accurate. This issue has been addressed at some length in several prior publications based on the *Citizen Sort* project (Crowston & Prestopnik, 2013; Prestopnik & Crowston, 2011; Prestopnik, Crowston, & Wang, 2014).

In this study, we further address this issue, albeit briefly, by retrieving play data from our experiment participants and providing a brief comparison of accuracy data from these participants to data from players in the live *Citizen Sort* system (see Table 9):

Participants in our study played each game briefly and so made only a few classification decisions that could be checked against the gold standard data (*Happy Match* mean # of decisions 43, median 38; *Forgotten Island* mean # of decisions 37; median 16). It is possible overall accuracy could change given more extended play, but the accuracy of our study participants matches well with overall accuracy in the live system (*Happy Match* mean # of decisions 77.7, median 39; *Forgotten Island* mean # of decisions 137, median 36).

6. Discussion

6.1. Limitations

Before discussing our findings, we note several limitations to be considered before interpreting our results. First, we conducted an analysis of our qualitative data using basic categorization and interpretation strategies. The trends we describe in section five definitely stood out, even using relatively informal analyses. Nonetheless, more structured approaches such as content analysis could be useful for extracting subtle themes or counter-examples embedded within the data.

Second, *Happy Match* and *Forgotten Island* were envisioned by the same designer, developed by the same design team, and finished at the same time. This theoretically makes them comparable in terms of overall quality. However, the reality of game design is that it is a complex activity where even proven commercial designers and teams struggle to consistently achieve a high level of quality. It is possible to conceive of a points-based game that is better designed than a story-based game, despite what our data show about *Happy Match* and *Forgotten Island*.

Nonetheless, in this study, participants noted that both games were polished and felt professionally designed. The major complaint about the polish and professionalism of the games actually related to *Forgotten Island*, not *Happy Match*. This was a criticism that the movement controls for the player character in *Forgotten Island* were clumsy and fatiguing to use, however all but two participants preferred this game anyway. The main stated reason was that *Forgotten Island's* story and world created compelling motivations to do the science task, not because of any disparity in overall perceived quality between the two games.

Third, to approximate a population of "gamers" we recruited students who were taking computer science courses. It would be interesting to narrow our definition of "gamers" and explore how this population compares to a specifically sampled population of naturalists in their reactions to various game design approaches. Though we detect some differences in the live data from Citizen Sort system, comparing to our experimental study, a more controlled approach would help us identify different strategies for scientists to consider when deploying purposeful gaming projects (Bowser et al., 2014).

Finally, this study was conducted in a controlled environment with a relatively small number of participants. The research findings reflect participants' preferences for *Happy Match* or *Forgotten Island* after a short period interaction with the game. Although we asked for their intentions to continue playing, it is difficult to generalize any conclusive statements about their actual behaviors over time based on their answers to their intentions. We consider that asking about participant intentions is still a valuable way to collect opinions about the two citizen science games. However, we hesitate to make very broad claims about the nature of long-term engagement with story-based or points-based games with a purpose based on findings from this one study alone.

6.2. Story is a powerful tool for reframing tasks

A variety of participants indicated that *Forgotten Island's* story, world, and rewards structure made it a compelling experience that felt "more like a game" than *Happy Match*. Moreover, many participants also suggested that the story, world, and rewards made the science task feel easier and made them more interested in doing it when asked. Coupled with the strong overall player preference for our story-based implementation over our points-based one, this is a powerful indication that story-based games should have a place in the pantheon of games with a purpose, especially if they are intended to attract participants who have limited interest in the embedded science activity.

Our participant sample was drawn from college students taking computer science courses. Many of these students held only a passing interest in science and nature activities, but indicated that they were active players of video games. Our results suggest that such players are unlikely to be attracted to points-based citizen science games that emphasize the science task, but could be attracted by story-based games that diegetically motivate the science task as part of a larger play experience. Yet reframing a science task using story and diegesis may not be appropriate for all demographics.

For example, altruism, learning, and interest in science remain strong motivators for many citizen science participants. A story-based game may well be a distraction for these individuals, who will care less about the game world and story than they will about the task. Here, points-based games (or no game at all) may provide more efficient paths for becoming involved in a science activity of interest. However, by reframing the science activity as a diegetically motivated mission set in a rich, story-based world,

story-based games may hope to recruit new kinds of players who would otherwise remain uninterested and uninvolved.

6.3. Story-based games with a purpose present significant design challenges

Though there may be player experience advantages to story-based games with a purpose, there are also a number of significant challenges to be considered when designing and implementing such games. Unsurprisingly, story-based games require their designers to craft great stories. Though not a technical exercise, this is nonetheless non-trivial, and requires deep knowledge of story structure, rising and falling action, how protagonists and antagonists interact or conflict with each other, and much more (e.g., Vogler, 2007). This is to say that there are well-understood guidelines for creating engaging, interesting stories. A poorly crafted story is unlikely to accomplish much of value in the context of any purposeful game. Of course, implementing even a great story into an interactive game is no mean feat.

Every story is situated in a world, and in video games, this world is especially important. It is the environment in which players are exposed to narrative, undertake exploration, experience conflict, pursue goals, and overcome obstacles. Crafting great worlds can be as challenging as crafting great stories, both from an aesthetic and a technical standpoint. In terms of aesthetics, compelling worlds require great creativity to envision, and the combined skills of an architect, artist, cinematographer, and engineer to present convincingly. From a technical standpoint, great world design also requires great level design – great games are the result of deep thinking about how environments are organized and arranged to maximize their entertainment value and indirectly control the player experience (Schell, 2008).

In Section 5.3.3 we noted how many participants mentioned that the sensory experience of *Forgotten Island* had an impact on their game preference. This is another challenge of creating great games, both points-based and story-based. However, creating a great sensory experience is particularly challenging in story-based games because visuals and sound tend to be more sophisticated and also more representational. In story-based games, graphics and sound effects are mostly diegetic, and thus represent real things like scenery, flowing water, animated characters, spoken dialog, and much more. In points-based games, the graphics and sound, even when very well done, are more often are used to convey interface information like clicks, drags, rewards, and the like.

Adding a "real world" activity into an entertainment-oriented experience creates challenges of its own. The classification task in *Happy Match* and *Forgotten Island* was not perceived by most participants in our study to be inherently fun, giving us a wicked problem (Rittel & Webber, 1984) to contend with: embed a difficult science activity into a broadly appealing game without spoiling either the game experience or the science. Our solution in *Happy Match*, a non-diegetic points-based approach, appears to work for players who already have an interest in nature and science. Many players in our controlled study critiqued *Happy Match*, however, because to them it felt very little like a game. *Forgotten Island* was better liked by our participants because it's various play activities made the science task more tolerable. It would be hard to argue, however, that *Forgotten Island* made the science of taxonomic classification truly fun in itself. In the realm of games with a purpose, the task itself does matter; some tasks may be inherently more viable for crafting great play experiences than others. This may help to explain the success of some very successful point and badge-based games that feature tasks and activities that are more inherently fun for participants.

In short, designing and implementing a story-based game with a purpose is a true challenge, and not undertaken lightly. It requires a development team with a good mix of relevant skills, and very often a designer with an overarching vision for the entire experience. Addressing these challenges and requirements can be costly and difficult, but the payoff may be in attracting more participants, or participants who come to the purposeful activity with different motivations than enthusiast players.

6.4. The case for fantasy

Participant altruism and learning have long been a strong force for engagement in many citizen science projects (Bradford & Israel, 2004; Franzoni & Sauermann, 2014; King & Lynch, 1998; Raddick et al., 2009, 2010). Many citizen science participants get involved because they are interested in a specific scientific field, interested in science more generally, or interested in doing something to help scientists or their community. Given their existing interests, these participants may not even require the engaging power of games to get involved, or may be satisfied with points-based recognition for their effort and participation.

The proven success of the many citizen science games (see Section 2.1) that follow a points-based model demonstrates that large crowds can be attracted by non-diegetic, point-based games. However, the current focus on altruist and science enthusiast player demographics by contemporary designers leaves several interesting possibilities for games with a purpose untapped.

First, "gamers" – many of whom care far more about play than about science – make up a great reserve of potential participants, if only their interest can be attracted. More numerous than science enthusiasts, gamers seem more likely to be captured by games with diegetic features: immersive game worlds, progressively unfolding narratives, and sensory stimulation through rich graphics, animation, and sound. Many millions more people play video games than participate in citizen science projects, and players spend countless hours engaging with game worlds instead of the real world. This is the vast potential to do great things with games that inspires designers, scholars, and innovators alike, yet much of this promise has been left unrealized.

Second, the use of stories in games with a purpose may have implications for changing the typical distribution of effort seen in crowdsourced activities. Often, the majority of the work is contributed by a minority of players. Stories, if they are interesting enough, may be able to engage players in more work than they would normally contribute; stories, even in games, are fixed and finite. Players who are engaged in the story and intend to finish the experience, may find themselves similarly contributing in a fixed, finite, but substantive way, possibly "flattening" the usual distribution of work into just two groups: those who have contributed and those who have not.

Finally, a focus on stories, worlds, fantasy, and diegesis opens up a host of possibilities for the design of new kinds of engaging and meaningful play experiences that are as flexible as storytelling itself. As human beings, we craft and are engrossed by stories of all kinds, including action, adventure, romance, drama, mystery, and much more. These many storytelling genres can themselves underpin virtually endless possibilities for different styles of games: fighting, first-person shooters (FPS), role-playing games (RPGs), simulators, real-time strategy games, point-and-click adventures, racing, sports, strategy, puzzle, and more (Qin et al., 2009). A new paradigm of purposeful game design, one structured around games that feel more like mainstream entertainment titles than like gamified tasks, might begin attracting players with more interest in play than in science. This would be a truly sizeable crowd indeed.

7. Conclusions and future research

We designed this study to explore the differences between story-based and points-based citizen science games, imagining that different gamification approaches might vary in their impact on player experiences and perceptions science tasks.

We found that the story-based game, *Forgotten Island*, resulted in a significantly more compelling player experience for our predominantly "gamer" participants than did *Happy Match*, a points-based, non-fantasy game. Participants stated that this was because of the story, game world, and diegetic rewards designed into *Forgotten Island.* These things made the science task more palatable and interesting than in *Happy Match*, and also gave them various reasons to continue play.

Our findings lend credence to the supposition that story-based games can be a powerful tool for attracting players who are not inherently interested in science, nature, or whatever other task. Prior studies have identified that participants with various inherent interests may react differently to different gamification approaches (Bowser et al., 2014). For future research, investigating how these different approaches can affect players with varied interest in science, nature, and gaming over the long term may help us to provide a more complete picture of this complex design space. It is our hope that interactive storytelling will become a more common element of many citizen science games, engaging more people and resulting in more potent crowd-based scientific inquiry.

Acknowledgments

The authors would like to thank *Citizen Sort's* principal investigators Kevin Crowston and Jun Wang for their support and guidance, as well as the development team for their efforts on this project: Nathan Brown, Chris Duarte, Susan Furest, Yang Liu, Supriya Mane, Nitin Mule, Gongying Pu, Trupti Rane, Jimit Shah, Sheila Sicilia, Jessica Smith, Dania Souid, Peiyuan Sun, Xueqing Xuan, Shu Zhang, and Zhiruo Zhao. The authors would also like to thank the following for their advice and assistance: Jennifer Hammock, Nancy Lowe, John Pickering, and Andrea Wiggins. This work was partially supported by the US National Science Foundation under grant SOCS 09-68470.

Appendix A

A.1. Chronbach's Alpha for constructs in the Game Experience Questionnaire

	Items used in the study	*M*	*SD*	Chronbach's Alpha
Challenge 1	I thought it was hard	2.42	1.13	
Challenge 2	I felt challenged	3.13	1.21	0.72
Challenge 3	I felt time pressure	1.55	0.82	
Challenge 4	I put a lot of effort into it	2.89	1.12	
Competence 1	I felt skillful	2.76	1.15	
Competence 2	I felt competent	3.37	1.25	0.87
Competence 3	I was good at it	3.17	1.26	
Competence 4	I felt successful	3.20	1.12	
Competence 5	I was fast at reaching the game's targets	3.24	1.18	
Flow 1	I was fully occupied with the game	3.52	1.09	
Flow 2	I forgot everything around me	2.35	1.23	0.88
Flow 3	I lost track of time	2.39	1.31	
Flow 4	I was deeply concentrated in the game	3.57	1.14	
Flow 5	I lost connection with the outside world	2.52	1.31	
Immersion 2	It was aesthetically pleasing	3.74	1.01	
Immersion 3	I felt imaginative	2.69	1.26	0.89
Immersion 4	I felt that I could explore things	3.11	1.46	
Immersion 5	I found it impressive	3.20	1.05	
Immersion 6	It felt like a rich experience	3.11	1.25	
Neg. Affect 2	I thought about other things	1.94	0.86	
Neg. Affect 3	I found it tiresome	1.98	0.89	0.69
Neg. Affect 4	I felt bored	1.87	0.88	
Pos. Affect 1	I felt content	2.94	0.99	
Pos. Affect 2	I thought it was fun	3.13	1.04	0.90
Pos. Affect 3	I felt happy	3.00	0.98	
Pos. Affect 4	I felt good	3.21	1.06	
Pos. Affect 5	I enjoyed it	3.30	1.15	
Tension 1	I felt annoyed	1.63	0.76	
Tension 2	I felt pressured	1.87	1.01	0.72
Tension 3	I felt irritable	1.52	0.75	
Tension 4	I felt frustrated	1.63	0.76	
Interest 1	I enjoyed doing this activity very much	3.06	0.93	
Interest 2	I would describe this activity as very interesting	3.15	0.97	0.88
Interest 3	This activity did not hold my attention at all (rev. coded)	4.17	0.89	
Intention 1	I have no intention to play this game again (rev. coded)	2.93	1.23	0.88
Intention 2	I intend to play this game again	2.48	1.11	

References

Anderson, C. (2008). *The long tail: Why the future of business is selling less of more*. New York, NY: Hyperion.

Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD Research, 1*(1), 19.

Berkovsky, S., Coombe, M., Freyne, J., Bhandari, D., & Baghaei, N. (2010). Physical activity motivating games: Virtual rewards for real activity. *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*.

Bogost, I. (2011). Persuasive Games: Exploitationware. *Gamasutra: The Art and Business of Making Games*. <http://goo.gl/jK1VR>.

Bowser, A., Hansen, D., Preece, J., He, Y., Boston, C., & Hammock, J. (2014). Gamifying citizen science: A study of two user groups. *Paper presented at the Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & #38; social computing*, Baltimore, Maryland, USA.

Bradford, B. M., & Israel, G. D. (2004). *Evaluating Volunteer Motivation for Sea Turtle Conservation in Florida*. University of Florida, Agriculture Education and Communication Department, Institute of Agriculture and Food Sciences. Gainesville, FL.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624–634.

Chen, Z.-H., & Chen, S. Y. (2013). A surrogate competition approach to enhancing game-based learning. *ACM Transactions on Computer–Human Interaction, 20*(6), 1–24. http://dx.doi.org/10.1145/2524264.

Cohn, J. P. (2008). Citizen science: Can volunteers do real research? *BioScience, 58*(3), 192–197.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crowston, K., & Prestopnik, N. (2013). Motivation and Data Quality in a Citizen Science Game: A Design Science Evaluation. *Paper presented at the Hawaii International Conference on System Sciences (HICSS)*, Wailea, Maui, HI.

Costikyan, G. (2013). *Uncertainty in games*. Canbridge, Massachussetts: The MIT Press.

Cox, A., Cairns, P., Shah, P., & Carroll, M. (2012). Not doing but thinking: The role of challenge in the gaming experience. *Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems*, Austin, Texas, USA.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. New York, NY: Harper Perennial.

Deci, E. L., & Ryan, R. M. (2013). Intrinsic Motivation Inventory (IMI). Retrieved Feb. 27, 2013, from http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/.

De Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education, 46*(3), 249–264. http://dx.doi.org/10.1016/j.compedu.2005.11.007.

Deterding, S. (2012). Gamification: Designing for motivation. *Interactions, 19*(4), 14–17.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". *Paper presented at the proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, Tampere, Finland.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. *Paper presented at the CHI 2011, Extended abstracts on human factors in computing systems*, Vancouver, BC, Canada.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. http://dx.doi.org/10.3758/brm.41.4.1149.

Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., & Mandryk, R. L. (2011). Calibration games: making calibration tasks enjoyable by adding motivating game elements. *Paper presented at the Proceedings of the 24th annual ACM symposium on User interface software and technology*.

Franzoni, C., & Sauermann, H. (2014). Crowd science: The organization of scientific research in open collaborative projects. *Research Policy, 43*(1), 1–20.

Gajadhar, B., Kort, Y. d., & IJsselsteijn, W. (2008). Influence of social setting on player experience of digital games. *Paper presented at the CHI '08 Extended Abstracts on Human Factors in Computing Systems*, Florence, Italy.

Galloway, A. R. (2006). Gaming: Essays on Algorithmic Culture (Electronic Mediations): University of Minnesota Press.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467. http://dx.doi.org/10.1177/1046878102238607.

IJsselsteijn, W., de Kort, Y., Poels, K., Jurgelionis, A., & Bellotti, F. (2007). Characterising and measuring user experiences in digital games. *Paper presented at the ACE Conference*, Salzburg, Austria.

Jafarinaimi, N. (2012). Exploring the character of participation in social media: the case of Google Image Labeler. *Paper presented at the Proceedings of the 2012 iConference*, Toronto, Ontario, Canada.

King, K., & Lynch, C. V. (1998). The motivation of volunteers in the nature conservancy – Ohio chapter, a non-profit environmental organization. *Journal of Volunteer Administration, 16*(5).

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

Malone, T. W. (1980). What makes things fun to learn? heuristics for designing instructional computer games. *Paper presented at the Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*, Palo Alto, California, United States.

Malone, T. W. (1982). Heuristics for designing enjoyable user interfaces: Lessons from computer games. *Paper presented at the Proceedings of the 1982 conference on Human factors in computing systems*, Gaithersburg, Maryland, United States.

Malone, T. W., & Lepper, M. (1987). *Making learning fun: A taxonomy of intrinsic motivations for learning*. Hills-Dale, NJ: Erlbaum.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. New York: Penguin Press.

Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers, 22*(5), 336–343.

Overmars, M. (2004). Teaching computer science through game design. *Computer, 37*, 81–83.

Prensky, M. (2005). Computer games and learning: Digital game-based learning. *Handbook of computer game studies, 18*, 97–122.

Prestopnik, N., & Crowston, K. (2011). Gaming for (Citizen) Science: Exploring Motivation and Data Quality in the Context of Crowdsourced Science Through the Design and Evaluation of a Social-Computational System. *Paper presented at the 7th IEEE International Conference on e-Science*, Stockholm, Sweden.

Prestopnik, N., & Crowston, K. (2012). Purposeful Gaming & Socio-Computational Systems: A Citizen Science Design Case. *Paper presented at the ACM Group: International Conference on Supporting Group Work*, Sanibel Is., FL.

Prestopnik, N., Crowston, K., & Wang, J. (2014, 4–7 March, 2014). Exploring Data Quality in Games With a Purpose. *Paper presented at the iConference*, Berlin, Germany.

Qin, H., Rau, P.-L. P., & Gavriel, S. (2009). Measuring player immersion in the computer game narrative. *International Journal of Human–Computer Interaction, 25*(2), 107–133.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., et al. (2010). Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review, 9*(1), 010103–010118.

Raddick, M. J., Bracey, G., Carney, K., Gyuk, G., Borne, K., Wallin, J., & Jacoby, S. (2009). Citizen science: status and research directions for the coming decade. *AGB Stars and Related Phenomenastro 2010: The Astronomy and Astrophysics Decadal Survey* (vol. 2010, pp. 46P).

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development, 44*, 43–58.

Rittel, H. J., & Webber, M. M. (1984). Planning problems are wicked problems. In N. Cross (Ed.), *Developments in design methodology*. New York: John Wiley & Sons.

Rogers, S. (2010). *Level up: The guide to great video game design*. Chichester, West Sussex, UK: John Wiley and Sons Ltd..

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, MA: The MIT Press.

Schell, J. (2008). *The art of game design: A book of lenses*. Burlington, MA: Elsevier Inc..

Schild, J., LaViola, J., & Masuch, M. (2012). Understanding user experience in stereoscopic 3D games. *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*, Austin, Texas, USA.

Schouten, D., Pfab, I., Cremers, A., van Dijk, B., & Neerincx, M. (2014). Gamification for low-literates: Findings on motivation, user experience, and study design. In K. Miesenberger, D. Fels, D. Archambault, P. Peňáz & W. Zagler (Eds.), *Computers Helping People with Special Needs* (vol. 8547, pp. 494–501). Springer International Publishing.

Stam, R., Burgoyne, R., & Flitterman-Lewis, S. (1992). *New vocabularies in film semiotics*. London: Routledge.

Vogler, C. (2007). *The Writer's Journey*. Saline, MI: McNaughton & Gunn Inc..

von Ahn, L. (2006). Games with a purpose. *Computer, 39*(6), 92–94.

von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM, 51*(8), 58–67. http://dx.doi.org/10.1145/1378704.1378719.

Webster, J., Trevino, L. K., & Ryan, L. (1994). The dimensionality and correlates of flow in human–computer interactions. *Computers in Human Behavior, 9*(4), 411–426.

Wiggins, A., & Crowston, K. (2011, January 04–January 07, 2011). From conservation to crowdsourcing: A typology of citizen science. *Paper presented at the 44th Hawaii international conference on system sciences*, Kauai, Hawaii.