Entertainment Computing 2 (2011) 3–9

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: ees.elsevier.com/entcom

Considerations for the use of commercial video games in controlled experiments

Ryan P. McMahan *, Eric D. Ragan, Anamary Leal, Robert J. Beaton, Doug A. Bowman

Center for Human–Computer Interaction and Department of Computer Science, Virginia Tech, 2202 Kraft Drive, KWII Building (0106), Blacksburg, VA 24061-0106, USA

ARTICLE INFO

Article history:
Received 10 October 2010
Revised 17 January 2011
Accepted 3 March 2011
Available online 15 March 2011

Keywords:
Commercial video game
Controlled experiment
Ecological validity

ABSTRACT

While commercial, off-the-shelf video games have been used often to observe cognitive and social phenomena, few studies have taken advantage of such games as research tools for conducting controlled experiments. Providing evidence that unmodified, commercial video games can be used to conduct gaming evaluations with high levels of both experimental control and ecological validity, we designed and conducted an experiment that utilized *Mario Kart Wii* to evaluate the effects of natural interaction on player performance. Based on our experience from that experiment, we present several concerns that require attention when using commercial video games as research tools. Providing examples of design decisions and outcomes from our experiment, we identify some of the benefits, drawbacks, and challenges of using such tools.

1. Introduction

As research tools, commercial video games (i.e., unmodified, off-the-shelf titles [1]) have a history of being used to study phenomena such as factual learning [2], skill development [3], aggressive behavior [4], and cultural awareness [5]. This history is undoubtedly linked to the ability to use these games as cheap, immediate, and effective tools [6]. Even so, commercial video games have not been used often as tools for controlled experiments. In a controlled experiment, controlled groups and methods (i.e., independent variables) are manipulated to create different experimental conditions, in which the same observational methods are used to determine the causality of the independent variables on particular factors of concern (i.e., dependent variables) [7]. The ability to isolate and control each independent variable independently is necessary to avoid confounds. Confounds are uncontrolled characteristics across the different experimental conditions that correlate (positively or negatively) with both an independent variable and a dependent variable [7].

In general, with unmodified, off-the-shelf titles, it is difficult to isolate independent variables and address specific research questions. Yet, some commercial video games can be utilized to conduct gaming studies that have high levels of control, in addition to high levels of *ecological validity* – a rare combination in research. Ecological validity refers to how well the methods, materials, and setting of a study approximate the real-life situation under investigation [8], and it is normally difficult to achieve a high level of control without reducing the experiment's resemblance to real life. As an example of a highly controlled and ecologically valid experiment, we used *Mario Kart Wii*, a commercial racing game for the Nintendo Wii, to evaluate the effects of natural interaction on player performance [9].

The goal of our controlled experiment was to obtain a better understanding of natural interaction techniques, which mimic real-world interaction by using body movements or actions that are similar to those used for the same task in the physical world. Considering the trend started by the Nintendo Wii gaming console for producing realistic experiences through natural interaction techniques, we specifically were concerned with how natural interaction (e.g., swinging one's real arm to swing a virtual tennis racket) and non-natural interaction (e.g., pressing a button to swing a virtual racket) would affect the player experience. We particularly wanted to compare the effects of approximately equivalent natural and non-natural interaction techniques on player performance, since performance influences enjoyment [10] and anxiety [11].

We wished to use a commercial video game for our experiment to help establish a higher level of ecological validity and to avoid implementing our own game software. We knew that the game would need to support our research goals by providing the ability to directly compare natural and non-natural interaction techniques that were seemingly balanced. After surveying many commercial titles, we finally chose *Mario Kart Wii* as our research tool. This racing game provided well-designed and approximately equivalent natural and non-natural interaction techniques, which were essential for our research. By using the game's "Time Trial" mode, which eliminated variance due to extraneous game features such as artificial intelligence, we were able to perform a controlled evaluation and objective comparison of natural and non-natural interaction techniques in a moderately ecologically valid context – attempting

* Corresponding author. Tel.: +1 540 230 6634.
E-mail address: rymcmaha@vt.edu (R.P. McMahan).

1875-9521/$ - see front matter © 2011 International Federation for Information Processing Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.entcom.2011.03.002

to break course records. The analyses of our results demonstrated that non-natural interaction techniques outperformed the natural techniques, indicating players should benefit from using the non-natural controllers for *Mario Kart Wii* [9].

Though we were able to successfully use a commercial video game to conduct an evaluation with high levels of experimental control and ecological validity, we had to address several concerns during our experimental design and analysis to do so. Based on those experiences, in this paper, we present some of the issues that should be addressed when using unmodified, commercial video games for similar controlled evaluations and frame those issues with explanations of the decisions we made for our natural interaction study. We then conclude with a summary of advantages, disadvantages, and challenges of using commercial video games for controlled experiments.

2. Related work

Commercial video games have been used as research tools in various studies. One of the earliest examples is Bowman's analysis of *Pac-Man* in order to provide suggestions for classroom motivation and instruction [12]. Since then, commercial video games have served as research tools for many studies on learning, skill development, and training. Another early example was the use of Atari's *Army Battlezone* as a tactical trainer for the United States Army [13]. More recently, simulation-based commercial video games, such as *SimCity* and *Roller Coaster Tycoon*, have been used for teaching economics (managing city resources) and physics (building a rollercoaster) in classrooms [14]. Similarly, Squire and Barab have investigated using *Civilization III* for engaging students in learning world history [2].

Aside from learning, commercial video games have been used to observe psychological and sociological behaviors. Kolo and Baur observed and compared the online and offline behaviors of *Ultima Online* players to better understand the psychology and sociology of massively multiplayer online gamers, determining that these individuals are not anti-social but actively seek interactions with others [15]. Similarly, Smith observed *Dance Dance Revolution* players and examined how the popular game spawned a social culture of ardent players [5]. Schulzke analyzed *Fallout 3* and how it reinforced moral decision-making by providing opportunities and consequences for moral and immoral actions [16]. In a similar analysis, Paulk observed how *The Sims* promoted the sociology of interior design principles, such as positioning a bed for access from both sides for a couple [17].

Commercial video games have also been used to study user experience through observation. One example is the research conducted by Larssen et al. for understanding movement as input for interaction [18]. By using *Beat Freak* and *Kung Foo*, two Eyetoy games for the Playstation 2, Larssen et al. analyzed two frameworks for investigating the relationships between bodily actions and corresponding interactions. The researchers also determined that natural interaction techniques should be sensible to players in terms of expectations and effort expanded, based on their observations.

Despite many research studies involving commercial video games, few controlled experiments have utilized commercial, off-the-shelf video games. One rare example is an early study conducted by Silvern and Williamson in which the researchers evaluate the influence of *Space Invaders* on children compared to baseline aggressive behaviors [4]. They determined that the commercial video game increased aggressive behavior in children after being played. They also found similar results for children watching a violent cartoon.

We suspect that the lack of research utilizing commercial video games for controlled experiments stems from the fact that most commercial video games have too much variability and rely too much on player strategy to answer the types of research questions typical of controlled experiments. Some researchers may dismiss the concept of using a commercial video game for a controlled evaluation with the idea that a modified game engine or custom software would better serve their research needs. Though such custom tools are probably better in many cases, some researchers could benefit from first investigating commercial video games as cheap, immediate, and effective tools with high levels of ecological validity [6]. In the following sections, we present the advantages, disadvantages, and concerns researchers should consider while investigating the use of a commercial video game for a controlled experiment.

3. Inherent advantages and disadvantages

Before discussing some of the experimental design concerns of using a commercial video game in a controlled experiment, we present the general implications of using such research tools. We divide these into advantages and disadvantages.

3.1. Advantages

A key advantage of using a commercial, off-the-shelf video game for a controlled experiment is ecological validity. As mentioned in the introduction, it is rare to achieve a high level of experimental control and ecological validity within the same study, since each aspect is often traded for the other. For research on gaming, commercial video games can offer some degree of ecological validity since there should be some population of game players represented in the research. In a few cases, like our natural interaction study, researchers can use these games to conduct controlled experiments that have a high level of ecological validity. These experiments should be reflections of real life. For instance, if a controlled research study was to conclude that most people play first-person shooters better with keyboards and mice than game controllers, players with keyboards and mice should outperform players with game controllers in real-life situations.

Another easily discernible advantage of using commercial video games in controlled experiments is the lack of implementation time. Because commercial video games and their accompanying hardware are already produced and available for purchase, there is no need to implement either software or hardware for a study, if these tools can be utilized. Instead, researchers can progress directly to conducting their studies and analyzing results, which is the goal of the research (not implementing software or hardware).

Another advantage, related directly to the fact that commercial video games are finalized products, is the lack of external influence on implementation. Though researchers are supposed to strive to design and implement controlled experiments that are fair and balanced, sometimes external influences can impact those processes to produce a biased evaluation and biased results. For example, in our experience, researchers may spend more time or effort implementing a condition that they subconsciously (or consciously) favor, biasing the study toward that condition. By using commercial video games, researchers are unable to influence the implementation of their research tools and instead depend on the motivations of the game developers (who are trying to produce a marketable product rather than enable experimental conditions).

A fourth advantage of using a commercial video game as a research tool is *study reproducibility*. We use the term study reproducibility to refer to the ability of other researchers to repeat a given controlled experiment and reproduce the results. While not guaranteed, reproducibility is much more likely when using a commercial video game because other researchers should have easy access to the same game and gaming hardware. This is in stark

contrast to some studies that involve custom software and custom hardware. The publications describing such experiments often do not or cannot provide enough detail to allow the study to be replicated.

3.2. Disadvantages

Despite the advantages of using commercial video games as research tools, researchers should be aware of the disadvantages of this approach. Most of these disadvantages stem from the fact that commercial video games are finalized products, and therefore modifications to the software are essentially impossible. A resulting, major disadvantage is the inability to eliminate potential confounds that are inherent to the game software for answering the research questions. Essentially, confounds make it difficult to confirm causality based on only the independent variable. If a game has an inherent confound, such as random artificial intelligence, it may not be possible to eliminate the confound with software modification (unless the game is open source). In a similar manner, researchers may want to add features to a game to balance unbalanced features or to create a provision for evaluating a particular independent variable. Again, without the ability to modify the game, these options are unavailable.

Other disadvantages can be attributed to the specificity of commercial hardware. Most commercial games are designed to run on specific consoles that have accompanying hardware as controllers and sensors. Most of these commercial hardware devices are perfectly sufficient for entertainment purposes, but can lack the capabilities and precision of industrial-quality controllers and sensors. For example, the orientation tracking of the Wii Remote (even with the Wii MotionPlus attachment) is inferior to the orientation tracking of a professional tracking system such as the Intersense IS-900. Results of experiments based on such consumer-level devices may be skewed or false if generalized beyond the specific gaming hardware used.

Another disadvantage is the challenge of finding a suitable game. Today, there are many commercial video games available for a variety of console systems and personal computers. Due to the sheer quantity of available titles, it may take researchers a while to discover an appropriate game for their research. Worse, despite the large number of available games, an appropriate commercial video game may not exist for a given research question. We address these types of concerns in Section 4.1 on game selection.

4. Experimental design concerns

In this section, we identify and discuss the concerns researchers should be aware of during the experimental design process when using commercial video games for controlled experiments. We describe how we addressed these concerns for our controlled evaluation of natural interaction techniques and provide some details of our experimental design.

4.1. Game selection

The most important and obvious concern when using a commercial video game for a controlled experiment is deciding which game to use. The game selected must provide the ability to answer the research questions motivating the study. Consider the study conducted by Silvern and Williamson [4]. The researchers were concerned with the influence of violent entertainment on the aggressive behavior of children. At the time of the study, the researchers deemed *Space Invaders* to be one of the more violent commercial video games, due to the constant firing of laser beams at alien invaders. This aspect of *Space Invaders* provided Silvern and Williamson with another form of violent entertainment, in addition to a cartoon, to evaluate the influence of violent entertainment on aggressive behavior.

For our research on natural interaction techniques and how they affect player performance in comparison to non-natural interaction techniques, we had to determine if any commercial games provided both styles of interaction. At the time of our experimental design, the Nintendo Wii gaming console was already popular for the offering of games that involved bodily interactions (e.g., *Wii Sports*). Unfortunately, upon surveying the many commercial Wii titles available, we quickly found that most games did not provide the ability to compare natural and non-natural interaction techniques, because they offered only one style of interaction for completing game actions. Some of the video games we surveyed did offer both 3D and 2D interaction techniques for completing the same game actions, but we viewed many of these 3D interactions as poor substitutes for truly natural interactions (e.g., shaking the Wii Remote to get up after a boxing knockdown). After surveying many available Wii titles, we did eventually identify two commercial video games as potential research tools for our study.

The first of these was *Mercury Meltdown Revolution*, a puzzle game similar to traditional marble tilt-mazes. This commercial video game provides both natural and non-natural interaction techniques for tilting various 3D environments to manipulate and move a blob of mercury through a series of maze-like obstacles. For the natural interaction technique, a player holds a Wii Remote flat with both hands and tilts the remote to tilt the 3D environment in the same orientation, ultimately causing the mercury blob to move in the direction of the tilt. With the non-natural interaction technique, the player uses a Wii Classic controller to tilt the 3D environment by moving the joystick in the same direction.

Despite the fact that *Mercury Meltdown Revolution* met our research requirement of being able to evaluate both natural and non-natural interaction, we quickly determined that it was an inappropriate tool for our research. The basis for our determination was that the underlying game task, tilting an environment to manipulate an object, is not a familiar real-world task for most people. Though *Mercury Meltdown Revolution* was well-designed to mimic the physical actions of tilting an environment with the Wii Remote technique, the physical actions were unfamiliar to most players. In our experience with the game, we found it was much easier to complete the obstacles with the non-natural joystick technique due to our prior experiences of using joysticks to manipulate environments. This led us to realize that researchers must also consider the underlying game task in regard to their research questions. For our study, we required a game that provided both natural and non-natural interaction techniques for completing a *familiar* real-world task to truly differentiate the effects of natural interaction.

The concern of an appropriate game task brought us to our second potential research tool, *Mario Kart Wii*. This commercial racing game provides both natural and non-natural interaction techniques that players can use to drive a vehicle around various racing courses. Since driving is a common and familiar real-world task for many adults, we felt this was an appropriate game task for our research. In addition to providing the ability to directly compare natural and non-natural interaction techniques, we were pleased to discover that these techniques were well designed and approximately equivalent, leading to our decision to use *Mario Kart Wii* for our research study.

There are five interaction techniques that players can use to drive a vehicle in *Mario Kart Wii*. The most natural technique involves the use of the Wii Wheel accessory (a passive prop) with the Wii Remote to simulate using a steering wheel (see Fig. 1). For this natural technique, players hold and turn the Wii Wheel to

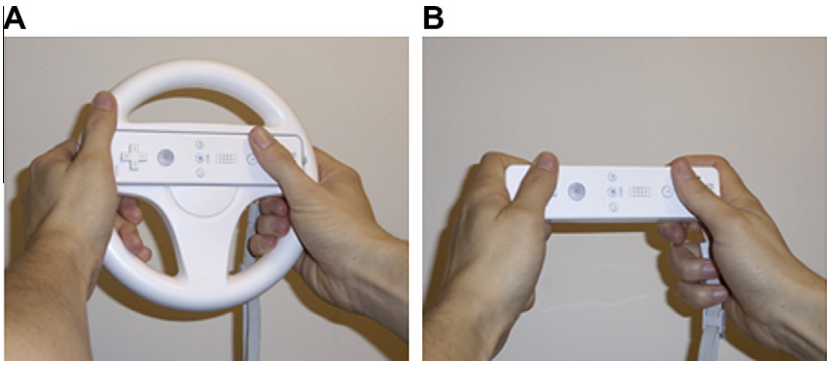

Fig. 1. (A) The Wii Wheel device and (B) the Wii Remote device.

steer. A second natural interaction technique, which we refer to as the Remote technique, works exactly the same as the Wheel technique except that the Wii Wheel is not used. The only difference is that the physical form factor of the underlying devices differ (i.e., players hold and turn a box-shaped Wii Remote as opposed to a wheel-shaped prop).

Two of the remaining interaction techniques are based on traditional schemes from prior versions of Mario Kart released on older consoles. One of these non-natural techniques uses a Wii Classic controller while the other uses a Nintendo GameCube controller (see Fig. 2). For both techniques, steering is achieved by pushing a spring-loaded analog joystick left or right. Like the Wheel and Remote techniques, the Classic and GameCube techniques are functionally the same but differ in physical form factor.

The last interaction technique provided by Mario Kart Wii is what we refer to as the Nunchuk technique. It uses the Wii Nunchuk connected to the Wii Remote. Similar to the Classic and GameCube techniques, the analog joystick on the Wii Nunchuk is used for steering, yet some game actions are performed by shaking the Wii Remote, like the Wheel and Remote techniques.

For our experimental design, we decided to evaluate the Classic, GameCube, Remote, and Wheel techniques (i.e., our independent variables), omitting the Nunchuck technique. We chose to evaluate both the Wheel and Remote as natural interaction techniques to learn more about the effects of form factor for natural interaction techniques. For a similar reason, we chose to evaluate both the Classic and GameCube techniques. We chose not to evaluate the Nunchuk technique due to its hybrid qualities.

4.2. Mode selection

Researchers intending to use commercial video games should also be aware of mode selection as an important design concern. Though some games still only offer one mode of play, many games today offer at least modes for single-player storyline and multiplayer competition. Some more fully featured commercial titles actually offer several other modes of game play to select from (e.g., single player sandbox, multi-player cooperation, online competition). Our research tool, *Mario Kart Wii*, provided four different game modes to select from: "Grand Prix," "Time Trial," "VS" (i.e., versus), and "Battle."

As with game selection, mode selection should be a careful process, as the mode selected should maintain the ability to answer the research questions without introducing other factors or confounds into the research. In our case, we immediately eliminated "Battle" as a potential mode to use for our research comparing natural and non-natural interaction. In this mode, the game task changes from racing a vehicle around a particular course to either attempting to pop balloons attached to other drivers' vehicles or running over coins to collect the most. As with game selection, we were concerned that a mode with an unfamiliar task could inherently affect our comparison of natural and non-natural interaction.

Another key consideration in selecting a game mode for research is how to balance experimental control and ecological validity. The balance should depend on the nature of the research questions motivating the study. If the questions are meant to provide an indication of how people interact with their environments, the ecological validity of the study will be more important than the experimental control. If the research is more concerned with isolating variables of influence, then experimental control is more important than the ecological validity.

For our study of natural interaction, we desired to conduct a more controlled evaluation of the interaction techniques provided by *Mario Kart Wii*. We wanted to eliminate external factors such as artificial intelligence, competitive wrecks, and randomly generated

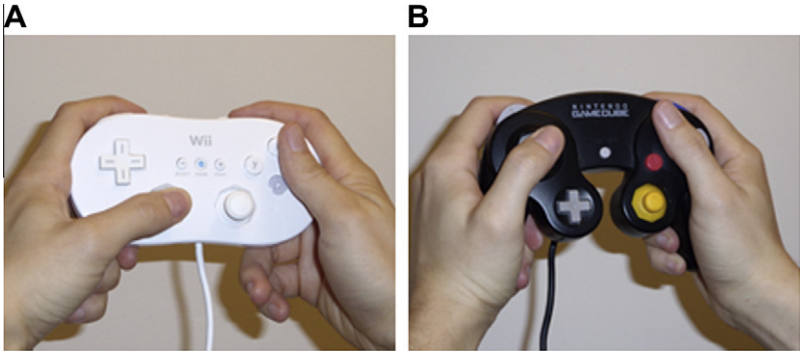

Fig. 2. (A) The Wii Classic controller and (B) the Nintendo GameCube controller.

effects. Based on this, we eliminated the "Grand Prix" mode due to the variances in the artificial intelligence of the other drivers and the randomly generated items that affect performance (e.g., homing red turtle shells that cause drivers to wreck). For similar reasons, we eliminated the "VS" mode, in which the performance of other players would affect the performance of the experiment participant. Although these modes are probably the most popular, and therefore would result in high levels of ecological validity, they did not provide enough control.

The remaining game mode was "Time Trial," in which a single player races a course alone and completes three laps in an attempt to break existing time records. Interestingly, the "Time Trial" mode was nearly a perfect match for our research. Aside from three mushroom items that provide short speed boosts when used, the game mode is a very controlled setting void of artificial intelligence and random effects. The "Time Trial" mode reports player performance metrics in the form of individual lap times and the total time to complete the three laps (i.e., our dependent variables). This meant that our primary objective metrics were perfectly accurate and we did not have to measure them ourselves (though we had to collect them). Additionally, the game mode could be played with any of the available character drivers, their associated vehicles, and on any of the courses. We decided to use an easy course and a difficult course to determine if there were any differences within the interaction techniques based on the level of game difficulty. We selected "Luigi Circuit," which offers no hazards, as the easy course, and "Wario's Gold Mine," which includes hazards (e.g., falling off the track and crashing into mining carts), as the difficult course.

4.3. Managing confounds

Despite making careful decisions about game selection and mode selection, researchers that use commercial video games for controlled evaluations should still be aware of potential confounds and plan accordingly. For instance, in our study, despite using the controlled "Time Trial" mode, there were still confounds based on when players used their three mushrooms and if they committed any racing errors (i.e., directional errors, fall errors, and crash errors). To account for these potential confounds, we developed distinct guidelines for identifying and tracking these occurrences per lap. We deemed the usage of the mushrooms as a potential independent variable and later analyzed the variable to determine that mushroom usage had no significant effect on our results. We also decided to utilize the racing errors as additional objective metrics since these errors are a direct result of the player's ability to use the interaction techniques.

Another consideration in managing confounds returns to the issue of balancing experimental control and ecological validity (as discussed above). For our study, we had to decide whether participants would be assigned a driver and vehicle (for greater control) or if the participants would be able to select any driver and vehicle they desired (for greater ecological validity). After careful consideration and reflection on our own experiences with *Mario Kart Wii*, we opted to allow selection of any character but limited selection to "karts" only (forbidding the selection of "bikes"). Our reasoning for this was to avoid frustrating participants experienced with *Mario Kart Wii* by forcing them to use an unfamiliar driver and "kart" combination (as opposed to their favorite). We only forbade "bikes" due to their additional feature of "popping a wheelie" to gain speed (something not possible with the "karts"). We later analyzed the selection of drivers and vehicles made by our participants and determined that there was not a significant effect of this selection option on our results.

Though we only used observational and post-analysis techniques for managing confounds in our study, there are alternative methods researchers can use. Participant instruction can be used to inform participants to avoid certain actions that cause confounds. For instance, we could have instructed our participants not to use any of the three mushrooms, hence eliminating the potential confound of when they use them. Another potential method is hardware alteration. We could have removed or disabled the buttons on our study controllers that activated the mushrooms to avoid the same issue. We suspect for other types of research similar hardware alterations could be used to eliminate other types of confounds. A third potential method for managing confounds is eliminating confounded trials from analysis. For example, we could have ignored trials that involved racing errors, but this would be less than desirable due to the frequency of such errors and the fact that they are a direct result of the player's ability to use an interaction technique.

4.4. Participants

Researchers using commercial video games should definitely consider the backgrounds of their participants. A major concern is whether a participant has prior experience playing the game used as the research tool. Participants with prior game experience will typically excel and perform better than participants who have never experienced the commercial video game before. This can affect both the enjoyment and anxiety of participating in the research study. Additionally, if the research questions depend on a baseline of experience, participants may require training before the experiment, particularly if they have no prior experience with the game. Therefore, it is important that the researchers account for participants' experiences in order to prevent additional confounds.

Experience with a game may also influence a participant's attitude, goal, and motivation. For example, if participants do not like particular aspects of a game used, their willingness to perform to the best of their abilities may decline. Similarly, since games are meant to be fun, it can be difficult to predict how participants' mindsets may change during the course of an experiment. Participants may be more interested in having fun with the game than in completing the experimental tasks exactly as needed for the controlled study. From our observations and formal analyses, we do not believe that this occurred in our study, but the possibility was certainly a concern.

There are at least two approaches researchers can take to account for participant experience and attributes. The first approach is to actively screen (include or exclude) participants based on their experience with the research tool or other selection criteria (e.g., gender or age group). Another approach is to passively collect data about the participants, using a background survey. This does not guarantee a balance of experienced and inexperienced players, but it does provide the means to conduct analyses based on their experience. This is how we accounted for participant experience in our *Mario Kart Wii* study (as we discuss in Section 5.1).

Another issue relating to participants and the use of commercial video games is recruitment. For research studies that rely on voluntary participation, commercial video games can be a great attractor and motivator. This is especially true for studies that involve a newly released title or a popular classic. But researchers should also be aware of the negative influence a commercial video game can have on recruitment. If a particular game did poorly in sales and received bad reviews, a research study utilizing the game may be hindered by the inability to recruit participants due to the bad image of the game.

5. Post-study issues

After designing and conducting an experiment that utilizes a commercial video game, there are still concerns that should be addressed post-study. These issues involve the analysis of the results and follow-up research studies.

5.1. Analysis

The first concern during the analysis of a commercial video game study is to ensure that confounds were properly managed by checking for any significant effects of those confounds. For our study, we were lucky to find no significant effects of mushroom usage or driver selection on our results. We did find a direct link between racing errors and lap times, but we accepted this, as racing errors are a reflection of a player's ability to use the interaction technique, so these results contributed to our understanding of the overall research question.

Another concern researchers should consider again during analysis is participant experience. In our first analysis of lap times on the easy course, we did not find a significant difference between the natural and non-natural interaction techniques. We then conducted the analysis again after removing inexperienced participants and determined that there was a significant difference between the two styles of interaction for experienced participants. Had we not accounted for participant experience with our experimental design, we would not have found this interesting result.

Aside from statistical analysis, researchers also need to consider the implications of using a commercial video game when they try to explain their results. For example, in our study, we determined that the non-natural interaction techniques significantly outperformed the natural techniques overall. In our publication [9], we went on discuss the potential reasons for this phenomena and provided three separate hypotheses. One of these hypotheses – latency – was a direct result of considering the implications of using *Mario Kart Wii* for our research tool. We felt that the natural interaction techniques may have suffered from the consumer-level hardware used for determining steering motions while the non-natural techniques relied on the values reported by the controller joysticks. Had we been able to substitute higher-quality sensors for the natural interaction techniques, this potential latency would not have been an issue.

5.2. Follow-up studies

As mentioned in Section 3.1, reproducibility is an advantage of using commercial video games for research. Because other researchers have access to the same software and hardware implementation, reproducibility for these studies relies on researchers reporting enough details to conduct the experiments again. When publishing study results, researchers should take this into account and provide enough information so that others can reproduce the study. This provides the opportunity for studies to be repeated and for results to be confirmed (or debated).

Aside from reproducibility, another follow-up consideration is the ability to quickly redesign controlled evaluations that failed or extend evaluations to account for additional independent or dependent variables. Consider for example a hypothetical study evaluating the effects of terrain (or environment) on the capabilities of first-person shooter players. Suppose that in an initial study, the researchers find no effect of terrain on the performance metrics, but from their observations, realize that they chose game levels that were not contrasting enough. By selecting new levels, the researchers could quickly re-run the same experiment and possibly find a significant effect of terrain.

6. Summary and conclusions

In this paper, we have presented several concerns and issues when using an unmodified, commercial video game as a tool for a controlled experiment. We have divided these into advantages, disadvantages, and challenges.

6.1. Advantages

- A gaming study should have some degree of ecological validity since there should be some population that plays the game.
- Researchers can progress directly to conducting their trials and analyzing results without the need to implement software or hardware.
- Researchers will not subconsciously (or consciously) influence the results of a study through implementation details.
- Other researchers should be able to reproduce the study, as they will have access to the same software and hardware.
- Failed evaluations or overlooked variables can be quickly remedied by conducting new studies with the existing game.

6.2. Disadvantages

- Confounds that are inherent to the game cannot be eliminated through software modifications.
- New features or conditions cannot be added through software modifications.
- Consumer-level hardware devices may not be as precise or capable as professional-quality devices.
- An appropriate commercial video game may not exist for a particular research study.

6.3. Challenges

- It may be difficult to select a commercial video game that provides the ability to answer the research questions motivating the study.
- Researchers need to choose a game mode that properly balances experimental control and ecological validity.
- Confounds must be managed through external methods (e.g., observational analysis or participant instruction).
- Experimenters must properly account for participant experience with the game.

In addition to presenting some of the concerns of using commercial video games for research, we have motivated and shown that these games can be used, in certain situations, to conduct studies that have both a high level of experimental control and a high level of ecological validity. Normally, these aspects of research compete, and one is traded off for the other. By carefully considering how to design a controlled experiment using a commercial video game, it is sometimes possible to achieve high levels of both. As an example, we designed and conducted a controlled experiment that evaluated the effects of natural interaction on player performance in an ecologically valid context by utilizing *Mario Kart Wii* and its "Time Trial" mode.

References

[1] R. Stanford, Teaching with games: COTS games in the classroom, in: JISC Innovating e-Learning 2006: Transforming Learning Experiences Online Conference, 2006.
[2] K. Squire, S. Barab, Replaying history: engaging urban underserved students in learning world history through computer simulation games, in: International Conference on Learning Sciences, International Society of the Learning Sciences, 2004, pp. 505–512.
[3] B.R. Lowery, F.G. Knirk, Micro-computer video games and spatial visualization acquisition, Journal of Educational Technology Systems 11 (1982) 155–166.
[4] S.B. Silvern, P.A. Williamson, The effects of video game play on young children's aggression, fantasy, and prosocial behavior, Journal of Applied Developmental Psychology 8 (1987) 453–462.
[5] J. Smith, I can see tomorrow in your dance: a study of dance revolution and music video games, Journal of Popular Music Studies 16 (2004) 58–84.
[6] R.E. Chatham, Games for training, Communications of the ACM 50 (2007) 36–43.
[7] J. Pearl, Causality: Models, Reasoning, and Interface, Cambridge University Press, 2000.

[8] M. Brewer, Research design and issues of validity, in: H. Reis, C. Judd (Eds.), Handbook of Research Methods in Social and Personality Psychology, Cambridge University Press, Cambridge, 2000.
[9] R.P. McMahan, A.J.D. Alon, S. Lazem, R.J. Beaton, D. Machaj, M. Schaefer, M.G. Silva, A. Leal, R. Hagan, D.A. Bowman, Evaluating natural interaction techniques in video games, in: IEEE Symposium on 3D User Interfaces (3DUI), 2010, pp. 11–14.
[10] P. Sweetser, P. Wyeth, GameFlow: a model for evaluating player enjoyment in games, Computers in Entertainment (CIE) 3 (2005) 3.
[11] D. Johnson, J. Wiles, Effective affective user interface design in games, Ergonomics 46 (2003) 1332–1345.
[12] R.F. Bowman, A Pac-Man theory of motivation: tactical implications for classroom instruction, Educational Technology 22 (1982) 14–17.
[13] B.J. Schachter, Computer Image Generation, Krieger Publishing Co., Inc., Melbourne, FL, 1983.
[14] J. Kirriemuir, A. McFarlane, Use of computer and video games in the classroom, in: Level up Digital Games Research Conference, 2003.
[15] C. Kolo, T. Baur, Living a virtual life: social dynamics of online gaming, International Journal of Computer Game Research 4 (2004).
[16] M. Schulzke, Moral decision making in fallout, International Journal of Computer Game Research 9 (2009).
[17] C. Paulk, Signifying play: the sims and the sociology of interior design, International Journal of Computer Game Research 6 (2006).
[18] A.T. Larssen, L. Loke, T. Robertson, J. Edwards, Understanding movement as input for interaction – a study of two eyetoy games, in: Annual Conference of Australian Computer–Human Interaction Special Interest Group (OZCHI), 2004.