Pathfinder: The Behavioural and Motivational Effects of Collectibles in Gamified Software Training

TIM NAGLÉ, Eindhoven University of Technology, Netherlands

SCOTT BATEMAN, University of New Brunswick, Canada

MAX V. BIRK, Eindhoven University of Technology, Netherlands

Designers of instructional software use gamification to help motivate and engage learners. Typically focusing on gamifying a single task, designers aim to provide a straightforward path through learning. In contrast, video games frequently provide optional secondary tasks using collectibles. Collectibles—like coins—are secondary, non-essential goals that encourage players to selectively take on additional challenges and engage more with a game. While research supports the idea that by increasing engagement learning can be improved, exactly how collectibles—an extremely common element in games—might be employed in gamified learning and how it might affect the play experience is underexplored. We present the results of a study comparing a gamified photo-editing training game that uses collectibles to one without collectibles. Our results show that learners choose to engage more when collectibles are present, and that this has a positive effect on software skills applied to a representative out-of-game challenge. Our findings provide a nuanced view of the tradeoffs in motivation and experience when collectibles are used.

CCS Concepts: • Human-centered computing → Human computer interaction (HCI); *Empirical studies in HCI*; • Human-centered computing → Human computer interaction (HCI) → Interaction paradigms; *Graphical user interfaces*;

KEYWORDS: Video games, tutorials, learning, gamification

ACM Reference format:

Tim Naglé, Scott Bateman, Max V. Birk. 2021. Pathfinder: The Behavioural and Motivational Effects of Collectibles in Gamified Software Training. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, CHI PLAY, Article 264 (September 2021), 23 pages, <https://doi.org/10.1145/3474691>

1 INTRODUCTION

Learning design software such as Adobe Photoshop, Rhinoceros 3D, or Autodesk's AutoCAD, is time intensive and requires new users to explore a myriad of features. The complexity of design software can turn users away or prevents them from learning about powerful software features [28]. Tools that aid users to learn design software have taken different forms, including online tutorials [40], software forums [38], or simplified interfaces [31]. Designers have used gamification—the use of game elements in a non-game context [17]—to increase engagement with interactive tutorials. Tutorials built into the target application (such as Jigsaw being built into Photoshop [18] or GamiCAD being built into AutoCAD [43]) have demonstrated the potential of game-based design to foster software learning. Motivation and engagement with instructional software are crucial to facilitate effective learning [22]. However, in most cases, gamified learning

This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2021 Copyright held by the owner/author(s).

2573-0142/2021/9-ART255. <https://doi.org/10.1145/3474691>

tools have offered a single sequential path through learning activities. While successful, previous work in gamified tutorials have most often provided simple learning paths; e.g., in GamiCAD [43] and Jigsaw [18] players progress step-by-step through a list of predefined challenges. This approach stands in contrast to many games that provide optional challenges designed to allow players to engage with optional game challenges, e.g., jumping to the top of the flagpole in *Super Mario Bros.* (Nintendo, 1985) or collecting Band-Aids in *Super Meat Boy* (Team Meat, 2010). These optional challenges allow players to choose to take on more challenge, to increase their own engagement, if they feel the need. When learners are not challenged or otherwise engaged with learning tasks as much as they could be, they might become disengaged, which is detrimental for learning [22]. This is because it is theorized that engagement affects motivation to learn new skills [61]. Thus, by increasing engagement, we can increase motivation, which could have a positive effect on learning.

Optional game challenges are often incentivized by providing non-essential rewards to players for taking on additional challenges: providing reward animations and sounds, revealing new story details, or unlocking characters to provide replay value [60]. Collectibles are in-game items that are often connected to non-essential rewards (e.g., coins, gems, etc.) that can be gathered while playing. Once enough collectibles have been gathered a reward is given. In most games, collectibles are non-essential for completing the core game [20] and game designers use them for a wide range of reasons, including: improving player engagement, encouraging exploration of unknown areas of a level, giving players breaks between missions, reinforcing certain behavior or play patterns, and encouraging practice of important new skills [13,20]. For example, Korok Seeds in *The Legend of Zelda: Breath of the Wild* (Nintendo, 2017) are hidden throughout the world and often require players to practice and creatively employ game mechanics to find and gather them. Gathering enough Korok Seeds allows the player to trade them for additional weapon-carrying or item capacity (a useful advantage during gameplay). However, collecting all 900 seeds provides the players with a non-usable, non-essential trophy [58].

Considering the ability of collectibles to foster engagement and encourage practicing new skills in commercial games, it is reasonable to believe that they could similarly be used in gamified instructional software to increase learner engagement.

While the results are promising [18,43], our understanding of the motivational and behavioural effects of individual game design elements such as collectibles is still scarce [1,24,53,61,68]. Previous investigations of game design elements have shown the motivational implications of badges [30], behavioural implications of avatar customization [6], or the effects that health bar manipulations can have on learning and performance [79]. Researchers on educational systems are concerned that when game-elements are misunderstood and incorrectly applied, that they might end up with systems that “... neither instruct nor engage ...” [23]. Hence, understanding the impact of non-essential game-elements on gamified software tutorials requires further empirical research to evaluate the effects of game-elements on motivation, experience and behaviour.

To deepen the understanding of the effect that collectibles can have for instructional software, we present Pathfinder, an instructional game that builds on the work of Jigsaw [18] and GamiCAD [43], and on empirical work on collectibles [1,24]. Pathfinder closely reproduces the interface and functionality in Adobe Photoshop (using the Unity 3D game-engine), enabling us to use game-development tools (e.g., physics engine, asset libraries, web export) and re-create basic photoshop editing features (i.e., move, brush, and layers).

Using Pathfinder, we investigated the effects of collectibles (i.e., coins) on user behaviour and motivation in a between-subject study. 60 participants were recruited on Amazon Mechanical Turk and were randomly assigned to play Pathfinder either without collectibles or with collectibles. After completing 12 levels of Pathfinder, both groups performed an out-of-game editing challenge.

Our results show that adding collectibles to Pathfinder increased time spent and behavioural engagement. We show that the addition of coins leads to improved performance in an out-of-

game challenge, and that the number of coins collected predicts out-of-game performance. Interestingly, the demand of the coin condition was experienced as higher, while perceived enjoyment and effort did not differ between groups. Our work contributes to our understanding of designing game-based tutorials, the use of collectibles in game design, and our understanding of game-based motivational design.

2 Related Work

Instructional software is an effective way to train people to use software. Our background research highlights design approaches for tutorial design, and focuses on game-based software tutorials. We investigate motivational design approaches and their role in designing engaging software tutorials. We focus on the use of collectibles, and coins in particular, to situate our work and highlight previous research efforts.

2.1 Software Tutorials

Software tutorials teach people the essentials of software using instructional examples. Adobe, for example, offers a range of example courses for Photoshop¹ such as “Work with layers”, “Make selections”, or “Change image size”. These tutorials provide example files and step by step instructions to complete the exercise. Tinkercad² pursues an approach integrated into the software itself. Tinkercad leverages game design strategies such as 1) the statement of a challenge, e.g., “Place it!”, referring to a simple box shape that should be placed; 2) feedback to guide the learner, e.g., such as a semi-transparent box that signals where to place the item; 3) the expectation to act instead of passively observing; and, 4) a clear signal when the goal, i.e., placing the box, is achieved.

Within HCI, research on computer-based learning [54] and educational software [12] has explored education and learning systems broadly; examples are Massive Open Online Courses (MOOCs) [29], or specific learning environments such as online tutorials [40] or software forums [38]. The focus here is primarily on fostering knowledge about a specific topic. Educational software has also drawn from game-based approaches [71], for example, by exploring challenge-based, competitive coding [80], or by providing narrative [72], visual embellishment, progressing goals, and achievement feedback [12].

Software tutorials focus on teaching learners specific skills or procedures that can then be applied. For example, learning how to use layers and object selection enables user to apply these skills in a wide range of image editing tasks. Besides understanding the software and the graphical user interface (GUI), learners also need to understand the specifics of the software, e.g., that the top layer indicated in the layer menu corresponds to a group of items that are placed on top of all other items.

While instructional videos that are used alongside an application may be the most common form that people learn software with, e.g., watching a YouTube video, tasks integrated into software have advantages for transfer learning. In instructional videos users watch the content, extract relevant information, and then try to apply what they have learned within the software. However, unless the learner is repeating the exact same task as in the video, there is no feedback provided and the quality of the execution is up to the learner’s judgement. Researchers have also addressed affective aspects of software learning by aiming to boost confidence and increase engagement through feedback [21] and community support [40].

¹ <https://helpx.adobe.com/photoshop/tutorials.html>, accessed February 16th, 2021

² <https://www.tinkercad.com/learn/designs>, accessed February 16th, 2021

Software learning has been described as the transitional process from being a novice to becoming experts [14]. Research on user interfaces [67], keyboard shortcuts [35], or improving command access [76], has highlighted the importance of developing a conceptual model of the software, building up muscle memory, and understanding the difference between tools and techniques within software. Research on video game tutorials shows that tutorials are particularly valuable for games with mechanics that cannot simply be discovered through experimentation [3]—suggesting that tutorials are particularly valuable for complex software.

Similar to the previously described research, work on advancing software tutorials has also focused on increasing engagement, improving the learning transfer between tutorial and software use, and supporting people to more rapidly transition from novices to experts [14]. Learning in video games, and hence video games tutorials, share several common properties with software tutorials, e.g., the high complexity of commands and the need for rapidly progressing learning. Because of the alignment between goals, game-based tutorials have been a fruitful source for innovating software tutorials.

2.1.1 Gamification of Software Tutorials. Video game tutorials are comparable to software tutorials, but focus on quickly picking up game mechanics typically in the context of entertainment, as opposed to skills to improve productivity. Learning how to shoot, change to a grenade, jump and crouch, or use the shops in *Call of Duty: Warzone* (Activision, 2020) requires learning a complex sequence of commands. This is not dissimilar to the interfaces for changing the size, colour, and position of an image, and understanding image exporting functions in photo editing software. Video games also leverage design patterns that have been shown to be effective for goal-setting [41], to create reward schedules [45], or to foster voluntary engagement [6]. While video game tutorials offer insights into designing tutorials, game-elements have also been directly applied within instructional software.

Dong et al. [18] investigated *Jigsaw*, a game implemented in Photoshop, that asks users to solve a number of jigsaw puzzles with the use of Photoshop tools. Their results show that using exploration and discovery based learning in software tutorials can be a useful tool for learning when combined with other demonstration tools, such as in-game help and step-by-step demonstrations for complete beginners [18]. Li, Grossman & Fitzmaurice [43] developed *GamiCAD*, a gamified AutoCAD tutorial that uses a path approach—users were asked, for example, to draw paths using line drawing tools or trim lines using the trim tool. The results show that *GamiCAD* increased performance and user engagement compared to traditional tutorials [43]. The 3D modelling software *TinkerCAD*³ (Autodesk, 2011) implements game-based approaches similar to *GamiCAD* in a productive environment, showing the advantages of game-based tutorials for software learning in commercial software.

Gamified tutorials such as *GamiCAD* [43] and *Jigsaw* [18] combine the motivational interface design elements of video games and create learning opportunities that draw learners in and assure the use of skills within the software context, and therefore reduce transfer-based skill loss when moving from a video tutorial to the software.

Designing tutorials is a complex task and the influence individual design patterns have on engagement is not well understood. However, past research has convincingly demonstrated that employing game elements (i.e., gamifying learning tasks) shows positive effects on self-reported and behavioural engagement [7,21,57].

2.2 Motivation and Learning

Motivational theories such as goal-setting [44] or self-determination theory [62,63], provide actionable insights for designing learning environments. Goal-setting theory [44] suggests that

³ <https://www.tinkercad.com/>, accessed February 16th, 2021

specific goals lead to performance improvements. Goal-setting theory is commonly applied by formulating SMART goals (Specific and Measurable, Achievable, Relevant, and Time-bound). Collectibles fulfill most of the SMART criteria, they are: specific (e.g., find all coins); measurable (collected or not); achievable (the goal can be difficult, but not unachievable); and time-bound (have an end, predictable in the time commitment needed). How relevant a collectible is depends on (1) the ascribed value of the collectible in the game [1], e.g., a single star in *Super Mario Bros.* (Nintendo, 1985) has more relevance than a single coin, (2) the person playing the game, and (3) the ratio between expectancy of success and value of an object [77].

Self-determination theory suggests that people are more likely to seek out environments where they can flourish [62]. The theory suggests that individuals are more motivated to search out environments that facilitate the experience of *competence*—mastery over a challenge, *autonomy*—engagement under one's own volition, and *relatedness*—feeling connected to others. The theory has been applied to learning [16] and has shown positive effects on engagement [6] and learning outcomes [61]. Previous work suggests that positive outcomes are due to an increase in learner's willingness to spend time, to complete more challenges, or due to increased interest and focus [18,43]. To achieve better engagement, learning tutorials should: a) provide opportunities for meaningful learning, b) provide a context that fosters motivation, c) promote the evaluation of new ideas, and d) support users in linking new concepts to already known concepts [37,50].

Motivation is associated with multiple positive aspects of learning, such as spiking curiosity, exploration, reflection, and self-regulatory behaviour [48]. People who are motivated to participate in an activity are more actively engaged and perform better than those who are not [16]. Game design has demonstrated the ability to foster motivation and engagement [6,7]. Because game design can foster motivation, game-elements have been successfully applied in non-game contexts, for example, to increase motivation in serious games such as MindLight [66], to engage people in citizen science [15], or in instructional software [18,43].

Researchers have investigated the effects of game elements such as points, badges, and leaderboards. Mekler et al. [52] showed the effect of points and badges on performance, but highlighted that neither increased intrinsic motivation. Landers et al. [41] and Hamari [30] demonstrated that leaderboards and badges improve goal-setting, leading to increases in performance. Serious games and gamified applications have been shown to outperform traditional learning methods in certain situations, and the inherent motivational aspects have been shown to improve intrinsic motivation and engagement in educational games [36].

2.3 Collectibles and Secondary Objectives

The motivational and behavioural effects of gamification, common game elements, and their applications have been studied extensively over the years. Previous work has shown that gamification can improve the learning effect of software tutorials [18,43], and that there are motivational advantages of using collectibles [1,24]. Previous work has suggested that players that like to engage with collectibles also enjoy other “incentives” including badges and achievements [73].

Work by Anderson et al. [1] shows differential responses to secondary objectives: 1) non-essential objectives that require too much effort to be obtained can have a negative effect on player retention, while 2) objectives that are along the path of the main objective can have a positive effect on player behaviour and level completion. Sarkar et al. [65] showed that participants in an online experiment kept on following their preferred path when no coins were present. This work suggests that coin placement strategies affect the length of time played and number of levels completed when coins are present.

Gaston and Cooper [24] showed that providing performance feedback in the form of three performance levels (three stars) reduced the number of moves, increased time taken per move, and increased the number replays per level, in the early and instructional levels of the human

computation game *Foldit* [15]. The shape of secondary objectives (e.g., coin, band-aid, sugar cane) did not change player behaviour, but secondary objectives that provided additional advantages (e.g., a large boost or a better chance to survive a fight) affected player behaviour [25]. Large-scale research has also shown that coins can potentially be perceived as distracting, resulting in decreased play time and retention [2].

In commercial games, collectibles are used in a multitude of ways to engage players [74]. The most common way is to directly reward players with extra lives, in-game currency, or information about the game, examples of this include gathering life orbs that increase the players maximum health in *Ori & the Blind Forest* (Moon Studios, 2015) or scanning data points to gather information about the game world in *Horizon Zero Dawn* (Guerrilla Games, 2017). In games that are more focused on exploration, rewards can also be used to show players which areas are still unexplored or to guide them in a certain direction. This can be done with markers on a player's mini-map [78] or by simply placing coins over a gap to show players' that they can jump over the gap. Lastly, designers often use collectibles to add replay value or extra challenge to their games. In the game *Celeste* (Matt Makes Games, 2018), most levels feature collectibles in the form of strawberries. No advantage can be gained from gathering the strawberries, but they allow for skill expression within the game and provide the player with an extra objective.

In summary, tutorials play an important role in software learning and previous work has demonstrated the advantages of game-based design strategies [18,43], gamification [12,68], and the application of motivational theories in game design [6,53,55]. Researchers and game designers have identified collectibles as valuable strategies to increase engagement and alter player behaviour [1,24]. However, to understand the value of collectibles for the design of game-based tutorials we require integrated studies that provide comparative insights into the effect of collectibles on experience, behaviour, and on potential transfer effects of collectible supported tutorials to applied tasks. Our work provides insights into these topics by comparing the effect of coins in a custom-built, game-based tutorial and investigates potential transfer effects to an out-of-game challenge.

3 Methods

Pathfinder is a custom built, 2D platform-based gamified software tutorial. We used Pathfinder to study the effects of collectibles on motivation and behaviour in two different conditions (i.e., with or without collectibles) and tested learning effects in a complex out-of-game challenge (Mandala challenge; see Figure 3). We used validated scales to measure play experience and cognitive demand, and we measured in-game behaviour to assess engagement.

3.1 Gaming System: Pathfinder

Pathfinder uses graphic editing tools as core game mechanics and was implemented in Unity 3D (Unity, 2020). The editor environment was simulated by creating an interactive screenshot of the Adobe Photoshop user interface (See Fig. 1). Just as in Photoshop, players of Pathfinder can select tools, move tiles, draw lines, and change the position of tiles on the z-axis, i.e., change a tile's layer. To implement Pathfinder, we used Unity and replicated aspects of Photoshop's editing interface.

While implementing games using software extensions has been done in GamiCAD [43] and Jigsaw [18], we see several advantages of leveraging a game engine for creating gamified tutorials: 1) we have access to purpose-built game development tools, difficult to implement features (such as physics), and game assets from the Unity assets store; 2) we can define our own logging events, e.g., we can log when a player clicks a button, moves the mouse, and their overall performance; and 3) we can export our system as a web-based HTML 5 game, which provides advantages for recruitment, especially when labs are inaccessible as during a health pandemic [19].

Fig. 1. Two panels: The left panel displays a solved level in Pathfinder including coins. The right panel displays a solved level of Pathfinder without coins.

3.1.1 Gameplay. In Pathfinder, players need to create a path to help a bird reach safety on a target platform. Throughout 12 different levels, players need to use the ‘move tool’, the ‘brush tool’, and ‘layer selection’ to arrange tiles and help the bird to safely reach the target platform without getting stuck or missing a platform and falling. We used the Unity physics engine to simulate gravity and collision. Pathfinder is played using a mouse, and interface elements are activated using the left mouse button or the corresponding keyboard shortcuts (“m” = move; “b” = brush). Each level presents the bird on a starting platform and the target platform. In addition, the player has access to a variety of tiles to create a path from the starting platform to the target platform. With the move tool, players can move tiles on the x- and y-axis. With the brush tool, players can draw lines on the pre-populated tiles, but not on the canvas. By changing the order of layers, participants can arrange tiles on the z-axis, changing how paths overlap. Participants first arrange all tiles, draw lines and then start the level by pressing the “start” button. Using the graphic editor tools, we gave players the freedom to create their own path, allowing for a large number of solutions per level. Players can retry a level in case the bird gets stuck or misses the target platform by pressing the “retry” button. Pressing “reset” puts the tiles back into their original position and removes drawn lines. Finally, players can skip a level entirely after three minutes by pressing the “skip” button.

3.1.2 Progression. New tools are gradually introduced while players progress through the game. We separated the game into three main stages corresponding to the move tool, the brush tool, and the layer system. Each stage starts with an initial instruction on how to use a new tool. We introduced tools using simple illustrations that show the process of selecting the right tool and how to use each tool (Fig. 2). Later levels require players to combine knowledge gained in previous levels in order to progress; e.g., in a later level, players need to draw barriers or platforms on the tiles, and then change the layer of the tiles to complete the level.

Fig. 2. Process illustration of selecting and using the brush tool showing three panels. The first indicating to press the brush button, the second to select the layer one wants to draw on, and the third to draw by dragging the mouse.

3.1.3 Conditions. We created two versions of Pathfinder: one version providing coins as collectibles, and another version without collectibles. To complete a level in the coin condition, participants needed to collect at least one coin, which assured that players engaged with the coin mechanic. To ensure that coins could be easily engaged with in the coin condition, we placed one coin near the start platform—allowing participants to collect the coin without much effort. The other coins required deliberate effort to be collected and were placed with the intention to encourage exploration, and—by further engaging with the task—additional practice of newly learned editing skills. Collected coins were indicated at the top of the user interface. It should be noted that players could “skip” to the next level, regardless of the condition, and could therefore always progress and complete the experiment, even if for some reason they failed to capture the “easy” coin.

Previous work [1] has shown that players collect collectibles close to the path they take to complete the main goal of a level, but that difficult to reach collectibles are often neglected. By assuring that one easy coin is collected, we assure that players are aware of the coin concept, and found affordances to form a meaningful connection with the collection mechanism and all other interface elements focused on coin collection. Due to the position of the other coins, we interpret coin collection beyond the easy coin as volitional and deliberately invested effort.

3.2 Out-of-Game Mandala Challenge

To investigate if playing through Pathfinder increases performance in an out-of-game challenge, we created a challenge that requires participants to apply skills they learned when playing Pathfinder to complete a complex geometric structure—the out-of-game Mandala challenge (see Figure 3) Mandalas are geometric figures that are well suited for our study, because they are visually complex, allow for line drawings, and are emotionally neutral.

In the Mandala challenge, players are asked to assemble a fragmented Mandala to match a reference image. The Mandala challenge was implemented using the same technical setup as used in Pathfinder and gave participants access to the brush, move, and layer tool. The Mandala is fragmented into four tiles and lines that are moveable and need to be placed over a reference image.

To complete the Mandala challenge, participants need to rearrange layers and move tiles to remove red areas using the move tool and arranging layers. They further need to position the lines across tiles by using the move tool and arranging layers, and draw missing lines using the brush tool. Participants could submit the final challenge at any point in time, and we used their logged in-challenge interactions and their level of success as measures of volitional engagement [6].

Fig. 3. Figure of the Mandala challenge showing the Mandala fragmented in four pieces on the left and the target on the right.

3.3 Measurements

We used standardized questionnaires (i.e., NASA TLX, IMI, PENS) to measure play experience and instrumented the game to measure in-game performance. All questionnaires have been used before to measure play experience [8,26,69].

3.3.1 Play Experience. Need satisfaction was assessed using the Player Experience of Need Satisfaction scale (PENS, [63]). Based on Self-Determination Theory, PENS assesses competence—the experience of mastery over a challenge; autonomy—the experience of engaging in a task under our own volition; and relatedness—the experience of feeling connected to others. In addition to the foundational three constructs, PENS also assesses two game-specific constructs: immersion/presence—the experience of being immersed in a game; and, intuitive control—the intuitiveness of the game controls. PENS asks participants' agreement ranging from "strongly disagree" (1) to "strongly agree" (7). PENS has been widely used before to measure and model player experience [8,34].

To measure motivation we used the Intrinsic Motivation Inventory (IMI, [51]). The IMI assesses participants' enjoyment, effort, competence, and tension. Participants are asked to state their agreement to 18 statements on a 7-point Likert-scale [51]. The IMI has been commonly used in research on player motivation [6,57,59].

To measure perceived workload, we used the NASA Task Load Index (NASA TLX, [32]). The NASA-TLX assesses mental demand, physical demand, temporal demand, performance, effort, and frustration. To indicate task demand, participants indicated their experience using a 21-point slider for their task experience from very low (1) to very high (21). 21 points was the originally published scale to measure increments of low, medium, and high levels of a given dimension [32].

3.3.2 Performance: Pathfinder. Based on previous work [1,6,24], we assume that increased engagement would result in increased time spent in the coin condition, as well as an increase in interactions with the tools, and players choosing to pursue coins. We assumed that these differences would lead to improved learning outcomes. To measure a participant's in-game performance, we looked at 7 behavioural metrics:

1. *Completion time*—the total time it took for a participant to complete each level.

2. *Frequency of moves*—the number of actions with a tool, i.e., moving a tile, drawing line with the brush, or dragging a layer to change its position.
3. *Frequency of layer selections*—the number of times a layer was selected to either move the layer or draw on it.
4. *Frequency of tool selection*—number of times a new tool was selected, e.g., selecting the move tool (+1), then selecting the brush tool (+1).
5. *Frequency of retries*—the number of times a participant retried a level.
6. *Frequency of skips*—the number of times a player skipped a level after three minutes of play time.
7. *Collected coins (coin condition only)*—the number of collected coins per level, minimum = 1, maximum = 3.

3.3.3. *Performance: Out-of-Game Mandala Challenge.* To assess the performance in the final challenge, and to understand differences in skill with the newly learned tools, we measured six player metrics:

1. *Completion time*—the total time to complete the challenge.
2. *Frequency of moves*—the number of actions with a tool.
3. *Frequency of layer selections*—the number of times a layer was selected to either move the layer or draw on it.
4. *Frequency of tool selection*—the number of times a new tool was selected.
5. *Frequency of retries*—the number of times a participant retried a level.
6. *Score*—the accuracy of positioning the tiles and ordering the layers, ranging from 0 to 100% accuracy.

For this study we have chosen to separate the game from the final out-of-game Mandala challenge for two reasons: 1) to separate play experience of Pathfinder and performance in the Mandala challenge; and 2) to investigate transfer effects to a task other than playing Pathfinder, since a gamified tutorial aims to allow skills to transfer to new non-game scenarios.

3.4 Participants and Deployment

We recruited a total of 60 participants from the crowdsourcing platform Amazon Mechanical Turk (MTurk). MTurk acts as a broker between workers who want to engage in digital labour and requesters who propose and compensate for Human Intelligence Tasks (HITs). We used the Bride-of-the-Frankensystem (BOFS) [33] to deploy the study online. Participants received USD\$7.50 for around 45 minutes of work, which is above the national minimum wage in the US. We paid above minimum wage to assure that participants received minimally fair compensation for their time and would be less likely to experience a time pressure while engaging in the study. MTurk has previously been used before for studying the effects of alternative game designs [7,27].

Before analysis, we removed 4 participants due to evidence of low effort, based on the following pre-established criteria: they skipped every level without engaging with the tasks (2 participants), showed zero variance between survey items in at least two questionnaires (2 participants), or showed ratings $\pm 3SD$ in more than 2 questions (0 participants). Removing participants based on quality criteria is a standard approach; we added skipping behaviour as an exclusion criteria to the metrics suggested by [49].

The remaining 56 participants ranged in age from 22 to 66 years ($M = 40.56$, $SD = 11.801$) with 61.4% identified as male and 38.6% as female. Participants were also given the option to identify as non-binary, not disclose their gender, or describe their gender themselves. 57% had post-secondary or higher education, 89.5% were employed for wages or self-employed. 33.3% played platformer games and the majority identified as a gamer ($M = 6.86$, $Md = 8$, $SD = 2.869$). Ethical approval was obtained from the ethical review board at Eindhoven University of Technology, and participants were asked to provide their consent. The MTurk task was only made available to workers in the US and workers younger than 18 were excluded. We only recruited workers with an approval rate above 90%.

Participants were asked about their previous game experiences as this has shown to be able to influence their performance in other game related tasks [56]. Participants replied on a 5-point scale from "not at all" to "very much". Kolmogorov-Smirnov tests for each question shows no differences between both conditions, suggesting that prior experience with Photoshop, interest to learn Photoshop, and game experience are equally distributed between groups.

3.4.1 Procedure. After providing consent, participants filled in demographics questionnaires and questionnaires inquiring about their previous experience with games and graphics editing. Participants were then randomly assigned to either the coin or no-coin condition and received a brief introduction to the game. Each participant played 3 "stages" (a group of levels) with 4 levels each. Each stage introduced one of the three editing tools; i.e., move, brush, layer. After each stage, participants responded to single item questions regarding their effort and enjoyment. After finishing all 12 levels participants filled in questionnaires regarding the mental task load (NASA TLX), and play experience and motivation (PENS, IMI). Finally, participants were asked to complete the Mandala challenge. In total, the study took around 45 minutes.

3.5 Analysis

Collected data was analyzed using SPSS 25 (IBM, 2019). Individual IMI and PENS items were averaged across construct scale. For experience data, we removed data points ($\pm 3SD$). Play behaviour data was collected per trial for each level, including retries, data points ($\pm 3SD$) were removed before analyzing the data. For frequency data, we aggregated the data and calculated the sum of all trials; e.g., the sum of all moves in one level. This exclusion criteria resulted in just 1 or 2 users being removed for each dependent variable.

To address family-wise errors, we controlled our results using a MANOVA include all variables reported in Tab. 1. In SPSS, MANOVAs correct family-wise error for multiple ANOVAs. However, participants with missing data—due to removal based on outlier data—would be removed from the analysis. The MANOVA results confirm individual ANOVA results. Hence, we report ANOVA results per measure.

Kolmogorov-Smirnov tests show that the majority of our measures are not normally distributed ($p < 0.5$). However, with a sample size larger than 20, ANOVAs are robust against non-normality [9]. Considering the level of measurement for Likert-scales (interval scale) and behavioural data (ratio scale), the tradition of the broader HCI community [46], and the central limit theorem we decided to perform parametric tests. Compared to non-parametric tests, parametric tests are more conservative and less likely result in false positives (Type I Error).

To test the effect of collectibles on play experience and play performance we performed univariate ANOVAs and regression analysis to investigate the relationship between the performance playing Pathfinder and engaging in the Mandala challenge. Degrees of freedom vary due to the removal of outlier trails (See Tab. 1). In Fig. 4 and Fig. 5, we present Z-Scores to account for the vastly different maxima between measures; mean values and standard deviations can be found in Tab. 1. We report partial η^2 as the effect size measure.

4 Results

We present analysis of participant behaviour and self-reported demand, experience, and motivation while playing Pathfinder. Further, we explore behaviour and experience in Pathfinder as predictors of performance in the follow-up Mandala challenge.

Fig. 4. Four panels of bar graphs showing Z-Scores on the y-axis and behaviour in Pathfinder (top left), and scales from NASA TLX (top right), PENS (bottom left), and IMI (bottom right) as dependent measures on the x-axis. Asterisk indicates significant results; $p < .05$. Error bars = 95% CI.

4.1 Pathfinder: Behaviour

We find that for game behaviour during all levels combined, individuals in the coin condition spent significantly more time playing ($F_{1,55} = 31.954, p < .001, \eta_p^2 = .376$), changed tools more often ($F_{1,55} = 6.926, p = .011, \eta_p^2 = .116$), changed layers more often ($F_{1,55} = 16.289, p < .001, \eta_p^2 = .235$), and made more moves ($F_{1,55} = 17.645, p < .001, \eta_p^2 = .250$). See Tab. 1 for descriptive details, and Fig. 4 for a visualization of our findings.

In total, participants in the coin condition spent an average of 654 seconds more in the game, performed 9 tool changes more, selected layers 87 times more often, and used 50 more moves. Our results suggest that the manipulation is successful and increases the time spent in each level and the overall player activity. Using a one-sample t-test, we investigated coin collecting behaviour in the coin condition. We tested the achieved score against the required 12 coins and found that the majority of participants collected around 20 coins ($t_{1,25}, p < .001, M = 20.73, Md = 20$), suggesting that the majority engaged in play and did not just try to finish the task as efficiently as possible. Our results suggest that using collectibles affects behaviour in the expected way—people engage more, and we assume that more time spent playing Pathfinder will also lead to better learning and hence better performance in the Mandala challenge

4.2 Pathfinder: Experience

Overall, the coin condition was experienced as more mentally demanding ($F_{1,54} = 3.99, p = .051, \eta_p^2 = .069$), participants felt more rushed ($F_{1,54} = 5.428, p = .024, \eta_p^2 = .091$), and that they worked harder ($F_{1,54} = 3.955, p = .052, \eta_p^2 = .068$). They also felt less successful ($F_{1,54} = 3.918, p = .053, \eta_p^2 = .068$) and less competent measured by IMI ($F_{1,54} = 4.960, p = .030, \eta_p^2 = .084$) and PENS ($F_{1,54} = 4.96, p = .03, \eta_p^2 = .084$). These findings are interesting, because they show how small manipulations affect our

experience of mastering a task. See Tab. 1 for descriptive details; Fig. 5 for a visualization of our findings.

Investigating effort and enjoyment after every 4 levels, we find no main or interaction effects—the results during play are in line with the results for effort and enjoyment after playing Pathfinder.

Our results show that participants in both groups experience Pathfinder differently. The differences show primarily in the increased experience of demand and decreased the experience of competence. When creating challenges, game designers aim to create challenges in such a way that players can experience mastery over the challenge. Mastering a challenge should lead to positive experiences, but Pathfinder instilled experiences of demand and decreased competency, which requires discussion from a game design point of view. The effects of collectibles on experience and behaviour are promising, but the aim of instructional software is to show transfer effects outside of the games. We next look at results in the Mandala challenge.

	<i>N</i> (<i>n1/n2</i>)	Coins		No Coins		<i>F</i>	F- Test		
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>df</i>	<i>p</i>	η_p^2
Pathfinder									
*Time spent (s)	55 (26/29)	944.41	405.66	1598.30	452.32	31.954	1,53	< .001	.376
*Tool change	55 (26/29)	33.00	13.58	41.89	11.17	6.926	1,53	.011	.116
*Layer selection	55 (26/29)	121.24	79.59	208.08	79.75	16.289	1,53	< .001	.235
*Moves	55 (26/29)	115.48	47.22	165.89	41.07	17.645	1,53	< .001	.250
+TLX Mental Demand	56 (26/30)	13.17	4.98	15.62	4.05	3.990	1,54	0.051	.069
TLX Physical Demand	56 (26/30)	8.47	9.49	9.62	10.54	0.184	1,54	0.670	.003
*TLX Rushed	56 (26/30)	3.30	3.76	6.38	6.03	5.428	1,54	0.024	.091
*TLX Success	55 (26/29)	18.24	4.14	15.15	4.22	7.49	1,53	0.008	0.124
+TLX Effort	56 (26/30)	12.97	6.02	15.77	4.21	3.955	1,54	0.052	.068
TLX Stress	56 (26/30)	5.73	6.15	8.69	6.50	3.055	1,54	.086	.054
+PENS Competence	54 (26/28)	5.70	0.86	5.19	1.03	3.936	1,52	0.053	0.07
PENS Autonomy	55 (26/29)	5.56	1.14	5.42	1.05	0.222	1,53	0.639	0.004
PENS Relatedness	56 (26/30)	3.96	1.26	3.92	1.736	0.007	1,54	0.936	0
PENS Presence	56 (26/30)	3.80	1.49	3.84	1.60	0.012	1,54	0.913	0
PENS Intuitive Controls	55 (26/29)	5.10	1.23	5.23	1.15	0.16	1,53	0.694	0.003
IMI Enjoyment	56 (26/30)	5.13	1.47	4.83	1.37	0.6	1,54	0.442	0.011
*IMI Competence	56 (26/30)	5.37	1.41	4.55	1.31	4.96	1,54	0.03	0.084
IMI Importance	56 (26/30)	5.97	0.83	5.78	0.86	0.649	1,52	0.424	0.012
IMI Tension	56 (26/30)	3.14	1.72	3.46	1.73	0.479	1,54	0.492	0.009
Mandala challenge									
Time	54 (25/29)	216.14	128.89	202.88	116.77	0.155	1,52	0.696	0.003
*Tool changes	55 (26/29)	2.45	1.94	1.46	0.86	5.722	1,53	0.02	0.097
Layer selection	55 (26/29)	16.86	11.34	16.08	9.87	0.074	1,53	0.786	0.001
Moves	56 (26/30)	29.07	20.06	25.19	16.66	0.607	1,54	0.439	0.011
Score	56 (26/30)	53.67	26.78	59.73	34.53	0.547	1,54	0.463	0.01

* = p < .05; + = p > .05 & P < .06

* = $p < .05$; + = $p > .05$ & $P < .06$

Tab. 1. Mean values, standard deviations, and F-test statistics for Pathfinder and the Mandala challenge.

4.3 Out-of-Game Mandala Challenge

While our results demonstrate that collectibles modify behaviour in gameplay, we were interested in transfer effects to the Mandala challenge. We removed eleven participants (19.6%) who showed a lack of effort in the Mandala challenge, indicated by a score lower than or equal to 20% accuracy. We find no differences in time spent ($p = .928$). However, participants in the no-coin condition showed fewer tool changes ($F_{1,43} = 5.552, p = .023, \eta_p^2 = .114$). See Figure 5.

To investigate if the pursuit of coins had an effect on performance in the Mandala challenge, we investigate the relationship between collected coins and Mandala challenge score in the coin

condition only. Our results show that collected coins predict the final score of the challenge $F_{1,24} = 4.407, p = .046, R^2 = .155$). See Figure 5.

Our results suggest that collectibles positively affect performance in a realistic editing task. The relation between collected coins in Pathfinder and score in the Mandala challenge suggest transfer potential of increased engagement and learning through collectibles to transfer to more complex tasks.

Fig. 5. Left panel shows bar charts for behaviour and performance in the Mandala challenge. The right panel shows a scatter plot with regression line for the relationship between the Pathfinder score and the Mandala challenge score.

5 Discussion

We studied experience and behaviour in a gamified software tutorial and the effects on performance in a transfer challenge. Our results provide three insights:

1. Collectibles affect in-game behaviour of participants, evident in an increase in time spent, tool changes, layer changes, and moves made in the coin condition.
2. Collectibles affect the experience of participants, evident in an increase in experienced mental demands, invested effort, and a sense of feeling rushed and stressed, while feelings of less success in the coin condition.
3. We show that pursuing collectibles when playing Pathfinder alters behaviour in the out-of-game Mandala challenge.
4. We show that performance in Pathfinder is predictive of the performance in the out-of-game Mandala challenge.

Our results provide insights into behavioural motivation and experience of Pathfinder, and demonstrate that performance in Pathfinder changes behaviour in the Mandala challenge. We discuss several theoretical and practical perspectives to contextualize our results.

5.1 Explanation of Results

To understand our findings, we need to consider the coin manipulation, the specific properties of Pathfinder and the out-of-game Mandala challenge, and the context of our participants. The effect of collectibles on movement in a game has been shown before [1,2,24], but not in the context of a software tutorial where software tools are the primary game mechanics. By tasking players to use software tools to progress through a level, we assure that the tools are constantly used to achieve the goal of the level while collecting user data on tool use.

First, we show that collectibles affect user behaviour. Participants were randomly assigned to the coin and no-coin condition, suggesting that the changes in behaviour are due to the coin placement. We find that participants pursued more than the required coins, suggesting that they

played the game, accepted the performance goal of collecting coins, and behaved accordingly. The change in behaviour is reflected in the individual behavioural measures, and varied depending on the level's difficulty. In addition, to use the different tools efficiently, players also needed to decide to choose a path that would allow them to gather all coins, further adding to the time spent in a level.

Second, we show that collectibles affect experience. The results show that when coins are presented the experienced demand of the game increases. This is in line with research defining video games as cognitive, social, emotional, and physically demanding experiences [11]. Cognitive demand suggests that playing is demanding on attentional resources, which has been shown to predict increased enjoyment [11]. We show that adding coins increased self-reported mental demand and decreased feelings of success. Our results show there is no difference in enjoyment between conditions, but that scores were slightly elevated with means above 5 (center of the scale is 4) suggesting a positive experience in both conditions. The relationship between play complexity, experienced demand, and self-reported enjoyment requires further investigation to better understand the role of enjoyment. These results show how gameplay creates challenge that facilitates a range of experiences but does not necessarily affect enjoyment.

Previous work has, for example, investigated the role of meaning in games [75], negative emotions [10], or experiences of challenge [4]. Further, the NASA-TLX and competence results suggest that Pathfinder is more demanding in the coin condition, but the motivational indicators in our data show no sign of changes in motivation. We interpret these findings as the emergence of challenge that could foster experience of mastery over time. Our results are in line with previous findings that suggest that game design elements in a non-game context increase performance, but not necessarily intrinsic motivation [52]. From a learning perspective, we add to Mekler et al.'s [52] findings that within a learning context, behavioural engagement is sometimes sufficient.

Third, we show that playing Pathfinder alters behaviour in the out-of-game Mandala challenge. The data shows that tool changes are increased in the coin condition. All other performance measures show no differences in terms of player interest. The results raise questions about the value of increased behaviour engagement for training. Is it necessary to add additional challenge to train people using basic software functionalities such as drawing and movement? Engaging in the basic challenges might simply be enough to understand the tool itself, because people are familiar with moving objects in software (e.g., drag and drop in the file system), and likely have used a pen or brush tool before (e.g., when digitally signing a document). The layer tool is likely the most novel concept, but conceptually not too difficult. These findings would suggest that coin placement is not necessary and the challenge to safely transport the bird from start to finish platform is sufficient to train people. However, the coin condition requires participants to make use of the entire tool range and use tools fluently to gather all coins. We assume that throughout the game people learn to frequently change tools and methodologically adjust the tiles on the screen to achieve their goal, which transfers into their tool changing behaviour in the out-of-game challenge. While this might not bring a direct performance advantage, it illustrates that in-game behaviour transfers to out-of-game challenges, which could, for example, be used to train different problem-solving strategies, e.g., interleaving tool use or using tools sequentially.

Fourth, our results show that within the coin condition the amount of collected coins predict performance in the Mandala challenge. Considering the lack of performance difference between both conditions (coin vs. no-coin), we assume that this result is primarily driven by differences in task commitment and performance-willingness. People who try to do well in Pathfinder also try to do well in the Mandala challenge. We interpret these findings as being influenced by the crowdsourced nature of the study. Previous work has shown that workers enjoy playing games online and produce good quality work [39], and that those whose primary interest is not financial perform better in complex tasks [47,64]. Workers are a heterogeneous groups of people, but previous work has shown that many are not primarily interested in earning money, but enjoy the different tasks they are challenged with [49].

5.2 Collectibles in Software Tutorials

As all approaches, using collectibles in software tutorials has its own strengths and weaknesses, and presents opportunities and threats. A key strength for the use of collectibles in tutorials is their versatility; e.g., they can be used to indicate different degrees of completion like three-star systems, guide players on specific paths, or provide achievements that can be shared with others. Further, the positive feedback collectibles provide are particularly attractive to players who strive for completion. However, collectibles add a different system to a game, which must be balanced with the main goals of the game; e.g., adding an extra life through coin collection to Super Mario Bros—too many coins may make some parts of the game too easy. If the correct balance between challenge, frequency, and side effects is not struck it may affect why and how people engage with collectibles. Collectibles are also another element that needs to be designed and considered in all other aspects of the game; e.g., the sound of collecting a coin needs to be audible, but not distracting. Opportunities are in using collectibles to manipulate behaviour, in creating tutorials that are more enjoyable, in creating better learning environments, and in helping people to reach their goal to become proficient faster. However, previous research has shown that collectibles can be distracting [2], simply ignored [65], or can diminish engagement and retention [2].

In summary, the value of collectibles in tutorials is ambiguous and depends on individual differences such as play preferences and the specific design case. However, in this work we provide evidence that when carefully integrated, we find that through collectibles learners experience complex demands that could be turned into more exciting experiences within the context of a tutorial and show improved out-of-game skill transfer. Additionally, we find no differences in performance, suggesting that in our research we did not observe the harmful effects of collectibles mentioned in previous literature [1,2].

5.3 Goal-Setting

Goal-setting theory has been previously applied to gamification principles [74]. Our findings suggest that a small number of achievable goals can alter behaviour. For example, many people collected at least two coins per level and one player even spent the time to collect all 3 coins, in all 12 levels. Goal-setting theory suggests that choice, effort, persistence, and cognition are affected by goals [42]. When people actively pursue a goal, e.g., collecting all coins, goal-setting theory suggests that when goal-relevant and goal-irrelevant choices are presented, people focus attention towards fulfilling the goal and increase goal-relevant behaviour; e.g., positioning tiles to collect all coins. Goal-setting theory also suggests that the effort to achieve a goal would increase, which in our study is evident in increased task load but not in self-reported effort: participants found Pathfinder more challenging with coins, but did not express that they invested more effort compared to the no-coin group. Goal-setting theory suggests that when a goal is set persistence increases (i.e., people tend to stick with and strive for their goals). In our study, some of our participants showed clear signs of pursuing the goal of collecting coins persistently; e.g., by retrying a level 10 times until all coins were collected. We also find that people adapted their behaviour and engaged more as shown by an increase in the number of moves.

In summary, our results suggest that people engaged with the collectibles in the intended way and accepted them as additional goals that they pursued, altering their behaviour. Our findings are relevant for experience design and gamification because they provide additional insight into how collectibles modify behaviour and experience.

5.4 Advantages of Game Engines for Integrated Game-Tutorials

By implementing an interface that is commonly used in photo editing in the Unity game engine, and with the goal to create game-based tutorials using photo editing tools as mechanics, we created a novel design context. Previous work has either implemented game design elements

within existing software [43], created external learning experiences [12], or layer tutorials on top of the existing software [18]. While these approaches enable designers to create games, from a game designer's perspective they are limited due to the features of the technology they are implement in. Creating new levels, adding physics, shaders, particles, or 3D effects is rather challenging within software built for photo editing. The emphasis in previous projects on the integration of tutorials within the target software being taught, has benefits for the available functionality, but also requires access to the software code itself or to be able to otherwise modify the software through plugins, API calls, or extensions. The limitation of Pathfinder is in the available features, which are either part of the game engine (e.g., drawing) or need to be implemented (e.g., layer tool).

Once the target software interface context has been developed, level creation is rather straight forward and allows researchers to focus on level design. In the study setup we presented, the in-game measurement of behaviour is also crucial. Game engines enable researchers to integrate logging hooks into the game code, which facilitates behaviour logging at different levels of granularity—something that within commercial software would be more difficult to setup, unless one has accesses to the code of the software; e.g., the open-source graphics editor GIMP [70].

Game engines also allow to export games to be played within web browsers using the WebGL standard, which allows for easy distribution online, and enables the use of crowdsourcing, marketing panels, or online recruitment. Researching gameplay at home is advantageous when studying motivation in-game play, because people can use equipment they are familiar with and have a lower barrier to quit or show unmotivated behaviour in game-play.

We illustrate several advantages of studying game design elements in the context of software tutorial by recreating existing software in a game engine over implementing the same game within the target software.

5.5 Design Considerations

Our work provides several guidelines for designers of learning tutorials and for applications of gamification. Our research suggests that when designing tutorials, and especially when the goal is to increase behavioural engagement, non-essential rewards can help to increase activity. While we are presenting work on collectibles in a gamified environment, our insights translate to traditional tutorials as well. A Photoshop tutorial focused on teaching skills about repair and retouching, could for example propose a challenge to embellish a set of predefined images and ask users to compare their result to benchmark solutions—without adding any visible game-elements.

While game designers have extensive experience with using collectibles, our results provide a foundation to discuss the effects of collectibles on user behaviour and experience. Our results can also, for example, be used to discuss the effect of collectibles on success. Combined with work on the Zeigarnik effect [81], which suggests the innate need to complete a task once started, our results provide insight into the condition of activities of completion and the tradeoff of progression vs. completion in gameplay.

We provide further evidence that the gamification of learning environments has potential and provides insights for the use of collectibles as game-elements. The most prominent category of non-essential collectibles in gamification are badges [30], which are usually provided by serious achievements and are not immediately visible within the game. Our work contributes to our understanding of smaller, more direct, and less spaced-out achievements that allow for a finer control of user behaviour during play—similar to coins in *Super Mario Bros.* to facilitate exploration.

Investigating tools such as move, brush, and layer selection, we focused on the basics of graphic editing and did not investigate more complex concepts like retouching and repairing. However, we could envision games played using the stamp tool to fix damaged bridges, using colour adjustments to alter the environment (similar to the use of shadows in *Black & White: Bushido*; Good Catch Games, 2015), or history tools such as Photoshop's history brush that could be used to

simulate time mechanics as used in Braid (Number None, 2008). Using collectibles as a concept is very powerful, because collecting is versatile and integrates well with a number of game concepts and mechanics.

5.6 Limitations and Future Work

Our work has several limitations that provide opportunities for future investigations. We only present a single study with a fairly small sample. Due to our small sample size and the lack of power to detect smaller effects, we might have missed relevant insights; e.g., all measures of behaviour in the Mandala challenge (see Figure 3) show descriptively a decrease of time spent and interface interactions, except for score. A study with a larger sample, more statistical power, and further replications of our insights would strengthen the findings we present.

Pathfinder uses basic editing tools; i.e., move, brush, layer—leveraging more complex tools would provide more nuanced insights into the effects of secondary rewards, transfer effects, and open up opportunities for more complex gameplay. The Mandala challenge is a single challenge that requires the use of all taught skills, but it is not a common challenge that users of photo editing software must approach regularly. Different challenges such as common tutorials or standard editing tasks like masking an image would add to our understanding of transfer effects.

Our approach was effective in that we were able to provide users with a tool that looked similar to the real Photoshop software. That is unless a user tried to interact with any part of our interface that was unimplemented. Our approach is akin to vertical prototyping [5] in that we offer only a very thin slice of the full software. This made our project tractable, as otherwise fully implementing many aspects of an extremely complex and feature-rich piece of software would be a huge undertaking. Our approach may be limiting to the adoption of more expert behaviour. For example, if there are other behaviours and paths that users might adopt to achieve similar outcomes, those would not be possible. While we do not believe this is the case for the simple tools we introduced, this might be a limitation of the approach in some contexts. In these cases, it would prevent novices from exploration and experimentation; however, we believe that this a reasonable tradeoff. We expect that users would have access to full versions of the software that they are learning and could be told the difference between the training environment and the actual software. The real software could still be used for exploration and experimentation. Future work might consider the tradeoffs of presenting a limited set of functionalities that are introduced through game-based tutorials, and investigate how exploration and experimentation of non-taught functionality might be supported, and whether providing this support for exploration in the tutorial software might be worth the effort.

To further our understanding of collectibles as a tool for serious game design, in the future we will investigate the effects of embellishment on behaviour. We also aim to investigate more complex tools such as the history brush. One direction we are evaluating is the use of gamification instructions in productive environments; e.g., in teaching university students photo editing basics. We see Pathfinder as a game design and research tool that allows us to investigate game design in software learning environments and hence investigate behavioural, experiential, and transfer effects of engaging with a game.

6 Conclusion

Designers of instructional software commonly evoke straightforward pathways through learning, but potentially miss out on the promise of non-essential goals to stimulate engagement and productive learning behaviour. After we studied the effects of non-essential goals in the gamified software tutorial Pathfinder, we can confirm direct positive effects of using collectibles on engagement with the game and find that the engagement also transfers to out-of-game challenges. While the training results provide further support for the potential of gamified tutorials, our results provide “food for thought” for designers and user researcher involved in the creation of

gamified learning environments by showing that increased challenge and decreased competence can potentially lead to motivationally relevant experiences of mastery. When we learn, we need to explore, retry, succeed, and fail—what better environment to grow our skills than video games?

REFERENCES

- [1] Erik Andersen, Yun En Liu, Richard Snider, Roy Szeto, Seth Cooper, and Zoran Popović. 2011. On the harmfulness of secondary game objectives. *Proceedings of the 6th International Conference on the Foundations of Digital Games, FDG 2011*: 30–37. <https://doi.org/10.1145/2159365.2159370>
- [2] Erik Andersen, Yun En Liu, Richard Snider, Roy Szeto, and Zoran Popović. 2011. Placing a value on aesthetics in online casual games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1275–1278. <https://doi.org/10.1145/1978942.1979131>
- [3] Erik Andersen, Eleanor O'Rourke, Yun En Liu, Richard Snider, Jeff Lowdermilk, David Truong, Seth Cooper, and Zoran Popović. 2012. The impact of tutorials on games of varying complexity. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Pajitnov 1984: 59–68. <https://doi.org/10.1145/2207676.2207687>
- [4] Craig G Anderson, Kathryn Campbell, and Constance Steinkuehler. 2019. Building Persistence Through Failure The Role of Challenge in Video Games. <https://doi.org/10.1145/3337722.3337741>
- [5] Michel Beaudouin-Lafon and Wendy E. Mackay. 2007. Prototyping Tools and Techniques. In *The Human-Computer Interaction Handbook*. CRC Press, 1043–1066. <https://doi.org/10.1201/9781410615862-66>
- [6] Max V. Birk, Cheralyn Atkins, Jason T. Bowey, and Regan L. Mandryk. 2016. Fostering intrinsic motivation through avatar identification in digital games. *Conference on Human Factors in Computing Systems - Proceedings*: 2982–2995. <https://doi.org/10.1145/2858036.2858062>
- [7] Max V. Birk and Regan L. Mandryk. 2018. Combating attrition in digital self-improvement programs using avatar customization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3173574.3174234>
- [8] Max V. Birk, Regan L. Mandryk, Matthew K. Miller, and Kathrin M. Gerling. 2015. How self-esteem shapes our interactions with play technologies. In *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 35–46. <https://doi.org/10.1145/2793107.2793111>
- [9] María J. Blanca, Rafael Alarcón, Jaume Arnau, Roser Bono, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema* 29, 4: 552–557. <https://doi.org/10.7334/psicothema2016.383>
- [10] Julia Ayumi Bopp, Elisa D Mekler, and Klaus Opwis. 2016. Negative emotion, positive experience? Emotionally moving moments in digital games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2996–3006. <https://doi.org/10.1145/2858036.2858227>
- [11] Nicholas David Bowman, Joseph Wasserman, and Jaime Banks. 2018. *Development of the Video Game Demand Scale*. Routledge.
- [12] Achilleas L.D. Buisman and Marko C.J.D. Van Eekelen. 2014. Gamification in educational software development. In *Proceedings - CSERC 2014: Computer Science Education Research Conference*, 9–20. <https://doi.org/10.1145/2691352.2691353>
- [13] Josh Bycer. 2019. The Crafting of Collectibles in Videogames. *Medium*. Retrieved September 12, 2020 from <https://medium.com/@GWBycer/the-crafting-of-collectibles-in-videogames-1ca37f39ccff>
- [14] Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys* 47. <https://doi.org/10.1145/2659796>
- [15] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and Foldit Players. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466, 7307: 756–760. <https://doi.org/10.1038/nature09304>
- [16] Edward L. Deci, Richard M. Ryan, Robert J. Vallerand, and Luc G. Pelletier. 1991. Motivation and Education: The Self-Determination Perspective. *Educational Psychologist* 26, 3–4: 325–346. <https://doi.org/10.1080/00461520.1991.9653137>
- [17] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: Defining “gamification.” In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, 9–15. <https://doi.org/10.1145/2181037.2181040>
- [18] Tao Dong, Mira Dontcheva, Diana Joseph, Karrie Karahalios, Mark W Newman, and Mark S Ackerman. 2012. Discovery-based games for learning software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2083–2086. <https://doi.org/10.1145/2207676.2208358>
- [19] N Donthu, A Gustafsson - Journal of business research, and undefined 2020. Effects of COVID-19 on business and research. *ncbi.nlm.nih.gov*. Retrieved September 16, 2020 from <https://www.ncbi.nlm.nih.gov/pmc/PMC7280091>
- [20] Dorknado. 2017. Collectables In Video Games: How And When To Use Them. *dorknado.com*. Retrieved September 12, 2020 from <http://dorknado.com/collectables-video-games-use/>
- [21] Jennifer Fernquist, Tovi Grossman, and George Fitzmaurice. 2011. Sketch-sketch revolution: An engaging tutorial

- system for guided sketching and application learning. In *UIST'11 - Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 373–382. <https://doi.org/10.1145/2047196.2047245>
- [22] Michael Filsecker and Michael Kerres. 2014. Engagement as a Volitional Construct: A Framework for Evidence-Based Research on Educational Games. *Simulation & Gaming* 45, 5: 450–470. <https://doi.org/10.1177/1046878114553569>
 - [23] Rosemary Garriss, Robert Ahlers, and James E. Driskell. 2002. Games, motivation, and learning: A research and practice model. *Simulation and Gaming* 33, 4: 441–467. <https://doi.org/10.1177/1046878102238607>
 - [24] Jacqueline Gaston and Seth Cooper. 2017. To three or not to three: Improving human computation game onboarding with a three-star system. *Conference on Human Factors in Computing Systems - Proceedings 2017-May*: 5034–5039. <https://doi.org/10.1145/3025453.3025997>
 - [25] Alison Gazzard. 2012. Game Studies - Unlocking the Gameworld: The Rewards of Space and Time in Videogames. *Game Studies* 11, 1: 9–13. Retrieved September 12, 2020 from http://gamestudies.org/1101/articles/gazzard_alison
 - [26] Kathrin M. Gerling, Regan L. Mandryk, and Michael R. Kalyn. 2013. Wheelchair-based game design for older adults. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2013*. <https://doi.org/10.1145/2513383.2513436>
 - [27] Tobias Greitemeyer. 2018. The spreading impact of playing violent video games on aggression. *Computers in Human Behavior* 80: 216–219. <https://doi.org/10.1016/j.chb.2017.11.022>
 - [28] Tovi Grossman, George Fitzmaurice, and Ramtin Attar. 2009. A survey of software learnability: Metrics, methodologies and guidelines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 649–658. <https://doi.org/10.1145/1518701.1518803>
 - [29] Philip J Guo, Juho Kim, and Rob Rubin. 2014. How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. *dl.acm.org/Paperpile*: 41–50. <https://doi.org/10.1145/2556325.2566239>
 - [30] Juho Hamari. 2017. Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior* 71: 469–478. <https://doi.org/10.1016/j.chb.2015.03.036>
 - [31] Kyle J. Harms, Dennis Cosgrove, Shannon Gray, and Caitlin Kelleher. 2013. Automatically generating tutorials to enable middle school children to learn programming independently. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 11–19. <https://doi.org/10.1145/2485760.2485764>
 - [32] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52, C: 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
 - [33] Colby Johanson. 2019. Bride of Frankensystem. <https://doi.org/http://doi.org/10.5281/zenodo.3544693>
 - [34] Daniel Johnson and John Gardner. 2010. Personality, motivation and video games. In *ACM International Conference Proceeding Series*, 276–279. <https://doi.org/10.1145/1952222.1952281>
 - [35] Jussi P.P. Jokinen, Sayan Sarcar, Antti Oulasvirta, Chaklam Silpasuwanchai, Zhenxin Wang, and Xiangshi Ren. 2017. Modelling learning of new keyboard layouts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 4203–4215. <https://doi.org/10.1145/3025453.3025580>
 - [36] Bomni Kang, Sophia Tan, Sophia H Tan, and E Craig Wall Sr. 2014. Interactive Games: Intrinsic and Extrinsic Motivation, Achievement, and Satisfaction. *Article in Journal of Management and Strategy* 5, 4. <https://doi.org/10.5430/jms.v5n4p110>
 - [37] Liesbeth Kester, Paul Kirschner, and Gemma Corbalan. 2007. Designing support to facilitate learning in powerful electronic learning environments. *Computers in Human Behavior* 23, 3: 1047–1054. <https://doi.org/10.1016/j.chb.2006.10.001>
 - [38] Juho Kim, Benjamin Malley, Joel Brandt, Mira Dontcheva, Diana Joseph, Krzysztof Z. Gajos, and Robert C. Miller. 2012. Photoshop with friends: A synchronous learning community for graphic design. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 271–272. <https://doi.org/10.1145/2141512.2141598>
 - [39] Markus Krause and René Kizilcec. 2015. To Play or not to Play: Interactions between Response Quality and Task Complexity in Games and Paid Crowdsourcing. *HComp'15 Proceedings of the AAAI Conference on Human Computation*: 102–109.
 - [40] Ben Lafreniere, Tovi Grossman, and George Fitzmaurice. 2013. Community enhanced tutorials: Improving tutorials with multiple demonstrations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1779–1788. <https://doi.org/10.1145/2470654.2466235>
 - [41] Richard N. Landers, Kristina N. Bauer, and Rachel C. Callan. 2017. Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior* 71: 508–515. <https://doi.org/10.1016/j.chb.2015.08.008>
 - [42] Gary P. Latham, Marie-Hélène Budworth, and Marie-Hélène Budworth. 2014. The Study of Work Motivation in the 20th Century. 383–412. <https://doi.org/10.4324/9781315820972-28>
 - [43] Wei Li, Tovi Grossman, and George Fitzmaurice. 2012. GamiCAD: A gamified tutorial system for first time AutoCAD users. *UIST'12 - Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*: 103–112.
 - [44] Edwin A Locke and Gary P Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* 57, 9: 705–717. <https://doi.org/10.1037/0003-066X.57.9.705>

- [45] Robert C. Lorenz, Tobias Gleich, Jürgen Gallinat, and Simone Kühn. 2015. Video game training and the reward system. *Frontiers in Human Neuroscience* 9, FEB. <https://doi.org/10.3389/fnhum.2015.00040>
- [46] Scott Mackenzie. 2013. *HCI Research: An Empirical Research Perspective*.
- [47] Andrew Mao, Ece Kamar, Yiling Chen, Eric Horvitz, Megan E. Schwamb, Chris J. Lintott, and Arfon M. Smith. 2013. Volunteering Versus Work for Pay: Incentives and Tradeoffs in Crowdsourcing. *First AAAI Conference on Human Computation and Crowdsourcing*: 94–102. Retrieved from <http://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7497>
- [48] Rob L. Martens, Judith Gulikerst, and Theo Bastiaens. 2004. The impact of intrinsic motivation on e-learning in authentic computer tasks. *Journal of Computer Assisted Learning* 20, 5: 368–376. <https://doi.org/10.1111/j.1365-2729.2004.00096.x>
- [49] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods* 44, 1: 1–23. <https://doi.org/10.3758/s13428-011-0124-6>
- [50] Richard E. Mayer. 2014. Incorporating motivation into multimedia learning. *Learning and Instruction* 29: 171–173. <https://doi.org/10.1016/j.learninstruc.2013.04.003>
- [51] Edward D. 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: 48–58. <https://doi.org/10.1080/02701367.1989.10607413>
- [52] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Do points, levels and leaderboards harm intrinsic motivation? An empirical analysis of common gamification elements. In *ACM International Conference Proceeding Series*, 66–73. <https://doi.org/10.1145/2583008.2583017>
- [53] Elisa D. Mekler, Florian Brühlmann, Alexandre N. Tuch, and Klaus Opwis. 2017. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior* 71: 525–534. <https://doi.org/10.1016/j.chb.2015.08.048>
- [54] Daniel C Moos and Roger Azevedo. 2009. Learning with computer-based learning environments: A literature review of computer self-efficacy. *Review of Educational Research* 79, 576–600. <https://doi.org/10.3102/0034654308326083>
- [55] Rita Orji, Regan L. Mandryk, Julita Vassileva, and Kathrin M. Gerling. 2013. Tailoring persuasive health games to gamer type. In *Conference on Human Factors in Computing Systems - Proceedings*, 2467–2476. <https://doi.org/10.1145/2470654.2481341>
- [56] Karin A. Orvis, Daniel B. Horn, and James Belanich. 2008. The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior* 24, 5: 2415–2433. <https://doi.org/10.1016/j.chb.2008.02.016>
- [57] Wei Peng, Jih-Hsuan Lin, Karin A Pfeiffer, and Brian Winn. 2012. Need Satisfaction Supportive Game Features as Motivational Determinants: An Experimental Study of a Self-Determination Theory Guided Exergame. *Media Psychology* 15, 2: 175–196. <https://doi.org/10.1080/15213269.2012.673850>
- [58] Tom Phillips. 2017. The reward for collecting all 900 Zelda: Breath of the Wild Korok seeds is a bit poo. *eurogamer.net*. Retrieved September 12, 2020 from <https://www.eurogamer.net/articles/2017-03-15-theres-a-ridiculous-reward-if-you-collect-zeldas-900-korok-seeds>
- [59] Andrew K. Przybylski, Netta Weinstein, Richard M. Ryan, and C. Scott Rigby. 2009. Having to versus wanting to play: Background and consequences of harmonious versus obsessive engagement in video games. *Cyberpsychology and Behavior* 12, 5: 485–492. <https://doi.org/10.1089/cpb.2009.0083>
- [60] Christian Roth, Ivar Vermeulen, Peter Vorderer, and Christoph Klimmt. 2012. Exploring replay value: Shifts and continuities in user experiences between first and second exposure to an interactive story. *Cyberpsychology, Behavior, and Social Networking* 15, 378–381. <https://doi.org/10.1089/cyber.2011.0437>
- [61] Rob van Roy and Bieke Zaman. 2018. Need-supporting gamification in education: An assessment of motivational effects over time. *Computers and Education* 127: 283–297. <https://doi.org/10.1016/j.compedu.2018.08.018>
- [62] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 1: 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- [63] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion* 30, 4: 347–363. <https://doi.org/10.1007/s11031-006-9051-8>
- [64] Anurag Sarkar and Seth Cooper. 2018. Comparing paid and volunteer recruitment in human computation games. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3235765.3235796>
- [65] Anurag Sarkar, Varun Sriram, Riddhi Padte, Jeffrey Cao, and Seth Cooper. 2018. Desire path-inspired procedural placement of coins in a platformer game. *CEUR Workshop Proceedings* 2282.
- [66] Elke A. Schoneveld, Monique Malmberg, Anna Lichtwarck-Aschoff, Geert P. Verheijen, Rutger C.M.E. Engels, and Isabela Granic. 2016. A neurofeedback video game (MindLight) to prevent anxiety in children: A randomized controlled trial. *Computers in Human Behavior* 63: 321–333. <https://doi.org/10.1016/j.chb.2016.05.005>
- [67] Katherine Schramm, Carl Gutwin, and Andy Cockburn. 2016. Supporting transitions to expertise in hidden toolbars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 4687–4698.

- <https://doi.org/10.1145/2858036.2858412>
- [68] Katie Seaborn and Deborah I. Fels. 2015. Gamification in theory and action: A survey. *International Journal of Human Computer Studies* 74: 14–31. <https://doi.org/10.1016/j.ijhcs.2014.09.006>
 - [69] David Sharek and Eric Wiebe. 2012. Embedding secondary tasks in video games to measure real-time cognitive load: An approach to developing adaptive video games. In *Proceedings of the Human Factors and Ergonomics Society*, 2147–2151. <https://doi.org/10.1177/1071181312561455>
 - [70] Michael Terry, Matthew Kay, Brad Van Vugt, Brandon Slack, and Terry Park. 2008. ingimp: Introducing instrumentation to an end-user open source application. In *Conference on Human Factors in Computing Systems - Proceedings*, 607–616. <https://doi.org/10.1145/1357054.1357152>
 - [71] Armando M. Toda, Ana C. T. Klock, Wilk Oliveira, Paula T. Palomino, Luiz Rodrigues, Lei Shi, Ig Bittencourt, Isabela Gasparini, Seiji Isotani, and Alexandra I. Cristea. 2019. Analysing gamification elements in educational environments using an existing Gamification taxonomy. *Smart Learning Environments* 6, 1. <https://doi.org/10.1186/s40561-019-0106-1>
 - [72] Paula Toledo Palomino, Armando M. Toda, Wilk Oliveira, Alexandra I. Cristea, and Seiji Isotani. 2019. Narrative for gamification in education: Why should you care? In *Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019*, 97–99. <https://doi.org/10.1109/ICALT.2019.00035>
 - [73] Gustavo F Tondello, Alberto Mora, and Lennart E Nacke. 2017. Elements of gameful design emerging from user preferences. In *CHI PLAY 2017 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 129–140. <https://doi.org/10.1145/3116595.3116627>
 - [74] Zachary O. Toups, Nicole K. Crenshaw, Rina R. Wehbe, Gustavo F. Tondello, and Lennart E. Nacke. 2016. “The collecting itself feels good”: Towards collection interfaces for digital game objects. In *CHI PLAY 2016 - Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 276–290. <https://doi.org/10.1145/2967934.2968088>
 - [75] Mike Treanor, Bobby Schweizer, Ian Bogost, and Michael Mateas. 2011. *Proceduralist Readings: How to find meaning in games with graphical logics*.
 - [76] Md Sami Uddin, Carl Gutwin, and Andy Cockburn. 2017. The effects of artificial landmarks on learning and performance in spatial-memory interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3843–3855. <https://doi.org/10.1145/3025453.3025497>
 - [77] Allan Wigfield and Jacquelynne S. Eccles. 2000. Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology* 25, 1: 68–81. <https://doi.org/10.1006/ceps.1999.1015>
 - [78] Jason Wuertz, Scott Bateman, and Anthony Tang. 2017. Why players use pings and annotations in Dota 2. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1978–1982. <https://doi.org/10.1145/3025453.3025967>
 - [79] Jason Wuertz, Max V. Birk, and Scott Bateman. 2019. Healthy lies: The effects of misrepresenting player health data on experience, behavior, and performance. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3290605.3300549>
 - [80] Tao Xie, Nikolai Tillmann, and Jonathan De Halleux. 2013. *Educational Software Engineering: Where Software Engineering, Education, and Gaming Meet*. Retrieved February 17, 2021 from <http://channel9.msdn.com/achievements/visualstudio>
 - [81] Bluma Zeigarnik. 1938. On finished and unfinished tasks. In *A source book of Gestalt psychology*. 300–314. <https://doi.org/10.1037/11496-025>

Received February 2021; revised June 2021; accepted July 2021.