Title Page:

Investigating game attention using the Distraction Recognition Paradigm

Authors

Joe Cutting (Corresponding Author), Digital Creativity Labs, University of York, UK
joe.cutting@york.ac.uk, ORCID iD 0000-0003-1856-4915
Paul Cairns, Department of Computer Science, University of York, UK
ORCID iD 0000-0002-6508-372X

Keywords: Games, Attention, Engagement, Distraction, Self-paced

Word Count: 11,144 excluding references

Abstract

Digital games are well known for holding players' attention and stopping them from being distracted by events around them. Being able to quantify how well games hold attention provides a behavioral foundation for measures of game engagement and a link to existing research on attention. We developed a new behavioral measure of how well games hold attention, based on players' post-game recognition of irrelevant distractors which are shown around the game. This is known as the Distractor Recognition Paradigm (DRP). In two studies we show that the DRP is an effective measure of how well self-paced games hold attention. We show that even simple self-paced games can hold players' attention completely and the consistency of attentional focus is moderated by game engagement. We compare the DRP to existing measures of both attention and engagement and consider how practical it is as a measure of game engagement. We find no evidence that eye tracking is a superior measure of attention to distractor recognition. We discuss existing research on attention and consider implications for areas such as motivation to play and serious games.

Introduction

Digital games undoubtedly form a major part of modern entertainment with common claims of being financially bigger than Hollywood (Chatfield, 2009). These games are well known for holding players' attention completely such that they are "immersed" (Brown & Cairns, 2004) or "lost in the game" (Etchells, 2019) which prevents them from being distracted by events around them. This paper investigates how well digital games hold players' attention and stop them being distracted by external stimuli.

Attention is a key mechanism in how we perceive the world and so impacts many areas of human-computer interaction. Attentional focus determines which user interface elements receive mental processing resources and which are ignored (Carrasco, 2011; Desimone & Duncan, 1995). This allocation of attentional resources then determines whether a display element is likely to be noticed via the mechanism of *inattentional blindness* (Mack & Rock, 1998; Simons & Chabris, 1999; Wood & Simons, 2019) or whether it has changed via *change blindness* (Rensink et al., 1997; Simons, 2000; Simons & Levin, 1997). Attention moderates the higher-level effects of interactive media such as learning, emotional change and motivation. Optimal learning is created when attention is focused on a single element (Ayres & Sweller, 2005; Chandler & Sweller, 1992). Attentional focus can help participants reduce negative emotions and promote positive emotional affect (Nix et al., 1995; Wadlinger & Isaacowitz, 2011). Schull (Schull, 2005, 2012) describes how players are attracted to gambling machines which "require just enough of your attention that you can't really think about anything else" and Murch et al. (2017) found that participants with gambling problems were less likely to have their attention distracted from slot machines.

Being engaged in playing a digital game is commonly understood to focus players' attention away from distractions and onto the game. This property of holding attention and preventing distraction could be understood as a fundamental component of the more general notion of engagement in digital games (Boyle et al., 2012; Cairns, 2016). Game engagement is more widely studied in academia and industry than attention in games as it is the more immediately obvious self-reported experience of players (Brown & Cairns, 2004). Measuring engagement is of interest because developing games is still an uncertain process due to the difficultly in predicting what player experience will be produced by particular game features. Game developer Bruce Phillips (Phillips, 2006, p.22) says "I have a secret longing for the confidence in purpose that I imagine my colleagues working on productivity applications must feel. Their goals seem communicable and measurable—mine don't." The experience of playing a game is complex, multifaceted and difficult to design for (McCarthy & Wright, 2004) but being able to measure even just the level of engagement players experience could give game developers assurance that their game has successfully produced a meaningful player experience ahead of release which may then lead to commercial success.

Much of the work on engagement in games to date has relied on questionnaires which, while informative as a self-reported measure, lack objectivity and a link to the psychological literature on attention. Considering a behavioral measure of attention specifically rather than engagement more broadly, could provide an objective foundation for this aspect of engagement and allow insights from existing attention studies to be applied to the experience of playing games. This is what led to the work of both Brockmyer et al (2009) and Jennett (2010) in looking to relate attentional focus as measured by distraction from audio cues to questionnaire-based measures of engagement. However, their measures of attentional focus were quite coarse offering only a few opportunities for distraction. This means the measure can indicate fewer different levels of attention and is more likely to be inaccurate if there is variation in participants level of attentional focus. Furthermore, they both used action games which necessarily require some attention in order to be played at all and thus, in the lab context, do not necessarily capture the voluntary commitment of attention that players make when playing out of that context.

The main goal of this paper is to develop a more effective measure of how well games hold attention and stop players becoming distracted, (which we refer to as a "measure of attention"). A new measure of attention could provide new insights into the experience of playing games and how gameplay creates effects which have been previously linked to attentional focus such as learning, reduction of negative emotions and motivation. These insights could then inform the design of serious games (Baranowski et al., 2008; Susi et al., 2007) for learning and behavior change. Attention and distraction have been linked to problem gambling (Murch et al., 2017; Schull, 2005) so a better measure of attention could give new insight into these issues. A behavioral measure of attention could provide a baseline measure to ground existing questionnaire-based methods which rely on self-report. To achieve these goals a new measure should be suitable for lab-based studies, discriminate between different game experiences and be broadly comparable to an existing engagement measure. A secondary goal of this paper is to use this measure to investigate how well games hold attention. We chose to focus on "self-paced games" (Jennett et al., 2008), which have no requirement for players to respond within a particular time period. We chose this type of game because the majority of game experience studies look at fast-moving action games (Mekler et al., 2014) and there is some evidence (Jennett et al., 2008) that self-paced games are less engaging than action games. We aimed to determine whether this type of game can fully hold players' attention. In these games, players can pause gameplay at any time, so are they more likely to be distracted if not fully engaged in the game?

This paper describes three studies which develop such a measure which is known as the Distractor Recognition Paradigm (DRP). The first study creates a baseline measure of image recognition with no game present. The second study measures attention using the DRP in two very different variants of the same game. The final study uses the DRP to compare three more similar game variants and investigates eye tracking as an alternative measure of attention.

Background

Measuring game engagement

The experience of being engaged in a game is complex and difficult to define and most approaches to measuring it are based on one particular underlying feature of the feeling of being engaged. Initial approaches to measuring game engagement (e.g. Chen, 2007) were based on the idea that being engaged in games induces a state of *Flow* (Csikszentmihalyi, 1991, 2013). Flow has been described by Csikszentmihalyi as an optimal state where the level of challenge meets the level of performance. Csikszentmihalyi measured whether participants were in flow using experience sampling methods which automatically polled the participant at set intervals. However, flow is not an accurate representation of most players' experience of playing games. Most games involve periods where the level of challenge is higher than the players' performance which are then followed by easier periods where the level of challenge is much lower (Schell, 2008). Attempts to use experience sampling with games have found that the act of sampling can interrupt the player and change the experience it was trying to measure (Kaye et al., 2018).

The most widespread method of measuring game engagement is using post-game questionnaires. These questionnaires are based on different aspects of engagement with inspiration from a range of sources. Brown and Cairns (2004) performed a grounded theory investigation of what the concept of engagement meant to players and found the players described the idea of "immersion" which is the feeling of being totally engrossed by a game. This led into the production of a validated questionnaire of game enagement known as the Immersion Experience Questionnaire (IEQ) (Jennett et al., 2008). Ryan et al. (2006) were inspired by the self determination theory of motivation to create the PENS (Player Experience of Needs Satisfaction) game engagement questionnaire which considers engagement by how it motivates players to perform particular actions such as continuing to play. Brockmyer et al. (2009) considered the likelihood that a particular player would become engaged by a game. They used this to produce and validate the Game Engagement Questionnaire (GExp). Denisova et al. (2016) analysed all three of these questionnaires and found a high degree of correlation between the results which suggests that they all measure similar underlying concepts. Post-game questionnaires have also been used to measure particular aspects of the game playing experience such as *Challenge* (Denisova et al., 2017) and *Uncertainty* (Power et al., 2018). There are several drawbacks with questionnaires. One is that they are a self-reported measure so rely on participants' reflection on their experience which can be unreliable (Gutwin et al., 2016; Kahneman et al., 1993) Another drawback is that they are less accurate when measuring experiences with low engagement. Jennett et al. (2008) found a questionnaire reported moderate levels of engagement with even very boring box clicking tasks that could hardly be considered a game.

To overcome issues with questionnaires, some measures of engagement are based on how physiological properties of players' bodies vary depending on their game experience. These measures are taken directly from the players' body so do not rely on self-report. There are many different physiological measures which have been applied to different games. Ambinder (2011) measured players' skin conductance and found that it corresponded to players' level of arousal in the zombie shooter game *Left 4 Dead.* He then used this data to tune an AI director which ensured a varied level of arousal across the time of the game. Nacke and Lindley (2008) manipulated three different levels of the shooter *Half Life 2* and measured EEG (electroencephalography), heart rate and skin conductance. They found that different level designs created different physiological readings. Izzetoglu et al. (2003) used a technique known as fNIR, which uses infrared sensors to measure blood oxygenation in the cortex, to measure the cognitive load used in a game-like high intensity “Warship Commander” task. However Harmat et al. (2015) found no relationship between fNIR and flow in participants who played the game *Tetris*. Mandryk and Atkins (2007) created a measure which combined four different physiological measures. These were skin conductance, heart rate and two different forms of Electromyography which is a measure of the electrical activity of muscles. They combined these measures into a fuzzy logical model which then corresponded to emotional states such as boredom and excitement in the action game *NHL 2003.* Current physiological measures are developed using action games and many of them correspond in some way to increased bodily arousal that is created when participants play a fast moving and challenging action game.

However, many highly engaging games fall into the category of self-paced games (Jennett et al., 2008) which give players as long as they need to make their moves and are unlikely to lead to changes in arousal in the same way as action games. This means that many existing physiological measures are unlikely to be effective. It seems unlikely that playing the self-paced strategy game *Civilization* (in which players build a civilization from prehistory up to the near future) will have the same effect on players’ heart rate as playing action shooters like *Half Life.* Some measures such as EEG and fNIR may not be as dependent on arousal and more suitable for slower games. However, it is often unclear what exactly these techniques are measuring which may introduce additional confounds to the data. Measuring how well a game holds players’ attention has a much clearer link to engagement and a new game experience measure based on attention may have the advantages of physiological measures but be suitable for self-paced games that do not rely on physical reactivity.

Measuring game attention

Digital games are well known for holding player’s attention and stopping them being distracted. How this happens is not well investigated although the effects and mechanisms of attentional focus have been extensively investigated (See Carrasco (2011) for a review). Attention can be focused by both *bottom-up* and *top-down* mechanisms. In bottom-up mechanisms features in the stimulus, such as visual characteristics (Wolfe, 2014) or the task being performed, focus attention *intrinsically* without conscious control of the participant. Load theory (Lavie, 2005; Lavie et al., 2014; Lavie et al., 2004) shows that bottom-up attentional focus is increased by high perceptual load and decreased by high cognitive load. In top-down mechanisms attention is focused *extrinsically* due to conscious control of the participant which may result in them creating an *attentional set* (Most et al., 2001) of elements that they should focus their attention on. Many situations involve both top-down and bottom-up attentional focus, such as magic tricks (Kuhn et al., 2016), in which attentional misdirection happens by both top-down directions from the magician and bottom-up manipulation of the stimulus seen by the audience. Digital games work on a number of levels ranging from the high-level story to the low-level game mechanics (Arnab et al., 2015; Calleja, 2007; Sicart, 2008), so it is possible that they hold attentional focus through a combination of both top-down and bottom-up mechanisms. High level game elements such as the story and level progression would provide top-down motivation for focusing attention whereas low-level elements such as visual design and game mechanics would provide bottom-up intrinsic attentional focus.

Sustained attentional focus over a longer period of time has been measured in non-game contexts. Smallwood et al. (2008) investigated the phenomenon of mind wandering which happens when attention drifts off task. Participants read a Sherlock Holmes story and were asked to periodically indicate if they were on-task or if their mind had wandered. Those who indicated that they were off task more often were less likely to understand who was the villain of the story. Murch et al. (2017)

used a similar technique to measured attentional focus on slot machines. The slot machines were surrounded by two screens which contained red squares and white circles. Participants were instructed to disregard the white circles and press a particular physical button whenever they saw a red square. They found an inverse relationship between participants' accuracy at spotting the red squares and their risk of problem gambling. However, almost all participants spotted at least 80% of the squares which suggests that there may have been a ceiling effect to the distraction measure. Both of these measures have the disadvantage that they interrupt the activity that is being performed and Murch et al's measure in particular requires participants to perform another task which may detract from their main task of playing the slot machine. "Banner blindness", in which web users are unaware of advertising banners has been investigated using less intrusive measures of attention. Burke et al. (2005) showed banners while users were performing another task and found that in a post-task forced choice recognition test, users recognized few banners (~20%) because their attention was occupied by the task. Hervet et al. (2011) looked at banner blindness using eye tracking and found low rates of recall despite finding that participants fixated their gaze on almost all the banners.

Similar techniques have also been used for initial investigations into attention in digital games. Brockmyer et al. (2009) validated their GEQ game experience questionnaire by using a simple indicator of how well the game held players' attention. Participants played a game and heard a voice asking them if they had dropped their keys. Those who were more engaged in the game were less likely to respond to the voice. Similarly Jennett (2010) played ten different audio clips during gameplay; those players who were more immersed in the game were less likely to be distracted by outside events and remembered hearing fewer audio clips after the game. Both of these measures have the disadvantage that they rely on just a few distractors which makes the measure coarse grained. Jennett's measure has more distractors but this makes it more likely that participants will not remember an audio distractor even if they heard it.

A new measure of game attention

We designed a new measure of sustained attention which overcomes these issues and those of other measures from Smallwood et al. (2008) and Murch et al. (2017) mentioned earlier. This measure was inspired by a study by Kinoshita (1995) which found that participants only remembered seeing the parts of the stimulus that they had been paying attention to for the task. Our approach measures game attention using visual distractors, if participants' attention is focused on the game then they will not remember the distractors. However, if their attention drifts from the game they are likely to remember the visual distractors. This is a similar technique to the audio distractor technique used by Jennett (2010) and Brockmyer et al. (2009) except that using visual distractors has several advantages. One advantage is that visual distractors can be shown and comprehended much faster than audio distractors. Each audio distractor would have to be several seconds long and participants' attention would have to drift for all that time for them to be sure of hearing all of the sound, whereas a visual image can be processed almost immediately, in some cases this is less than 150ms (Thorpe et al., 1996). Another advantage is that visual distractors are more memorable than audio distractors. Miller and Tanis (1971) found that participants remembered only 75% of an audio stimulus whereas both Standing (1973) and Brady et al. (2008) showed participants thousands of images and found that they remembered over 90% of them. This means that using visual distractors will be more sensitive than audio distractors as participants are more likely to remember the distractor and also finer grained because images can be comprehended much faster than sounds, which allows more distractors to be used. Our new measure is known as the *Distractor Recognition*

Paradigm (DRP) and it works by showing many irrelevant changing distractor images during game play and afterwards using a forced choice recognition test to assess how many images are remembered.

Eye tracking has been used to measure where players look during gameplay (e.g. Alkan & Cagiltay, 2007; Johansen et al., 2008; Zain et al., 2011) and may be more accurate than a post-game recognition test for measuring whether attention was captured by the distraction images because it does not rely on memory of the images. Despite this Hervet et al. (2011) found that banner blindness was better measured by testing recognition rather than using eye tracking. As part of our development of the DRP we also compared eye tracking as a measure of attention to the DRP.

Research questions

RQ1 Is the DRP an effective technique for measuring how well games hold players' attention?

RQ2 Can this measure be used as a measure of game engagement with advantages over existing measures?

RQ3 Is eye-tracking a more effective measure of attention than post-game testing of distractor images?

We performed 3 experiments to answer these questions. Experiment 1 determines a baseline level of image recognition used in the other experiments. Experiments 2 and 3 investigate RQs 1 and 2, with experiment 3 also considering RQ3.

Experiment 1: Calibration memory test of distractor images

Aims

This experiment determines a baseline level of recognition of the intended distractors without any game present, which shows the potential sensitivity of distractor recognition for measuring player experience. This gives a ceiling for the measure when players might be presumed to be fully attending to the distractors. If the general recognition level is high, then that would indicate that the DRP may be an effective measure as participants who were not at all engaged by the game would recognize most of the distractor images.

Design and Materials

This was a single condition experiment. All participants were shown 60 distractor images and then tested on 30 to see how many they could recognize. Each image was shown for 5 seconds (as in Standing, 1973) in a random order which differed for each participant. Showing 60 images, each for 5 seconds, takes 5 minutes, which is approximately the length of a game level. The images were icons taken from the *Webdings* typeface as these provide a wide variety of images which are consistent in terms of styling, size and color. This reduces the chance that some distractors might be more memorable due to particular attributes or features. The images to be shown were chosen randomly for each participant from a pool of 90. After being shown the images participants were tested to see how many they recognized. This used a forced choice test (similar to Hervet et al., 2011; Standing, 1973) in which participants choose between one image that they had been previously shown and another that was completely new to them. Using a forced choice test avoids increased variance due to different levels of confidence in ability at the test. The questions were displayed in a random order, so it is theoretically possible for the last displayed image to be the first one tested but as 60 images are shown and 30 are tested this is very unlikely.

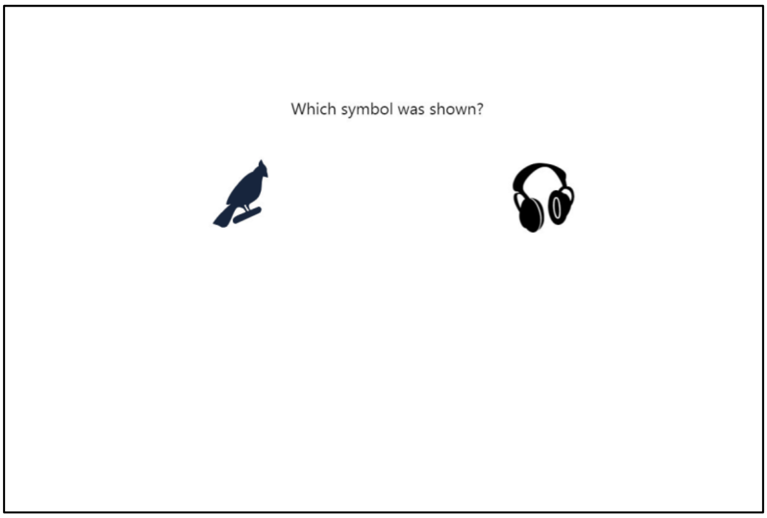

Figure 1 The distractor recognition test. Participants need to choose one image from the two. One of these images has been shown to the participant during the experiment. The other has not been shown before.

Participants

Standing (1973) used between 5 and 10 participants for his investigations into image recognition memory. Following this example, 10 staff and students from colleges in York, UK took part in the study. 8 were male and 8 were native speakers of English with ages ranging from 18-27 (Mean 20.3)

Procedure

Participants performed a consent procedure and were then told to watch the screen and that they would be asked some questions afterwards. The stimulus was shown on a 24" inch monitor with screen dimensions of 51.5 x 32.5cm. Participants kept their chin in a chin-rest which was positioned 95cm from the screen so that the screen display filled 31.5° of the participant's field of view. The participants were then shown 60 distractor images in the presentation phase of the experiment. Immediately afterwards they performed a forced choice recognition test on 30 images.

Results

The mean number of images recognized was 27.5 out of 30 (SD=2.88). This is equivalent to recognizing 91.6% of the images. A single sample t-test showed that the level of recognition is significantly different from the result that would be found by chance with a very large effect size. ($t(9)= 13.74$, $p<0.001$, $d=4.34$).

Discussion

Participants recognized almost all the images that they had been previously shown. They correctly recognized an average of 91.6% images which is consistent with Standing (1973)'s finding that participants recognized over 90% of images. This confirms that this set of *Webdings* distraction images has a high recognition rate if participants are paying full attention to them. The high rate of recognition ensures that there is sufficient variability available for the recognition rates to be significantly impacted by changes in attention. If the same icons are shown at the same time as participants are playing a game, then the number that they recognize afterwards is likely to be inversely related to how well the game holds their attention

Experiment 2: Comparing two very different games

Aims

The experiment aimed to investigate how well a simple self-paced game holds players' attention. This was an initial feasibility test of using the DRP as a measure of how well games hold attention. As an unproven measure there was no previous empirical data to form an estimate of likely effect sizes. We decided to use two very different games which were likely to give a large difference in how well they held attention. The games used were based on a popular self-paced puzzle game called *Two Dots* which is described in more detail below. These games were designed to be similar in graphics and interaction but to produce large differences in engagement to produce a significant difference in how well they held participants' attention. To ensure that players were indeed experiencing different level of engagement we also measured the level of immersion in the two games using the IEQ.

Hypothesis

The hypotheses of the experiment were:
H1: The number of distractors that participants recognize will be higher for the less engaging game condition than the more engaging game condition.
H2: Participants will have a higher immersion score for the more engaging game condition than for less engaging game condition.

Method

Design

This was a between-subjects design with two conditions. The independent variable was the game each participant played. The more engaging game was known as the *Full game* and the less engaging game known as the *Reduced game*. These are described below under *Materials.* The main dependent variable was the number of images that participants recognized after the game. Another, secondary dependent variable is the IEQ score for each participant's experience of the game.

Participants

A pilot study (n=16), which is not reported here, found a significant difference in immersion between these games, so for this initial attention experiment we decided to use 20 participants, who were students and staff from the University of York. 12 were men. Ages ranged from 21 to 50 (mean = 30.7). All participants received chocolate for their participation.

Materials

Both games used in this experiment were variants of the popular mobile puzzle game *Two Dots*. This game was chosen because it is a simple self-paced puzzle game with a minimal number of additional features. It is also very successful and has been installed by millions of players (Crook, 2014; Fine, 2015) which suggests that players find it very engaging. The game of *Two Dots* is played on a grid of dots of different colors. The player has to drag a line to join two or more dots of the same color which are next to each to other. When they release the mouse the dots that they have joined disappear. These gaps are filled by the remaining dots dropping down. Any gaps still remaining are filled by new dots dropping from the top of the screen. Each level has a set of targets at the top of the screen which indicate how many dots of each color need to be joined. For example, the screen in Figure 2 shows that the player needs to join 40 blue dots as well as 40 red, green and yellow dots. To succeed at the level, the player needs to join this number of dots within a move limit which is shown in the top left of the screen. If players join enough dots within the move limit, they complete the level and move onto the next level which has a different grid of dots with different targets. If players fail to remove that number of dots, then they fail the level and replay it from the start.

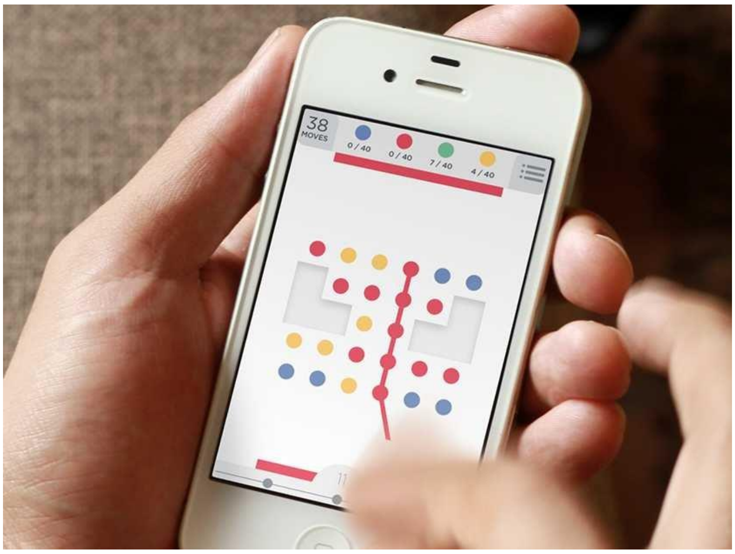

Figure 2 Two Dots being played on a phone. Players join the dots in the grid to meet the targets at the top of the screen within the move limit. Image © Playdots, Inc

Participants played one of two variants of the game; known as the *Full game* or the *Reduced game*. The *Full game* is a direct copy of *Two Dots* which runs on a Windows PC and is controlled by a mouse. (See Figure 3). The *Reduced game* is based on the *Full* game, but all the dots are the same color and the move counter, targets and levels are removed. Players can still join dots which disappear and drop down but there is no challenge or progression in the game. During the play section of the experiment both games were surrounded by repeated distractor images which changed every 5 seconds (See Figure 5). As in the initial experiment participants were shown 60 distractor images and tested on 30 of them. Each participant was shown the same set of distractors which were presented in the same order, in the next experiment this was randomized.

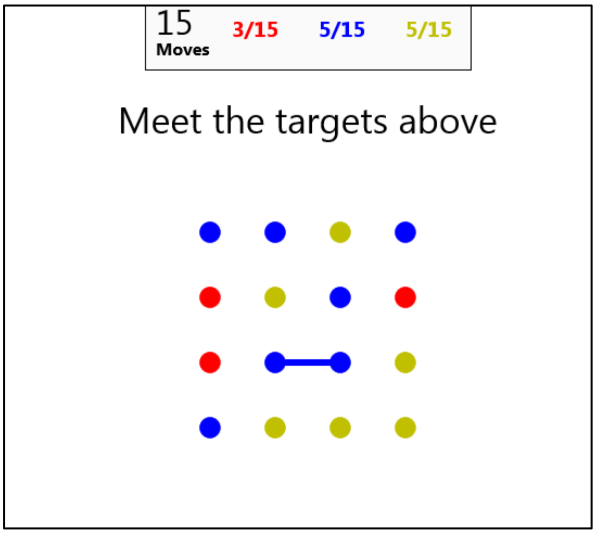

Figure 3 The *Full game* which is a clone of *Two Dots* which is played using a mouse on a computer

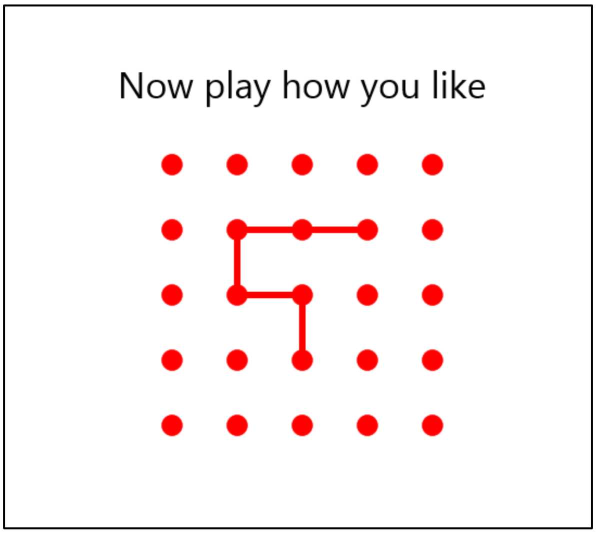

Figure 4 The *Reduced game*. Players can join dots to remove them, but there is no challenge or target

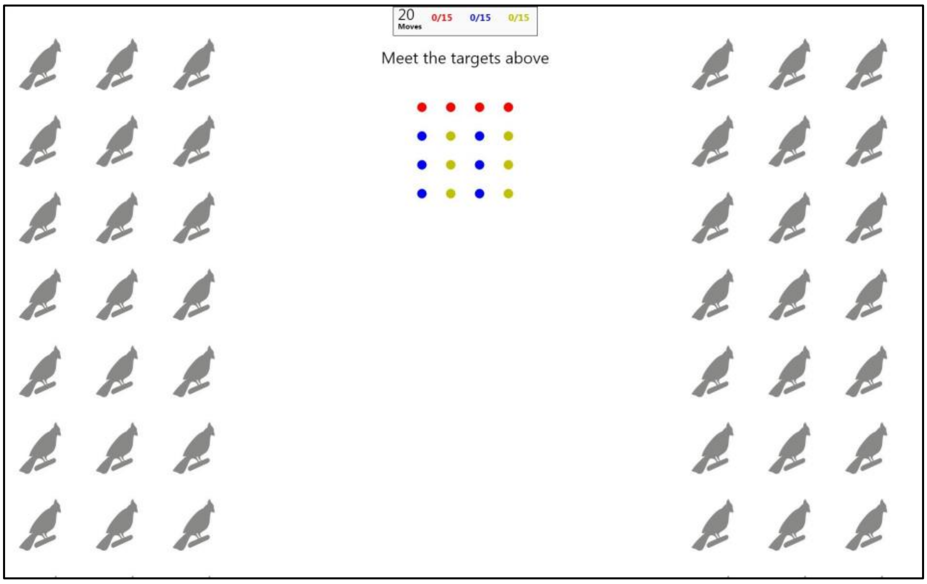

Figure 5 The Full game surrounded by distractor images

Procedure

Participants began by completing a consent procedure. They then played either the *Full game* condition or the *Reduced game* condition. They were told to play the game as well as they could and given no instruction regarding the images. Both games included a short tutorial to teach participants how to play. Participants played the game for 5 minutes. During play, participants rested their chin on a chin rest so that the distractors and game were in a fixed angle of view, with the distractors in the peripheral vision of the players. This means that for players to be able to remember an image, they would need to divert their attention away from the game and look at the distractors. After 5 minutes of play, the game stopped automatically, and the participants then completed the on-screen distractor recognition test which was identical to the test described in the first calibration experiment. This was done immediately after the game so that all participants in both conditions would not forget any distractors or have their memories confused by other tasks. After the distractor test participants came away from the chin rest and screen and filled in a paper based IEQ about their experience of the game.

Results

Distractor images recognized

There was a significant difference and large effect size in the number of distractors correctly recognized between the *Full game* (M=13.0, SD=3.1) and the *Reduced game* (M=19.4, SD= 3.9) conditions; F(1,18)= 16.28, p=.001, η_p^2= .475. A single sample one sided t-test found that the number of distractors recognized for the *Full game* was not significantly higher than the number which would be recognized by chance (15); t(9)= -2.05, p = .965.

We also calculated the chance of each individual image being recognized (Figure 6) and the chance of an image being recognized at each time period of the game (Figure 7). The top 10% of images most likely to be recognized for the *Full game* (75, 205, 15, 55, 25, 105) and the *Reduced game (*205, 85, 155, 165, 55, 75) had three images in common to both groups (205, 75,55). Removing these three images from the analysis still showed a significant difference and large effect size between conditions; F(1,18)= 13.6, p=.002, η_p^2= .430.

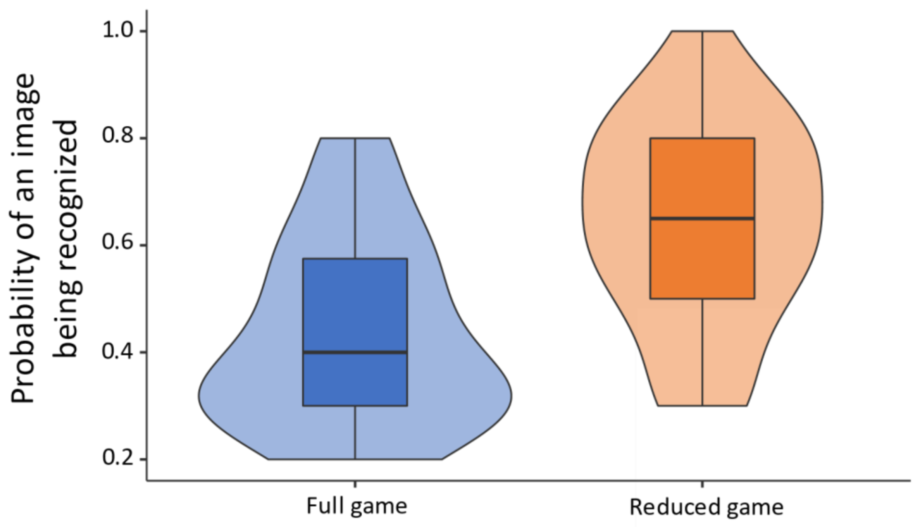

Figure 6 Violin plot of the probability that each individual image will be recognized

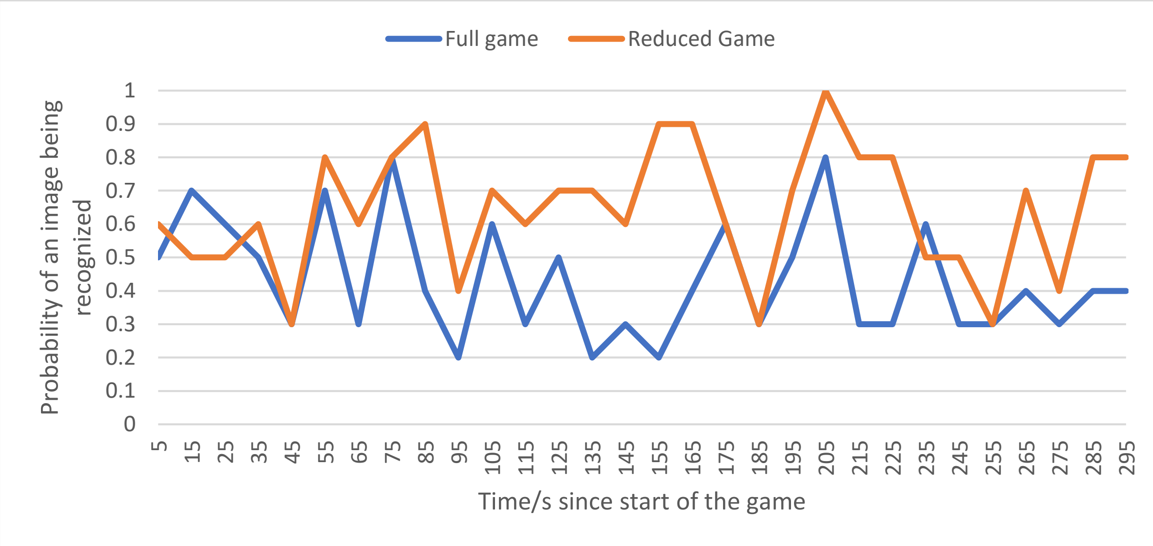

Figure 7 Probability of an image being recognised over the time of the game

Immersion Experience Questionnaire (IEQ)

There was a significant difference and extremely large effect size in the immersion scores between the *Full game* (M=110.2, SD=14.4) and the *Reduced game* (M= 84.6, SD= 10.8) conditions; F(1,18)= 20.30,p<0.001, η_p^2= 0.530.

Discussion

Participants recognized very few distractor images after playing the *Full game.* The recognition rates were not significantly higher than that which would have been achieved by chance and far fewer images than they recognized in the first experiment which did not have a game activity. This indicates that, despite being a self-paced game in which there is no requirement to respond quickly, the *Full game* held participants attention entirely and stopped them being distracted by events around them. This is also in notable contrast to the results found by Murch et al. (2017) in which participants responded to over 80% of the surrounding distractor images. The difference is likely to be because Murch et al told participants in advance that they had to respond to the distractors so their attention was split between the slot machine task and distractor task. In the DRP participants are not told about the distractors before the task so they only pay attention to them if the main game task does not hold their attention.

There was also a large difference between games in the number of distractors recognized. The first hypothesis that the number of distractors recognized by participants who played the *Reduced game* would be higher than the number remembered by those in playing the *Full game* was supported with an extremely large effect size. This indicates that the DRP may be an effective measure of how well games hold players' attention. The second hypothesis that the immersion score for participants who played the *Full game* would be higher than for those who played the *Reduced game* was also supported with an extremely large effect size. This indicates that participants who played different games had a significantly different levels of engagement in the game. The *Reduced game* had a similar visual stimulus (grid of dots) and the same dot joining interaction as the *Full game* so the difference in attention is likely to be due to other factors such as the level of engagement created by the games. This supports our initial proposal that how well games hold attention is linked to the level of engagement in the game rather than just the very low-level properties of the stimulus and the interactions that players are performing.

We were concerned that some images might be more memorable than others. Plotting the probabilities on a violin plot (Figure 6) showed some variation in the chance that each image would be recognized, but this may be due to random variation rather than differences between images. Looking at the images most likely to be recognized, 3 images were in the top 10% of both conditions. Removing these images from the analysis still shows a significant difference between conditions but with a reduced effect size because even the most memorable images are recognized more in the *Reduced game* than the *Full game*. The same images were shown in both conditions so even if some images were more memorable this would not threaten the validity of the results although it might change the variance in the data. To reduce the possibility that a particularly memorable image would be shown at the same time in the game, the next experiment randomized both the set of images which were shown and the order they were shown in. We also plotted the probability of images being recognized over the time of the game (Figure 7) but found no evident pattern. Boxplots (available in the supplementary materials) of the number of images recognized and immersion scores showed very few outliers, so these are unlikely to have biased the results.

Experiment 3: Comparing three more similar games with eye tracking

This experiment tested the DRP with three more similar games and tested eye tracking as an alternative measure of attention. A limitation of the previous experiment was that the two games compared were very different. Although both games used similar visual stimulus and input mechanisms there was a very large difference in engagement and attention between the two games. This was due to the differences in game mechanics between the two games. A useful measure of attention in games would be able to differentiate between more similar levels of engagement, so this experiment uses games which are more similar to each other. Another possible issue is that all participants saw the same distractor at the same time for both conditions. A boxplot of image recognition probability showed no evidence that any one distractor image was more memorable than another but if it was, this could distort the data if all participants saw a particularly memorable image at the same point in the game. For this experiment the order of distractors was randomized for each participant.

Another limitation of the previous experiment is the possibility that participants looked at the distractors but did not remember them afterwards. This uncertainty about whether participants recalled a particular distractor may introduce variance which would reduce the sensitivity and reliability of the measure. An alternative to testing participants' memory of the distractor images is to use eye tracking to see what proportion of time they are looking at the distractor images compared to the game. Many studies have found that gaze target corresponds to the target attention (Deubel & Schneider, 1996; Shepherd et al., 1986). In particular, Shepherd et al. (1986) found that although attention could be focused without eye movement, eye movement always resulted in a shift in attention. When playing a game that does not hold their attention, participants' gaze direction may drift to the surrounding distractor images. Modern eye tracking equipment has a very high degree of precision and it may create a more accurate measure than testing for recognition of distractors. Using eye tracking to record players' gaze direction may be a more robust measure due to being a more direct measure of attention which does not depend on memory.

Aims

This experiment aimed to see whether visual distractors could be used as a measure of how well a game holds players' attention for more similar games. Another aim was to see whether tracking participants' gaze could be used as a more accurate measure of how well a game holds players' attention than the number of distractors recognized after the game.

Hypotheses

The hypotheses of the experiment were:

H1: The number of distractors that participants remember will be different in the less engaging game variants than in the full game.
H2: The immersion score of the two game variants will be less than that of the full game.
H3: The amount of time that participants spend looking at the game rather than the distractors will be higher for the full game rather than the game variants.
H4: Eye tracking will be a more accurate measure of attention than the number of distractors recognized. This will be indicated by a higher effect size between conditions for the eye tracking measure compared to the DRP.

Method

Design

This was a between-subjects design with three conditions. The independent variable was the game each participant played; the three different games are described below under materials. The main dependent variable was the number of distractors that participants recognize after the activity. Another dependent variable was the IEQ score as a measure of each participants' level of engagement in the game. A final dependent variable was the percentage of time that participants were fixated on the central game area rather than the surrounding distractors.

Participants

We performed a power calculation to estimate how many participants to have in the study. The effect size of the previous experiment was η_p^2= 0.475. This is equivalent to a Cohen's f of 0.951. We expected this experiment to have a smaller effect size as the games were more similar. So, we divided the previous f value by 2 to give an expected f of 0.476. Using this effect size in a power calculation with a power of 0.8 (80%), an alpha of 0.05 and 3 conditions gives 15.24 participants per condition which we rounded up to 16 for each condition.

48 students and staff from the University of York took part in the study. 24 were male. Ages ranged from 18 to 62 (mean = 22.3). Previous game experience and attitudes varied between the participants, ranging from those who played games less than once per month, to those who played several times a week. Despite the wide range of ages and experience there was no noticeable difference in results between older or younger participants. In the previous experiment participants were paid in chocolate but this made it more difficult to recruit sufficient participants. So, in this experiment all participants were paid £6.

Materials

Participants played one of three games all of which are variants of *Two Dots.* In a change from the previous experiment all the games were changed so that they were in monochrome (levels of grey) rather than color. This was because participants were monitored by an eye tracker during these games which also measured pupil size which was used in a separate analysis not reported here. Pupil size changes depending on light levels and changing to monochrome graphics reduces large changes in light levels which may have affected participants' pupil size. The different colored dots were changed to different symbols. These changes were consistent across all the game variants and there is no evidence that they affected game engagement; the commercial version of *Two Dots* has a "color blind" mode which replaces the different dot colors with symbols in the same way.

The first variant is the *Full game* which is the same direct copy of *Two Dots* which was used in the previous experiment (see Figure 8). The other two games were variants of *Two Dots* but with different game elements removed to make them less engaging. These two variants were primarily designed to create an experience less engaging that the *Full game* but more engaging than the *Reduced game* used in the previous experiment. We also took the opportunity to use this as preliminary investigation into the effects of higher and lower level game mechanics on engagement and attention. The first variant removes some of the higher-level game mechanics which may create extrinsic motivation to pay attention to the game. This variant is known as the *No goals* game and shown in Figure 9. This variant keeps the different dots which need to be joined together and there are still counters at the top of the screen which keep track of how many dots of each type have been joined. These could be seen as lower level mechanisms which create intrinsic attention on to the game. However, the targets and move counter have been removed. Because there are no targets to meet in joining dots this means that there are also no levels in this game. So, players are given the instruction "Now play how you like" and just join dots which disappear and add to the dot totals until the time runs out. The second game variant removes some of the lower level mechanics which may create intrinsic attention on to the game. This is known as the *All dots the same* game and shown in Figure 10. In this variant all the dots are the same which makes it trivially easy to find some that are the same and join them together. However, this variant keeps the same high-level targets and levels as the *Full game,* so players still have the task of joining enough dots to meet the target which then takes them onto the next level. The next level is identical to the previous level except that the target number of dots to join has doubled.

All of the game variants were surrounded by distractor images in the same way as the previous experiment (See Figure 11). The only difference was that the images to be shown were chosen randomly for each participant from a pool of 90. The images were shown in a random order which differed for each participant.

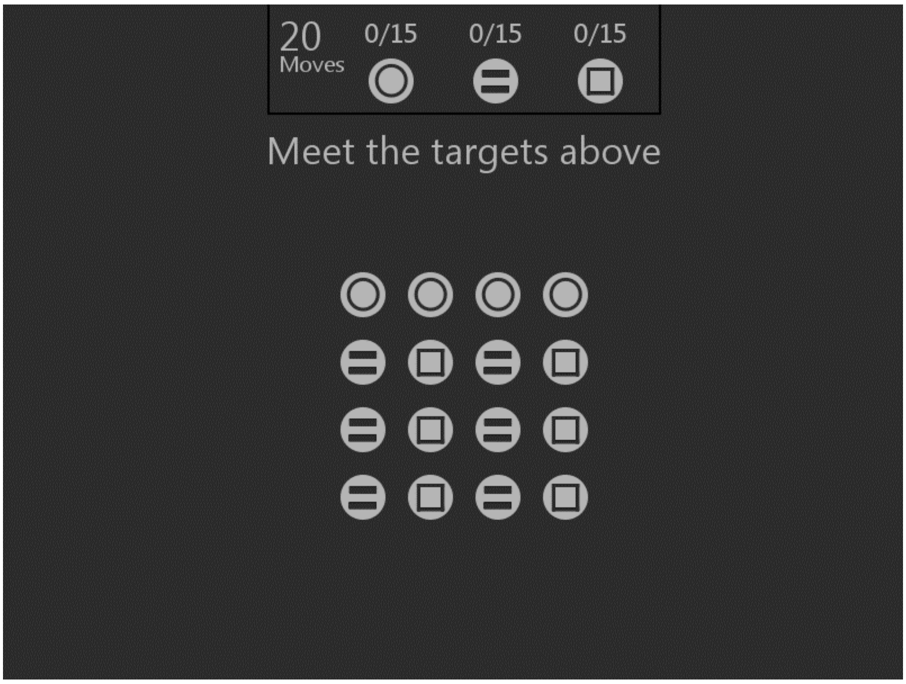

Figure 8 The monochrome version of the *Full game* of *Two Dots*

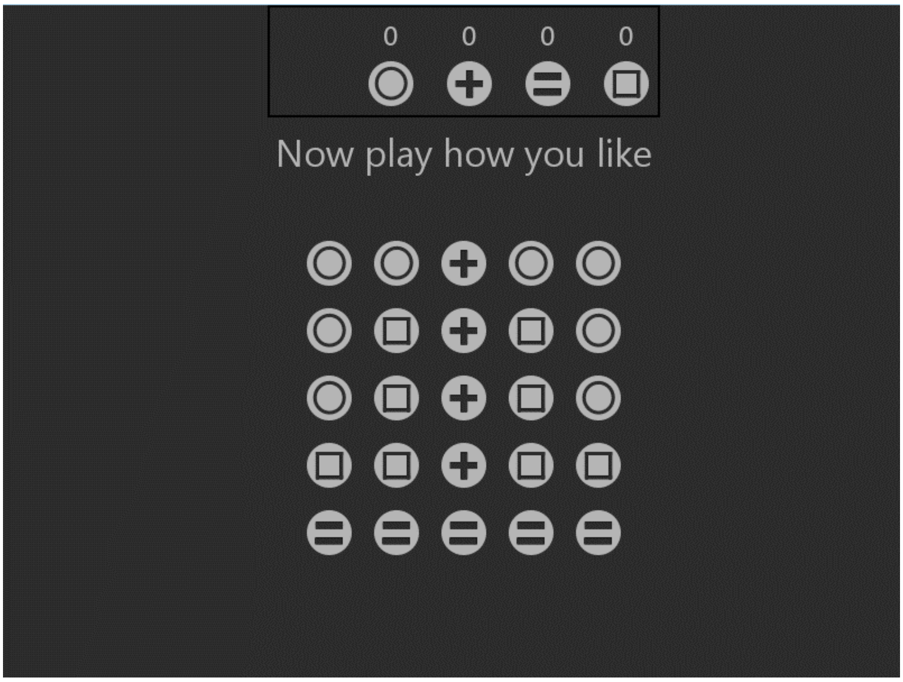

Figure 9 The *No goals game*. Players can still join the dots in the center of the screen. The game keeps track of how many of each dots of each type have been joined but there are no targets and no move limit.

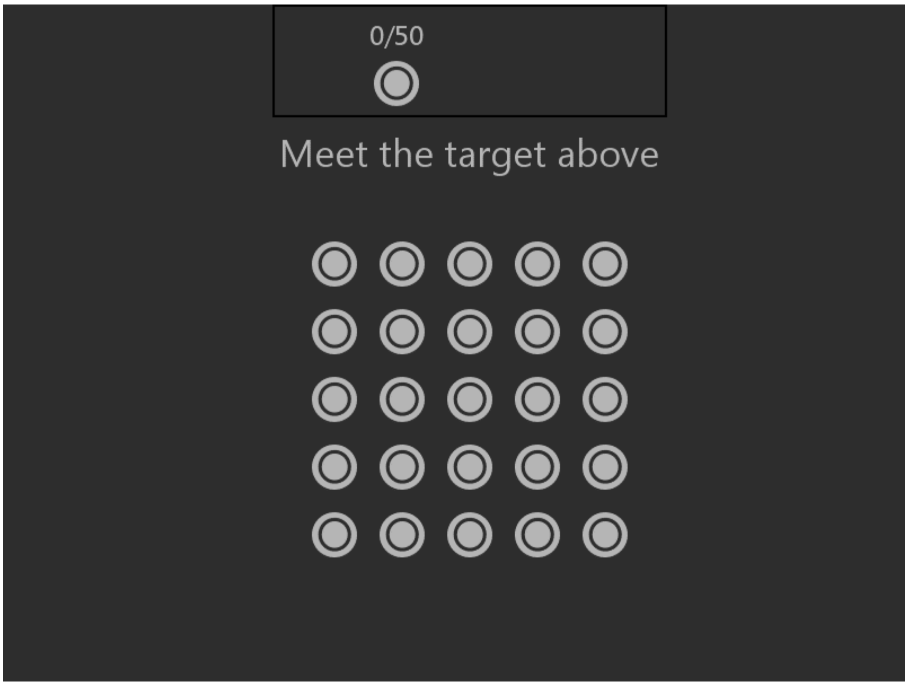

Figure 10 The *All dots the same* game has all the dots the same symbol. Players can still join the dots in the center of the screen. Players have to reach the target at the top of the screen to get to the next level.

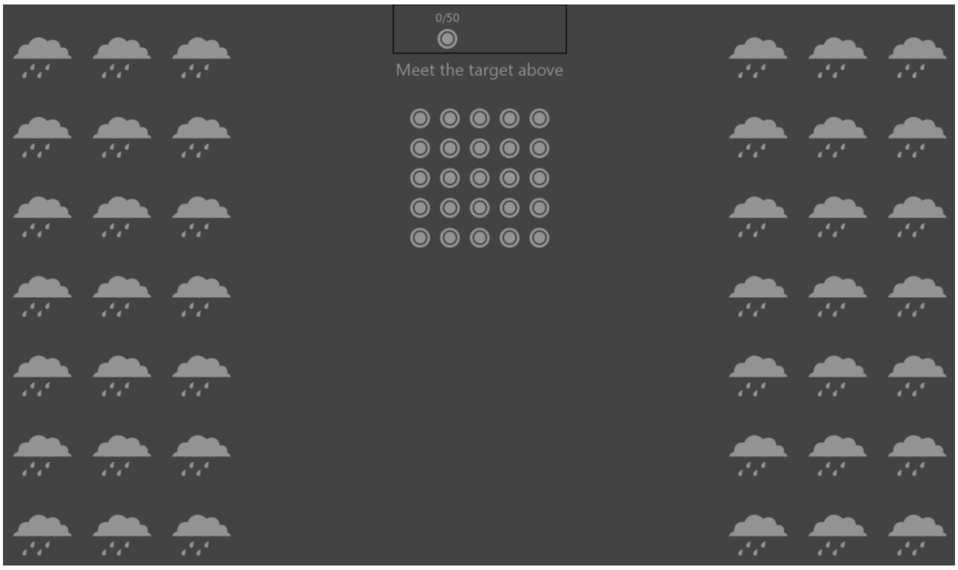

Figure 11 The *All dots the same* condition surrounded by distractors

Experimental setup and procedure

The experimental setup and procedure were the same as for the previous experiment except that participants' gaze position was recorded by an eye tracker. The eye tracking equipment used was an Eyelink 1000 Plus made by SR Research (https://www.sr-research.com/). This uses a desk mounted camera which tracks eye movements on a desktop computer screen. The camera recorded participants' eye position at a frequency of 250Hz which is more than adequate to accurately record fixations and saccades. The experimental procedure was also exactly the same except that after completing the consent procedure participants performed an eye tracker calibration task to configure the eye tracker.

Results

Distractor recognition

There was a significant difference in the number of correct distractors recognized between the *Full game* (M=14.13, SD=3.00), the *No goals* game (M=16.94, SD= 2.93) and the *All dots the same* game (M=16.31, SD=2.60) conditions; F(1,46)= 4.336, p=0.019, η_p^2= 0.162. A single sample one sided t-test found that the number of distractors recognized for the *Full game* was not significantly higher than the number which would be recognized by chance (15); t(15)= -1.181, p = .128,

We performed a Tukey's HSD post-hoc test to investigate which conditions were significantly different from each other.

Condition	*No goals*	*All dots the same*
Full game	p=0.020	p=0.085
No goals		p=0.808

Table 1 Tukey's HSD comparing distractors remembered for all conditions

There was a significant difference between the *Full game* and the *No goals* game. The difference between the *Full game* and the *All dots the same* game was not significant but tended towards significance (p=0.085). There was no significant difference between the *No goals* and *All dots the same* games.

We also calculated the chance of each individual image being recognized (Figure 12) and the chance of an image being recognized at each time period of the game (Figure 13). The top 10% of images most likely to be recognized for the *Full game* (168,134,119,116,85,66), *No goals* game (241,230,213,181,107,64) and *All dot the same* game (161,111,106,81,78,67) have no overlap between conditions.

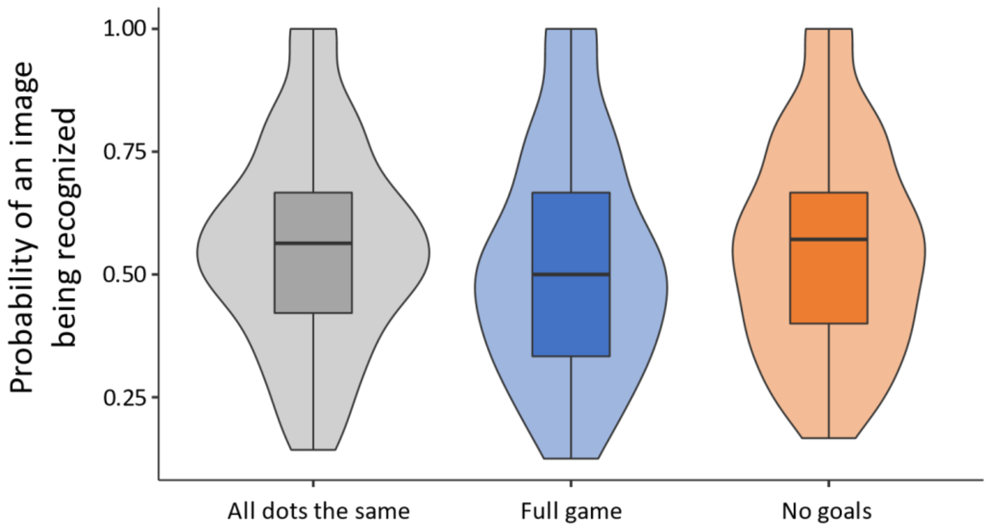

Figure 12 Probability of each individual image being recognised

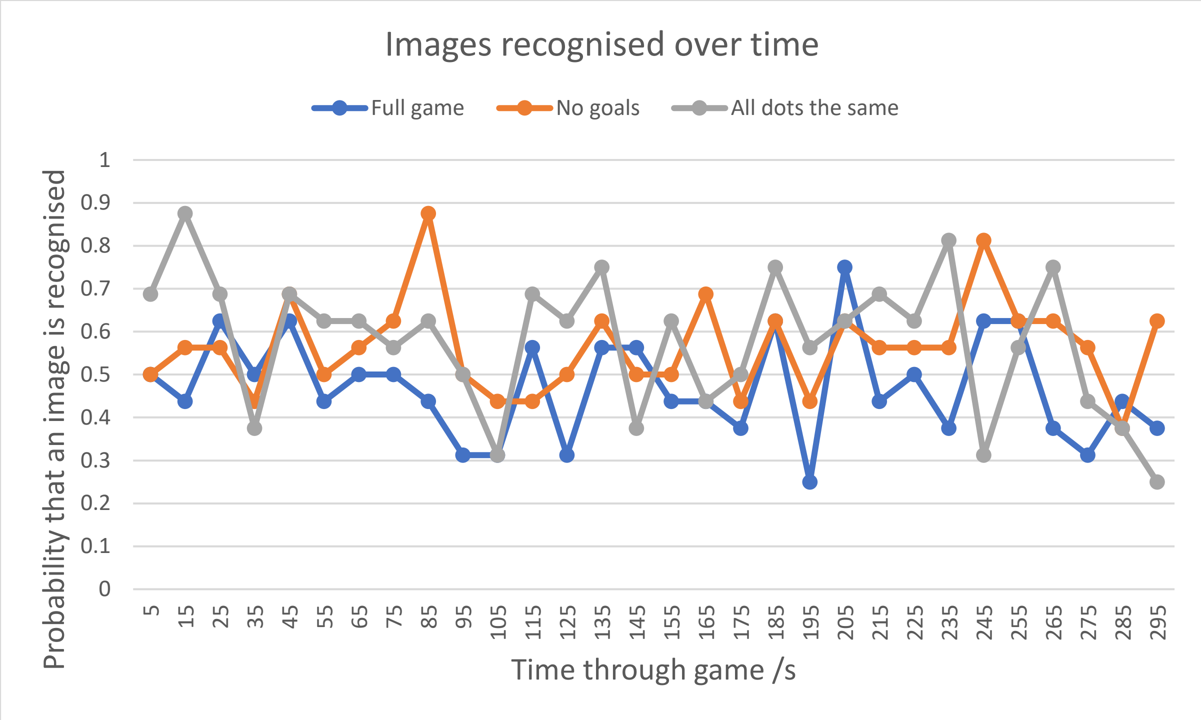

Figure 13 Probability of an image being recognised over the time of the game

Immersion Experience Questionnaire (IEQ)

There was a significant difference in the immersion scores between the *Full game* (M=107.13, SD=17.00), the *No Goals* game (M= 93.38, SD= 14.05) and the *All dots the same* game (M=93.56, SD=13.38) conditions; F(2,45)= 4.492,p=0.017, η_p^2= 0.166.

We performed a Tukey's HSD post-hoc test to investigate which conditions were significantly different from each other.

Condition	*No goals*	*All dots the same*
Full game	p=0.032	p=0.039
No goals		p=0.999

Table 2 Post-hoc test on IEQ scores between three game variants

There was significant difference between the *Full game* and the *No goals* game. There was also a significant difference between the *Full game* and the *All dots the same* game. There was no significant difference between the *No goals* and *All dots the same* games which had almost identical levels of mean immersion.

Eye tracking

To systematically analyze participants' gaze movements, we defined an Area of Interest (AOI) which filled the middle third of the screen that contained the game but not the distractor images (see Figure 14)

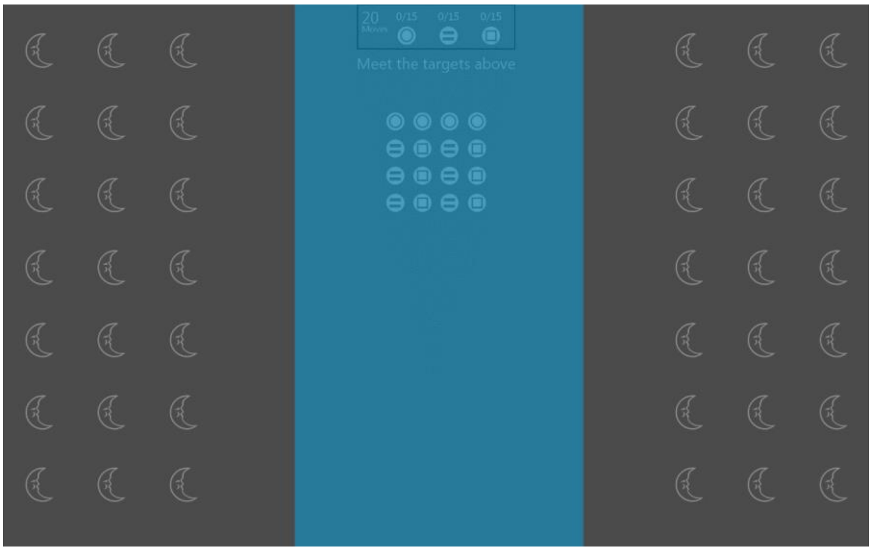

Figure 14 Game screen with distractors. The Area of Interest is shown in blue taking up the middle third of the screen. (This AOI is not visible to participants)

Eye tracking equipment divides eye movements into short fast-moving *saccades* and slower *fixations* on a particular area. The Eyelink system we were using defines a *fixation* as a period in which the eye velocity is less than 30 degrees/s and eye acceleration is less than 8000 degrees/s^2. Typically, fixations last between 50-600ms. We calculated the percentage of time that participants spent fixated on the central area rather than looking at the distraction images. The difference between conditions was not significant although there was a moderate effect size. *Full game* (M=98.42, SD=2.70), the *No goals* game (M=98.47, SD= 2.43) and the *All dots the same* game (M=96.20, SD=3.99); F(2,46)= 2.754, p=0.074, η_p^2= 0.109.

We were concerned that this analysis of gaze position may suffer from a ceiling effect, so we repeated this analysis using a smaller AOI which filled the middle quarter of the screen. The results were similar to before with no significant difference between conditions and a slightly smaller effect, *Full game* (M=97.41, SD=4.46), the *No goals* game (M=98.30, SD= 2.28) and the *All dots the same* game (M=95.35, SD=4.83); F(2,46)= 2.249, p=0.117, η_p^2= 0.091.

Discussion

In a replication of the previous experiment, participants did not recognize significantly more distractor images after the *Full game* than they would have achieved by chance. This confirms that self-paced games like *Two Dots* can hold players' attention completely and stop them being distracted by outside events. The first hypothesis that participants would remember more distractors for the game variants than the *Full game* was supported with a moderate effect size between conditions. A post hoc analysis showed a significant difference between the *Full game* and the *No goals* game. The difference between the *Full game* and the *All dots the same* game is not significant but does approaches significance. There was no difference between the two reduced variants of the game. The *No goals* game had fewer higher-level game mechanics such as levels and a score. The significant difference in attention between this and the *Full game* supports the idea that high level game mechanics are important in holding players' attention. Before carrying out the experiment we expected that the *All dots the same* game would be the least engaging so we were surprised that there was no significant difference in distracted recognized between that game and the *Full game.* However, the difference did approach significance so it is possible that with a larger sample size this would become significant and support the idea that lower level game mechanics are also important in how games hold players’ attention.

The second hypothesis that levels of immersion would be lower in the two game variants was also supported. There was again a moderate effect size between conditions which is of similar magnitude to the effect size between distractors remembered. A post-hoc analysis showed a significant difference between the *Full game* and both other games. There was no significant difference between the *No goals* and *All dots the same* game variants. These results suggest that players found both variants of the game to provide a similar level of immersion and that both higher and lower level game mechanics contribute to the experience of immersion.

The third hypothesis that eye tracking would show that participants fixated more on the distractors during the game variants was not supported although it does approach significance. There was a moderate effect size, but this was still smaller than that for distractors recognized or the IEQ. In all conditions, participants focused on the game for the majority (over 96%) of the time. We were concerned that the analysis may suffer from ceiling effects, so we repeated it with a smaller AOI which only covered the middle quarter of the screen. This reduced the proportion of time that participants were fixated on the AOI but there was a similar non-significant difference between conditions and moderate effect size between conditions. As a non-significant result, it is not possible to say whether this result would be the same with a much larger number of participants. However, in this experiment the fourth hypothesis that the effect size for the eye tracking would be higher than that of distractors recognized was not supported. Eye tracking requires an unsettling calibration and setup procedure which emphasizes that players’ eye movements are being watched by a camera which may then make participants feel watched and less likely to behave naturally as they would when playing a game. This experiment does not show evidence that these disadvantages of eye tracking are justified due it being a more sensitive way of measuring how well a game holds players’ attention than a post-game recognition test.

As in the previous experiment, a violin plot (Figure 12) of image recognition probability showed variation in the chance that a particular image would be recognized which could due to differences in image memorability or just random variation. As with Experiment 2, we looked at the top 10% of images recognized in each condition, but unlike Experiment 2, we found no overlap in the most recognized images in each condition. This suggests that randomizing the order of images and pool of images shown has successfully reduced confounds due to differences between images. Similarly, a plot of the probability of images being recognized over the time of the game (Figure 13) found no evident pattern. Boxplots (available in the supplementary materials) showed outliers in the immersion scores and eye tracking but not in the number of images recognized.

Conclusions

The goal of this paper was to introduce the DRP as a way to measure players’ attention while playing self-paced games and provide a step towards a more objective behavioral measure of player engagement which links to existing work on attention. The studies show that the DRP was sensitive to different features in games and that when games were likely to be more engaging through clearer and more restricted goals, players were less aware of the distractors. Moreover, attention as measured by the DRP supported the summative questionnaire measures.

RQ1 asked whether the DRP is an effective technique for measuring how well games hold players’ attention. Compared with previous approaches to measuring attention from Jennett (2010) and

Murch et al. (2017) the DRP shows promise as a particularly effective measure. Participants recognized over 90% of the distractors when their attention was fully on the distractors and almost none when their attention was held by playing the *Full game* of *Two Dots*. In contrast, Jennett's approach found that participants playing a fast-paced action game still recalled around 30% of the audio distractors. Murch et al found that instructing slot machine players to press a button when they saw a particular distractor resulted in participants in both groups responding to around 80% of the distractors. The DRP samples attention in 5 second "windows" which is a much finer grained measure of attention than approaches from Murch et al. (2017) which sampled every 2 minutes or Jennett (2010) which sampled every 30 seconds. Another advantage of the DRP over measures of attention from Murch et al. (2017) and Smallwood et al. (2008) is that it does not interrupt the main task of playing the game which may alter the experience which is being measured.

RQ2 asked whether the DRP could be used as an effective measure of game engagement. Attention is an aspect of the wider concept of game engagement and as a measure of attention the DRP also showed agreement with a questionnaire measure of game engagement. Results from the DRP were broadly similar to those of the IEQ although to make a reliable comparison, much larger sample sizes would be needed. However, the measures differ in their responses to different levels of engagement. The DRP had a distinct floor effect with almost no images being recognized after highly engaging experiences. This suggests that above a certain level of engagement attention is held completely. Conversely, even in the least engaging game participants still recognized fewer distractors than if there had been no game present which indicates that at low levels of engagement there may be more variation in attention and that the DRP may be a more effective measure of engagement for less engaging experiences than questionnaires such as the IEQ (Jennett et al., 2008). Unlike questionnaires the DRP has the potential to show changes in engagement over time, however this was not seen in these experiments, probably because the games used did not themselves have significant variation in engagement over time.

As the DRP is an effective measure of attention for less engaging experiences it may be a useful way of investigating more recent genres of video game which have extended the concept of game design far from the archetypal action shooter and which may have lower levels of engagement or hold attention less strongly. This would include self-paced casual games (Kultima, 2009) like *Two Dots,* but also "walking simulators" (Zimmermann & Huberts, 2019), "idle games" (Cutting et al., 2019) and interactive fiction. These experiences are also less likely to depend on factors such as the quality of controls which are examined by existing action game-oriented questionnaires. Similarly, the ability to measure low levels of engagement, via the effect of engagement on attention, may be useful when assessing the effect of individual game elements when divorced from much of the rest of the game. Previous researchers have considered the effect of differences in design of individual game elements such as power-ups (Denisova & Cook, 2019) and aesthetics (Andersen et al., 2011) as well as user interface elements such as feedback designs (Gouveia et al., 2016) in non-game contexts. There is the potential to test different variants of individual game elements within very simple interactions and then use the DRP to measure the effect that different variations of that design have on their ability to hold attention.

RQ3 asked whether eye tracking was a more effective measure of attention than a post-game distractor test. Eye tracking has been used to measure where players direct their attention during a game (e.g. Cox et al., 2006; Jönsson, 2005; Zain et al., 2011) but the DRP is a separate measure with different strengths and weaknesses. Rather than indicate *where* in the game attention is directed the DRP measures how well games *hold* attention. Eye tracking requires specialist equipment, calibration before use and may make participants feel “watched”. Experiment 3 used eye tracking to measure the difference in fixations on the game compared to the distractors but found no significant difference between games and with smaller effect sizes than the DRP. This may be an artefact of the moderate size of the sample but could also point to a more complex picture of players’ attention during games. Rensink (2015) describes several levels of attention and phenomena such as *inattentional blindness* (Mack & Rock, 1998; Most et al., 2001; Simons & Chabris, 1999) show that it is possible to be looking directly at a stimulus but not see it due to attention being focused elsewhere. It may be that eye tracking shows that participants are looking directly at the distractor images, but their attention is still on the game, so they do not remember the image afterwards. This would explain why the DRP appears to be a better measure of game attention than eye tracking. As the DRP relies on participants memorizing images it is dependent on participants giving more of their attention to the distractors whereas eye tracking only records where their gaze is directed.

Overall, then, the DRP does seem to have many of the features that we hoped in relating attention to the experience of games. It is an objective behavioral measure which both reflects the expected differences in attention between games and is aligned with existing measures of game engagement. Therefore, it may be of use as a practical measure of game engagement. Although these experiments provide evidence that it has similar results and effect sizes to questionnaire-based methods there are clearly a number of factors which make implementation more difficult. The game to be tested would need to be surrounded by distractor images, which may not be possible for game designed to fill a full landscape screen although it may be possible for mobile games by making software changes. Participants were positioned in a chin rest which reduces ecological validity although any game experiment in a laboratory setting rather than where players would actually play the game suffers from this issue. Similarly, for the DRP to be effective, participants cannot know that they will be tested on the distractors before they start the test. However, this is similar to many game experience experiments in that participants can only do the experiment once as prior knowledge of the game would change their response. In these experiments the game was played for 5 minutes which is shorter than most commercial games, but longer in length than both the crucial “onboarding” (Gaston & Cooper, 2017) initial section of a game and also a typical casual game level. Standing (1973) showed hundreds of images to participants over a period of several hours and still found a very high recognition rate, so it is likely that the DRP could be used to measure attention over much longer periods of gameplay which may also capture ebbs and flows in the level of attention over time. A possible limitation is that the DRP only shows an image every 5 seconds which limits the possible resolution of the measure to be lower than eye tracking measures. However, most game sections last for at least a few minutes this so this will not be an issue for many applications.

Attentional focus has been widely studied in the psychological literature and measuring attention in games allows links to be made between the experience of playing games and existing theories of attention. During the *Full game* recognition of distractors was no better than chance. This shows that even self-paced games can hold players’ full attention and stop them being distracted despite there being no requirement to make responses within a particular time period. It may be that the property of focusing attention is a fundamental part of the experience of being engaged in a game and that games are a particularly effective at holding participants’ attention and stopping their attention wandering. During the *Full game* participants recognized almost no distractors afterward which shows a more consistent level of attentional focus than many *inattentional blindness* experiments (e.g. Mack & Rock, 1998; Most et al., 2001; Simons & Chabris, 1999) where it is usual for at least half of the participants to notice the unexpected task-irrelevant event. For example, in

Simons and Chabris (1999)'s experiment, participants have to count the number of times a ball is passed between basketball players and many participants do not notice an unexpected person in a gorilla costume walking through the players. However, 54% of participants *did* notice the gorilla., Recognition rates in our experiment are more similar to results found in banner blindness studies. Burke et al. (2005) found that participants recognized only around 20% of banners they had been shown whilst performing another task.

Future work

There are many directions that this research could be developed. This paper describes an initial investigation into how self-paced games hold players' attention and how this can be measured using the DRP. These experiments used the matching puzzle game *Two Dots* which contains many of the same game elements as many popular "match 3" games such as *Candy Crush Saga* and *Bejewelled* so the DRP is likely to be effective with a wide variety of self-paced games. Future experiments could explore a wider range of games and other non-game interactive experiences in which measuring attention is important, such as driving a car (e.g. Kircher & Ahlstrom, 2018; Strayer et al., 2003). The experiments in this paper made use of three different measures; the DRP, the IEQ and eye tracking. However, the sample sizes used were not large enough to make detailed comparisons of the different measures or create reliable estimates of effect sizes. Repeating these experiments with much larger sample sizes would allow more robust comparisons to be made.

It is possible that some of the images used in the DRP were more memorable or distracting than others. Standing (1973) reported that "vivid" images were more likely to be remembered than others. In Experiments 2 and 3 the same images were shown in each condition so differences between images would not affect validity, but they might affect variance and reduce the sensitivity of the measure. Future experiments could investigate differences in memorability between images with no game present and then apply those findings to improve the DRP, either by weighting each image with a "memorability" factor or just by omitting images which were shown to be particularly memorable or distracting.

These experiments show that attentional focus is a key aspect of the experience of playing games which may also lead to future work which apply existing attention research to games. Attentional focus can help participants reduce negative emotions and promote positive emotional affect (Nix et al., 1995; Wadlinger & Isaacowitz, 2011). If most games hold attention as well as *Two Dots* then it would suggest that a possible motivation for game play is that it focuses players' attention onto the subject of the game which may then increase positive affect rather than letting players' attention wander into other areas may cause negative emotions. There is also the potential to investigate the effect that games have on learning through the lens of attentional focus. In the field of serious games and education there is evidence that attentional focus and avoiding split attention are important for effective learning (Ayres & Sweller, 2005; Chandler & Sweller, 1992). Considering the way that these games focus attention could lead to improved designs of serious games.

The DRP is a new way of thinking about game experience, which highlights the importance of how games hold our attention and stop us being distracted by outside events. This approach may yet reveal new insights into what happens when we play a game. We found that even a simple self-paced game like *Two Dots* can hold players' attention completely and make them "blind" to surrounding events. Self-paced casual games (Kultima, 2009) are sometimes seen as being less interesting than more complex action games. By looking through the lens of attention, these experiments show that even very simple games are still very successful in their aims to hold players' attention and give them a break from their lives.

Acknowledgements

This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Game Intelligence (IGGI) [EP/L015846/1] and the Digital Creativity Labs jointly funded by EPSRC/AHRC/InnovateUK under grant no EP/M023265/1.

Data availability

The authors confirm that the data supporting the findings of this study are available within the supplementary materials which are available from OSF at this address: https://osf.io/wza45/

Disclosure of interest

The authors report no conflict of interest

References

Alkan, S., & Cagiltay, K. (2007). Studying computer game learning experience through eye tracking. *British Journal of Educational Technology, 38*(3), 538-542.

Ambinder, M. (2011). *Biofeedback in gameplay: How valve measures physiology to enhance gaming experience*.

Andersen, E., Liu, Y.-E., Snider, R., Szeto, R., & Popović, Z. (2011). *Placing a value on aesthetics in online casual games.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., De Freitas, S., Louchart, S., . . . De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology, 46*(2), 391-411.

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. *The Cambridge handbook of multimedia learning, 2*, 135-146.

Baranowski, T., Buday, R., Thompson, D. I., & Baranowski, J. (2008). Playing for real: video games and stories for health-related behavior change. *American journal of preventive medicine, 34*(1), 74-82. e10.

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior, 28*(3), 771-780.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325-14329.

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.

Brown, E., & Cairns, P. (2004). *A grounded investigation of game immersion.* Paper presented at the CHI'04 extended abstracts on Human factors in computing systems.

Burke, M., Hornof, A., Nilsen, E., & Gorman, N. (2005). High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(4), 423-445.

Cairns, P. (2016). Engagement in digital games. In *Why Engagement Matters* (pp. 81-104): Springer.

Calleja, G. (2007). Digital game involvement a conceptual model. *Games and culture, 2*(3), 236-260.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision research, 51*(13), 1484-1525.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*(2), 233-246.
Chatfield, T. (2009, Sunday 27th Sept). Videogames now outperform Hollywood movies. *The Observer*.
Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM, 50*(4), 31-34.
Cox, A. L., Cairns, P., Berthouze, N., & Jennett, C. (2006). *The use of eyetracking for measuring immersion.* Paper presented at the Workshop on What have eye movements told us so far, and what is next.
Crook, J. (2014). Two Dots, The Sequel To Betaworks' Dots, Is A Beautiful Monster. Retrieved from http://techcrunch.com/2014/05/31/two-dots-the-sequel-to-betaworks-dots-is-a-beautiful-monster/
Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience* (Vol. 41): HarperPerennial New York.
Csikszentmihalyi, M. (2013). *Flow: The psychology of happiness*: Random House.
Cutting, J., Gundry, D., & Cairns, P. (2019). Busy doing nothing? What do players do in idle games? *International journal of human-computer studies, 122*, 133-144.
Denisova, A., & Cook, E. (2019). *Power-Ups in Digital Games: The Rewarding Effect of Phantom Game Elements on Player Experience.* Paper presented at the Proceedings of the Annual Symposium on Computer-Human Interaction in Play.
Denisova, A., Guckelsberger, C., & Zendle, D. (2017). *Challenge in Digital Games: Towards Developing a Measurement Tool.* Paper presented at the Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.
Denisova, A., Nordin, A. I., & Cairns, P. (2016). *The convergence of player experience questionnaires.* Paper presented at the Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play.
Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience, 18*(1), 193-222.
Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision research, 36*(12), 1827-1837.
Etchells, P. (2019). *Lost in a Good Game: Why we play video games and what they can do for us*: Icon Books.
Fine, S. (2015). Two Dots, Too Charming. *Gamasutra*. Retrieved from http://gamasutra.com/blogs/ScottFine/20151118/259618/Two_Dots_Too_Charming.php
Gaston, J., & Cooper, S. (2017). *To three or not to three: Improving human computation game onboarding with a three-star system.* Paper presented at the Proceedings of the 2017 CHI conference on Human Factors in Computing Systems.
Gouveia, R., Pereira, F., Karapanos, E., Munson, S. A., & Hassenzahl, M. (2016). *Exploring the design space of glanceable feedback for physical activity trackers.* Paper presented at the Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.
Gutwin, C., Rooke, C., Cockburn, A., Mandryk, R. L., & Lafreniere, B. (2016). *Peak-end effects on player experience in casual games.* Paper presented at the Proceedings of the 2016 CHI conference on human factors in computing systems.
Harmat, L., de Manzano, Ö., Theorell, T., Högman, L., Fischer, H., & Ullén, F. (2015). Physiological correlates of the flow experience during computer game playing. *International Journal of Psychophysiology, 97*(1), 1-7.
Hervet, G., Guérard, K., Tremblay, S., & Chtourou, M. S. (2011). Is banner blindness genuine? Eye tracking internet text advertising. *Applied Cognitive Psychology, 25*(5), 708-716.
Izzetoglu, K., Bunce, S., Izzetoglu, M., Onaral, B., & Pourrezaei, K. (2003). *fNIR spectroscopy as a measure of cognitive task load.* Paper presented at the Proceedings of the 25th Annual

International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439).
Jennett, C. (2010). *Is game immersion just another form of selective attention? An empirical investigation of real world dissociation in computer game immersion.* UCL (University College London),
Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies, 66*(9), 641-661.
Johansen, S. A., Noergaard, M., & Rau, J. (2008). *Can eye tracking boost usability evaluation of computer games.* Paper presented at the Proceedings of CHI.
Jönsson, E. (2005). If looks could kill–an evaluation of eye tracking in computer games. *Unpublished Master's Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden*.
Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological science, 4*(6), 401-405.
Kaye, L. K., Monk, R. L., Wall, H. J., Hamlin, I., & Qureshi, A. W. (2018). The effect of flow and context on in-vivo positive mood in digital gaming. *International journal of human-computer studies, 110*, 45-52.
Kinoshita, S. (1995). The word frequency effect in recognition memory versus repetition priming. *Memory & Cognition, 23*(5), 569-580.
Kircher, K., & Ahlstrom, C. (2018). Evaluation of methods for the assessment of attention while driving. *Accident Analysis & Prevention, 114*, 40-47.
Kuhn, G., Teszka, R., Tenaw, N., & Kingstone, A. (2016). Don't be fooled! Attentional responses to social cues in a face-to-face and video magic trick reveals greater top-down control for overt than covert attention. *Cognition, 146*, 136-142.
Kultima, A. (2009). *Casual game design values.* Paper presented at the Proceedings of the 13th international MindTrek conference: Everyday life in the ubiquitous era.
Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences, 9*(2), 75-82.
Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1641), 20130205.
Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*(3), 339.
Mack, A., & Rock, I. (1998). *Inattentional blindness*: MIT press.
Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International journal of human-computer studies, 65*(4), 329-347.
McCarthy, J., & Wright, P. (2004). Technology as experience. *interactions, 11*(5), 42-43.
Mekler, E. D., Bopp, J. A., Tuch, A. N., & Opwis, K. (2014). *A systematic review of quantitative studies on the enjoyment of digital entertainment games.* Paper presented at the Proceedings of the 32nd annual ACM conference on Human factors in computing systems.
Miller, J. D., & Tanis, D. C. (1971). Recognition memory for common sounds. *Psychonomic Science, 23*(4), 307-308.
Most, S. B., Simons, D. J., Scholl, B. J., Jimenez, R., Clifford, E., & Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological science, 12*(1), 9-17.
Murch, W. S., Chu, S. W., & Clark, L. J. P. o. A. B. (2017). Measuring the slot machine zone with attentional dual tasks and respiratory sinus arrhythmia. *31*(3), 375.
Nacke, L., & Lindley, C. A. (2008). *Flow and immersion in first-person shooters: measuring the player's gameplay experience.* Paper presented at the Proceedings of the 2008 Conference on Future Play: Research, Play, Share.

Nix, G., Watson, C., Pyszczynski, T., & Greenberg, J. (1995). Reducing depressive affect through external focus of attention. *Journal of Social and Clinical Psychology, 14*(1), 36-52.

Phillips, B. (2006). Talking about games experiences: A view from the trenches. *interactions, 13*(5), 22-23.

Power, C., Cairns, P., Denisova, A., Papaioannou, T., & Gultrom, R. (2018). Lost at the Edge of Uncertainty: Measuring Player Uncertainty in Digital Games. *International Journal of Human–Computer Interaction*, 1-13.

Rensink, R. A. (2015). A function-centered taxonomy of visual attention. *Phenomenal Qualities: Sense, Perception, and Consciousness*, 31.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological science, 8*(5), 368-373.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 344-360.

Schell, J. (2008). *The Art of Game Design: A book of lenses*: CRC Press.

Schull, N. D. (2005). Digital gambling: The coincidence of desire and design. *The Annals of the American Academy of Political and Social Science, 597*(1), 65-81.

Schull, N. D. (2012). *Addiction by design: Machine gambling in Las Vegas*: Princeton University Press.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *The Quarterly journal of experimental psychology, 38*(3), 475-491.

Sicart, M. (2008). Defining game mechanics. *Game Studies, 8*(2), n.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7*(1-3), 1-15.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception-London, 28*(9), 1059-1074.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in cognitive sciences, 1*(7), 261-267.

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition, 36*(6), 1144-1150.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly journal of experimental psychology, 25*(2), 207-222.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*(1), 23.

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games: An overview. In: Institutionen för kommunikation och information.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *nature, 381*(6582), 520.

Wadlinger, H. A., & Isaacowitz, D. M. (2011). Fixing our focus: Training attention to regulate emotion. *Personality and Social Psychology Review, 15*(1), 75-102.

Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided search. *The Oxford handbook of attention, 11*, 35-44.

Wood, K., & Simons, D. J. (2019). Processing without noticing in inattentional blindness: A replication of Moore and Egeth (1997) and Mack and Rock (1998). *Attention, Perception, & Psychophysics, 81*(1), 1-11.

Zain, N. H. M., Razak, F. H. A., Jaafar, A., & Zulkipli, M. F. (2011). *Eye tracking in educational games environment: evaluating user interface design through eye tracking patterns.* Paper presented at the International Visual Informatics Conference.

Zimmermann, F., & Huberts, C. (2019). From Walking Simulator to Ambience Action Game. *Press Start, 5*(2), 29-50.