Bot or not? User Perceptions of Player Substitution with Deep Player Behavior Models

Johannes Pfau  
Digital Media Lab, TZI, University of Bremen  
Bremen, Germany  
jpfau@tzi.de

Jan David Smeddinck  
Open Lab, School of Comp., Newcastle University  
Newcastle upon Tyne, UK  
jan.smeddinck@newcastle.ac.uk

Ioannis Bikas  
Digital Media Lab, TZI, University of Bremen  
Bremen, Germany  
bikasio@tzi.de

Rainer Malaka  
Digital Media Lab, TZI, University of Bremen  
Bremen, Germany  
malaka@tzi.de

ABSTRACT
Many online games suffer when players drop off due to lost connections or quitting prematurely, which leads to match terminations or game-play imbalances. While rule-based outcome evaluations or substitutions with bots are frequently used to mitigate such disruptions, these techniques are often perceived as unsatisfactory. Deep learning methods have successfully been used in deep player behavior modelling (DPBM) to produce non-player characters or bots which show more complex behavior patterns than those modelled using traditional AI techniques. Motivated by these findings, we present an investigation of the player-perceived awareness, believability and representativeness, when substituting disconnected players with DPBM agents in an online-multiplayer action game. Both quantitative and qualitative outcomes indicate that DPBM agents perform similarly to human players and that players were unable to detect substitutions. In contrast, players were able to detect substitution with agents driven by more traditional heuristics.

Author Keywords
Player Substitution; Game Disruption Prevention; Player Modeling; Neural Networks; Deep Learning; Games; Games User Research

CCS Concepts
•Human-centered computing → User models;  
•Computing methodologies → Neural networks; •Applied computing → Computer games;

INTRODUCTION
Match disruptions in online games are one of the major causes for frustration reported by players and make for a frequent occurrence given varying network quality depending on location and over time [23, 5]. Designing and deploying scalable online games that avoid interruptions remains an important challenge [17]. Even with recent advances in network stability, the complete prevention of any disruptions is highly unlikely [4]. Apart from unintended cut-offs, disconnecting on purpose can also occur due to a range of reasons, such as escaping, in which players avoid their loss to be recorded, resentful behavior (“rage-quitting”), in which players seek to deprive their opponent(s) of victory or intentionally hurt their own team in collaboratively competitive games, as well as forced disconnects of opponents via glitches or third-party tools [58, 57, 32, 60, 31]. To counteract purposely caused interruptions, some games record them as losses or penalize them, which can lead to even higher frustration for non-self-inflicted disconnects [40]. Other examples of successful commercial games substitute disconnected players by heuristic, computer-controlled bots that continue playing (in some examples only until the original player reconnects), e.g. Left 4 Dead [54] (an FPS), Heroes of the Storm [13] (a multiplayer online battle arena game), Super Smash Bros. 4 [46] (a Beat’em up), Mario Kart 8 [11] (a racing game), Civilization V [15] (a turn-based strategy game), Company of Heroes 2 [14] (an RTS), or Rocket League [39] (a sports game). However, such substitution is frequently criticized, since the replacing bot is usually under-performing and not able to compete with human players. While modern machine learning approaches have proven to master a variety of games by continual improvement through simulated play [30, 49, 43], over-performing bots would also miss the point of adequate, representative substitutions, since they would yield an obvious and considerable potential for abuse.

Challenging all of the aforementioned issues, we approach the bridging of temporary match disruptions with a novel method, utilizing Deep Player Behavior Modeling (DPBM) to
We hypothesize that a sufficiently accurate representation of individual behavior will be indiscernible from the original human player and that DPBM is capable of implicitly approximating the player’s game proficiency, leading to no significant perceived deviation in performance.

In order to establish a suitable test bed for the evaluation, we designed and implemented Korona: Nemesis [7], a platform fighter focused on player skills around prediction, learning and decision making. The game facilitates competitive skill-based play using an extended rock-paper-scissors mechanic to allow a broad range of play styles to arise by preference rather than encouraging dominant strategies. In an ecologically valid real-world field study \( n = 312 \), we simulated substitutions of players during online matches and assessed detection rates, awareness towards bot presence and DPBM fitness over the course of four weeks. Our study shows that participants were not able to discriminate DPBM behavior from original human players and – at the same time – that they were significantly more likely to detect replacements with classic heuristically-driven bots. Between players that successfully detected a DPBM bot and those who were unaware, there were no differences in perceived performance or predictability. Supported by additional qualitative results, we conclude that DPBM are a suitable method for temporarily substituting disconnected players in online games and generate adequate and desirably human-like behavior. These findings contribute to game user research and game development alike, by demonstrating a technically feasible and successfully evaluated approach that can lay the foundations for considerable advancements in the challenge of overcoming negative consequences of online match disruptions.

RELATED WORK

Network stability and connection maintaining are under steady improvement, both in terms of progress on physical connections, as well as through the development of architectures and protocols for tackling discontinuity issues [55, 27, 38] or prediction of traffic anomalies to counteract bandwidth- or connectivity-loss before it becomes critical [16, 22]. Yet, online games are still vulnerable to connectivity disruptions, since they can arise from a large variety of potential error sources, ranging from fast-paced real-time mechanics over massively large amounts of simultaneous players to vast connection distance differences that can span continents. In combination, these issues are improbable to be overcome completely and can significantly impact the motivation of affected players and of other players in the same play-session. Disconnected players in cooperative team fights for example, have to be compensated for by allies which – depending on the game and genre – is unlikely to be manageable beyond short durations [18].

Originating from the more general approach of user modeling [3, 56, 61], the relatively young field of player modeling has developed steadily during the last decade, with approaches rooted in applications of machine learning techniques for data mining large sets of game protocols for purposes of analysis, prediction or classification [8, 26, 48, 42, 12], informing game development with player-specific insights [9, 6, 25], or the reproduction of limited, atomic tasks [51, 47]. Holmgård et al. studied personas for player decision modeling [19, 20] that continually observe and adapt to human behavior in order to produce agents with different decision making styles. These personas were realized via evolutionary linear perceptrons and compared to heuristic agents in a test-bed 2D dungeon crawler game, resulting in a higher player-rated human-likeness that could be utilized for game analysis, testing or providing believable opponents. They also assessed player models as defined as “deviations from theoretically rational actions” in a study of Super Mario Bros. [21, 1] and clustered these by means of feature extraction. Using the same game, Ortega et al. [34] imitated human playing styles by means of neuroevolution and dynamic scripting, reaching higher scores of human-likeness than performance-directed AI agents, based on subjective judgments. Missura and Gärtner utilized player modeling in a 2D test-bed shooter via support vector machines as a predictor for difficulty mismatches and to enable dynamic difficulty adjustment (DDA) based on the results [29]. Transforming the tracks of a racing game, Togelius et al. successfully deployed player modeling as a method of assessing entertainment metrics [50]. In previous work, we were successful in showing that player modeling agents yield significantly higher motivation potential than heuristic opponents [36]. In addition, we contrasted different machine learning techniques in a player modeling study of the MMORPG Lineage 2 [33, 35], showing that deep learning offers the highest individual prediction accuracies with the ability to reproduce playing sessions that closely resemble the original behavior, as well as offering the potential to differentiate between players. Based on this, we embedded DPBM into a long-term DDA evaluation about competing against agents of own behavior on a daily basis in the MMORPG AION [37], in which DPBM opponents were perceived to be significantly more engaging than traditional DDA opponents adjusted by heuristic parameter tuning.

In computer generated behavior in general, human likeness or believability has been established as one of the most important metrics to facilitate engaging game play [52, 24, 2, 28, 53, 34, 19]. However, these approaches have focused on producing a general closeness to human behavior so far, not explicitly on representing behavior from specific individual players within
the same game session. Although player disconnects pose long-standing challenges, substituting disconnected players by means of player modeling bots has not been approached in openly published materials before, neither academically nor in the games industry, and – to the best of our knowledge – there is no prior scientific research on alternative temporary replacements.

**APPROACH**

In this section, we outline a description, critical design decisions and mechanics of the game utilized for the evaluation, and provide a detailed overview of the architecture, method and parameters of the DPBM approach.

**Game Design**

To provide a setting for studying crucial decision making in real-time, we designed a fast-paced physic-based platform fighter called Korona:Nemesis that extends the classic rock-paper-scissors scheme to seven types of element projectiles (cf. Table 1). In each level, players are placed in a 2D environment, start with 100 health points (HP) and face the objective of eliminating their opponents’ HP (last player standing wins). Players can move (left or right), jump, attack or switch actions using mouse and keyboard or an XBox or Playstation controller. Switching changes the current stance to one of the 7 elements. Giving the ability to chose any element at any time remedies potential balancing-issues, as the available action-spaces are – in principle – symmetric. Attack will launch an elemental projectile depending on the current stance. Getting hit by a hostile projectile subtracts 10 HP. Since this damage is doubled on a critical hit and projectiles can be destroyed, reflected or influenced by other projectiles (cf. Table 1), players constantly have to be aware of present projectiles, their own and enemies’ stances and adapt quickly to the situation. As in rock-paper-scissors, predicting the opponent is key to success and since players adapt and react constantly, there is no single dominant strategy.

- Exemplary game-play scenario:
  When facing an incoming Fire projectile, there are multiple viable choices. The player might react with a Water attack, since Water projectiles destroy Fire projectiles (cf. Figure 1). A more offensive choice would be to counter this attack with a Pain attack, which would not stop the incoming projectile, but critically hit and ignite the opponent. At the same time, the opponent has the opportunity to re-counter this, depending on making good predictions (e.g. if predicting a Water counter-attack and intending to counter it with Lightning. Yet again, this strategy may fail: If the Water prediction turns out to be wrong, attacking Pain with Lightning will incur a critical hit).

<table>
<thead>
<tr>
<th>Element</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fire</td>
<td>Cancels Restoration</td>
</tr>
<tr>
<td></td>
<td>Critically hits Restoration/Steel</td>
</tr>
<tr>
<td></td>
<td>Destroys Steel projectiles</td>
</tr>
<tr>
<td></td>
<td>Applies burning damage over time</td>
</tr>
<tr>
<td>Water</td>
<td>Immunity against burning</td>
</tr>
<tr>
<td></td>
<td>Critically hits Fire/Steel</td>
</tr>
<tr>
<td></td>
<td>Destroys Fire projectiles</td>
</tr>
<tr>
<td>Lightning</td>
<td>Immunity against suffering</td>
</tr>
<tr>
<td></td>
<td>Critically hits Water/Death</td>
</tr>
<tr>
<td></td>
<td>Destroys Water projectiles</td>
</tr>
<tr>
<td>Restoration</td>
<td>Restores 10HP</td>
</tr>
<tr>
<td></td>
<td>Converts Water projectiles into 10HP</td>
</tr>
<tr>
<td></td>
<td>Immunity against Pain</td>
</tr>
<tr>
<td>Steel</td>
<td>Reflects Lightning projectiles</td>
</tr>
<tr>
<td></td>
<td>Reflects Pain projectiles</td>
</tr>
<tr>
<td></td>
<td>Critically hits Lightning/Pain</td>
</tr>
<tr>
<td>Death</td>
<td>Inverts Restoration</td>
</tr>
<tr>
<td></td>
<td>Critically hits Restoration/Pain</td>
</tr>
<tr>
<td></td>
<td>Applies suffering damage over time</td>
</tr>
<tr>
<td>Pain</td>
<td>Self-ignites Fire</td>
</tr>
<tr>
<td></td>
<td>Critically hits Fire/Lightning</td>
</tr>
<tr>
<td></td>
<td>Applies 0.4 seconds stun</td>
</tr>
</tbody>
</table>

Table 1. Elements and their interactions in Korona:Nemesis.

Players need to learn not only the in-game element-interactions, but also their preferred way to counter attacks...
and maximize their chances, depending on the current situation. The presence of multiple viable choices, preferences and dislikes makes for a fertile ground for player modeling and decision making studies. For the evaluation of this work, participants were introduced to the mechanics via an in-game tutorial and were then able to play online matches consisting of 20 rounds in total.

**Deep Player Behavior Modeling**

Based on insights about expressive data and suitable modeling techniques from our earlier work [35, 36], we recorded all crucial player action decisions (attacking with – or switching to – a specific element and jumping) together with situational data from the current game state. After every level and for each player, the recorded behavioral data from all preceding levels was fed into a dedicated 24x10x10x9 feed-forward multi-layer perceptron with backpropagation and a logistic sigmoid activation function (cf. Figure 2). The network was initialized randomly and trained in a background thread over 1000 epochs, based on previous findings [35, 36] and benchmarks prior to the study that indicated diminishing returns beyond these parameters. When a DPBM bot substituted a player, it applied the trained model generatively to retrieve a set of action probabilities based on the given state description in real-time. After a weighted choice, it executed the most likely predicted skill and proceeded with querying the DPBM for the next situation, effectively approximating the learned behavior from the player’s decision making so far. Since movement characteristics are rather limited within the game, motion behavior is approached heuristically. This implementation realizes a model-free (bottom-up) player modeling approach mapping gameplay data to actions via preference learning and classification, employing the player modeling taxonomy of Yannakakis et al. [59]. According to the player modeling description framework of Smith et al. [44], DPBM directly utilizes game actions (domain) to generate (purpose) individually (scope) modeled behavior by means of induced (source) training of machine learning techniques.

**Heuristic Bots**

Instead of DPBM bots, heuristic bots substituted players in situations where no recorded behavior or trained model was available, i.e.:

- When players waited for over 2 minutes in the online multi-player lobby, heuristic bots filled the remaining slots to enable constant, comparable 4-player situations. Since DPBM training took place on the involved local machines parallel to the matches and the game followed a client-hosted design, no existing behavioral data could be acquired from a centralized server.

- When a player disconnected, but the background training thread for his DPBM counterpart was not completed. Yet, due to the considerably low temporal demand (cf. Results), this incidence occurred rarely.

- When a player deliberately disconnected before displaying enough behavior information for training.

Based on the insights of previous work [35, 36], we chose to endow the heuristic bots with random decision making between elements, since it yields a balanced performance level, analogous to random decision making in rock-paper-scissors. Thus, contrary to human and DPBM opponents, it was impossible for other players to predict this behavior. Movement was realized in the same heuristic fashion as for DPBM bots to avoid the detection of differences based on movement characteristics.

![Figure 2. DPBM architecture for a single player; mapping game state (information about player and closest target) to action probabilities.](image-url)
waited for other players to join their match. If less than four were able to enter the online multi-player lobby in which they were controlled by a bot after every completed match. Additionally, players were asked to complete a follow-up survey. At this point, informed consent was gathered and data was stored in a pseudonymized fashion.

**Measures**

In-game, we recorded state-action data for DPBM training (cf. Figure 2), local training times and prediction accuracies of the DPBM, and the player’s estimation whether and which players were controlled by a bot after every completed match. Additionally, players were asked to complete an online post-study questionnaire concerning demographics, subjective remarks and quantitative assessments of substitution awareness, asking the following set of 7-point Likert scale questions (separated by page transitions) that were constructed for this purpose:

- One of the players suddenly behaved differently.
- I felt that one player suddenly played better than they did before.
- I felt that one player suddenly played worse than they did before.
- I felt that one player suddenly became very predictable.
- I suspect that one of the players was switched for a bot.

**Procedure**

Participants could download *Korona:Nemesis* and play any number of matches without restrictions. Following a tutorial that demonstrated the basic mechanics of the game, they were able to enter the online multi-player lobby in which they waited for other players to join their match. If less than four players connected after two minutes, the remaining slots were filled by heuristic bots. During every active match, we intervened by substituting a random player by a DPBM bot that was trained in parallel to the playing session up until that point. If no trained model was available at that point, a heuristic bot took the place of the player. This replacement happened at a randomized point in time between round 5 and 15. To avoid discriminating the substituted players or diminishing their playing experience by being removed from play, they were immediately shifted into a new match that mirrored the original, differing only in the fact that the remaining three players were substituted in this version (cf. Figure 3).

The displayed appearance, name and score of replaced players was kept consistent in both matches at the time of the fork. After 20 rounds, players entered an end-screen depicting the ranking of all competitors, were encouraged answer the single in-game bot detection question and were then redirected to the main menu. In case they accepted the additional post-study questionnaire, they were referred to it using their standard browser.

**Participants**

During the study period, 1397 unique players downloaded *Korona:Nemesis*. (n = 312) submitted complete, pseudonymized game protocols and bot detection responses, encompassing 206 multi-player sessions in total. 24 of the players from these sessions (82.61% male, 17.39% female (self-identified), aged (M = 22.4, SD = 3.75)) completed the optional post-study questionnaire. 91.3% stated to be active gamers (playing multiple times a week), while 43.5% indicated that they only play occasionally (multiple times a month) and another 4.35% do not regularly play video games.

**Explorative Laboratory Study**

In order to pilot our approach and study design and to accumulate qualitative statements about reasons for detecting bots, the general perception of them and desirable behavior, we also conducted an explorative laboratory study (n = 7). Participants were publicly recruited on-campus of a university, asked to play a match of *Korona:Nemesis* and subsequently participated in a semi-structured interview. For reasons of clarity in our observations, only one of the four players necessary for a match was controlled by a participant, while trained experimenters filled the remaining slots, with one of them randomly being substituted. The experiment lasted about 30 minutes in total.
Measures
In addition to the aforementioned measures of the field study, a semi-structured interview assessed qualitative aspects of the player experience. Participants were able to provide free responses about the game, game experience and the behavior of their opponents, before the following directed questions were asked (in order and on separate pages).

- What do you think of the game?
- Did you notice anything strange during the game?
- Did you notice a change of behavior of other players?
- Do you think that there was a bot playing in this match?
- How can you tell that a player is actually a bot (in general)?
- How do you think bots in general should be improved to be (more) enjoyable?

Procedure
Following informed consent, participants were introduced to the game and asked to play the tutorial, without an enforced time limit. Once a player decided to proceed to visiting the online multi-player lobby, the experimenters joined soon thereafter, starting the match once the player count completed to four. All experimenters were kept spatially separated from the participants during the time of the match to avoid confounding factors from association or observation. The following procedure was analogous to the field study, only differing in the additional semi-structured interview that took place between match and post-study questionnaire.

Participants
(n = 7) subjects participated in the explorative pilot study (62.5% male, 37.5% female (self-identified), aged (\(M = 23.86, SD = 3.34\)). 42.86% self-identified as active gamers (playing multiple times a week), while 28.57% respectively indicated that they only play occasionally (multiple times a month) or do not regularly play video games.

RESULTS
The following quantitative outcomes resulted from the main field study, while qualitative insights of the laboratory pilot study are discussed at the end of the section.

<table>
<thead>
<tr>
<th>guessed behavior</th>
<th>actual behavior</th>
<th>human</th>
<th>DPBM bot</th>
<th>heuristic bot</th>
</tr>
</thead>
<tbody>
<tr>
<td>isHuman</td>
<td>87.18%</td>
<td>85.48%</td>
<td>32.75%</td>
<td></td>
</tr>
<tr>
<td>(68)</td>
<td>(53)</td>
<td>(75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>isBot</td>
<td>12.82%</td>
<td>14.52%</td>
<td>67.25%</td>
<td></td>
</tr>
<tr>
<td>(10)</td>
<td>(9)</td>
<td>(154)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Percentages (and absolute numbers in parentheses) of bot detection estimates, according to the responses to the in-game bot detection survey.

Using a chi-square test of independence with Yates-correction, a significant difference in guessing whether a player’s behavior stems from a human or bot could be found based on the groups of actual human players, DPBM bots and heuristic bots (\(\chi^2_{2,369} = 97.11, p < .001\), Kramer’s \(v = 0.36\), (cf. Table 2 for percentages and absolute estimate numbers). For differentiation between bot types, we further assessed the differences between the three particular groups:

- **Actual human players and DPBM bots:** \(\chi^2_{1,140} = .002\) (not significant)
- **Actual human players and heuristic bots:** \(\chi^2_{1,307} = 67.1, p < 0.001\) (significant), \(\phi = .47\)
- **DPBM bots and heuristic bots:** \(\chi^2_{1,291} = 52.95, p < 0.001\) (significant), \(\phi = .43\)

Concerning the awareness scale constructed for this study, we compared answers between players that managed to successfully detect a substitution and players unaware of substitution.

![Figure 4. Boxplot illustrating the results (medians, standard deviations as boxes, minima and maxima as whiskers, significant differences indicated with *). (i) I felt that one player suddenly behaved differently. (ii) I felt that one player suddenly played worse than he did before. (iii) I suspect that one of the players was switched for a bot.

Figure 4. Boxplot illustrating the results (medians, standard deviations as boxes, minima and maxima as whiskers, significant differences indicated with *).

“One of the players suddenly behaved differently.” \((t_{23} = 2.10, p = .04, \text{Cohen's } d = 1.3)\)
and “I suspect that one of the players was switched for a bot.” \((t_{23} = 3.11, p = .005, \text{Cohen's } d = 1.98)\).
The average DPBM training time (computed locally on each game client) amounted to \((M = 2.23, SD = 2.87)\) seconds. Within each iteration, 80% of the recorded data was used for training, while the remaining 20% allowed for following routine tests, resulting in a prediction accuracy of \((M = 82.17\%, SD = 23.17\%)\). There was a strong positive correlation between the amount of data points used for training and the prediction accuracy of the following test (Pearson’s \(r_{2871} = .64, p < .01\)).

**Explorative Laboratory Study**

Additionally, the laboratory pilot study yielded augmentative qualitative results. 6 of 7 participants remarked that they liked the game overall. None of them noticed anything generally strange in the session, nor a change in behavior of one of the players. Regarding the question whether they recognized a bot, no one managed to provide a correct answer (4 of them did not detect a substituted player, 3 incorrectly judged human players to be bots). In a notable contradiction to this finding, when asked, what they expect from the behavior of a bot, the participants consistently responded that bots are typically noticeable due to their bad performance (5x) or predictable strategies (3x). In response to the question “How do you think bots in general should be improved to be (more) enjoyable?”, they stated that they “would like them to be as human as possible”, would want bots that are “adaptive (like humans), but not with superhuman performance”, and that “playing with real people feels better”.

**DISCUSSION & FUTURE WORK**

With respect to the initial research question “Can disconnected players in running online matches be substituted by DPBM agents without being detected?”, we found quantitative as well as qualitative outcomes that support our hypothesis that DPBM yields a feasible approach for player substitution. The results of the bot detection estimation (cf. Table 2) indicate that participants were not able to differentiate between human and DPBM behavior, even if they were substituted during a running match. The significant difference of this finding to the frequent detection of heuristic bots answers “Is DPBM capable of providing measurably better substitutions than traditional (heuristic) methods?” in favor for DPBM and amplifies the expressiveness of the former results, since players evidently were able to detect bots, if their behavior was less human-like. Qualitative insights from the laboratory study complete the picture of a successful substitution, since participants stated to be unaware of changes in behavior after DPBM substitutions and were unable to correctly name replaced players.

The true positive rate of 87.18% for human behavior aligns fittingly with related research in which participants were asked to judge game sessions according to whether a human was playing *The Legend of Zelda* [10] (88.7%) or *Boulder Dash* [45] (80.7%) [24].

Regarding the remaining research question “Do DPBM yield an adequate, fair representation that does not improve or worse the original player’s performance?”, we provide evidence based on the awareness questionnaire constructed for the purpose of this study. Player proficiency or performance can develop during game play, but there was no significant increase or decrease or change in predictability between detected bots and undetected bots or regular players. Together with a considerably high DPBM prediction accuracy, this supports the claim that DPBM behavior does not significantly deviate from the original human player behavior. Additionally, our approach meets the desired ideal behavior of bots, according to the qualitative statements that players prefer to play against opponents that are as human-like as possible.

Still, this study faces limitations. In general remarks on the field study, 3 participants stated that they played *Korona:Nemesis* simultaneously with a friend who took part in the same session, while constantly communicating. The discrepancy between the original and the mirrored match (that could be communicated between the players) was the main cause of detecting the substitution for these players, as opposed to actually judging changes in behavior. We were not able to prevent this potentially confounding factor in the large-scale field study, as we aimed for maximizing the ecological validity of the approach. However, even if this introduced a bias to our results, it would have increased the correct bot detection rates, which actually would decrease the possibility of a non-significant result of the bot detection estimation between human and DPBM opponents. The result, that people were not able to discriminate human and DPBM behavior nonetheless indicates that this bias was not significantly confounding.

Furthermore, one player claimed that a real-time game might not be the best test bed for substitution awareness, since players are too focused on themselves. While we can not disprove this assertion or control for some extent of bias, we explicitly designed *Korona:Nemesis* in an extended rock-paper-scissors fashion in which players have to pay attention to their opponent. Moreover, we argue that artificial behavior would likely be even more indiscernible in many other types of games, such as turn-based games, since action decisions that might seem idiosyncratic or not human-like would likely be assumed to be part of larger complex strategies that are common to turn-based games. Altogether, our study can only provide high certainty that DPBM player substitution works adequately, fairly and indiscernibly as implemented for the game *Korona:Nemesis*. Yet, we designed the game to be complex enough to facilitate individual preference formation and to require attention, prediction, learning and tactical decision-making without incorporating dominant strategies. We argue that the insights formed in this approach can be extended and generalized to other games in the genres of fighting games and decision-making-focused action RPGs. We are looking forward to assess awareness, believability and representativity of DPBM opponents in these and further genres, including turn-based, cooperative games and games that encompass complex movement characteristics.

The DPBM architecture was kept as frugal as possible, in order to ensure feasible training times on the uncharted multitude of different hardware constellations that were able to acquire the game via *Steam*. The low average time required for network training, however, suggests some room for elaborating more ambitious deep player modeling techniques (e.g. recur-
rent, deep belief, GAN or context-driven LSTM networks) to further improve the proximity to human-like behavior, or to model more complex observation-to-action mappings. Since – to the best of our knowledge – no evidence in the field of player modeling exists that would give an estimation about the connection of prediction accuracy and perceived human-likeness, we seek to aggregate data for a large-scale evaluation in which participants are asked to watch game sessions of DPBM agents with different gradations of prediction accuracy, judge them according to their human-likeness and allocate them to the correct human player from which the behavior originated. Additionally, no prior research exists that evaluates the perception of fairness when it comes to substituting players. Thus, we plan to assess this from both the substituted player’s perspective, as well as the impression from involved team mates and opponents. Eventually, we envision DPBM as an effective instrument for elevating autonomous game testing and balancing, since realistic player behavior can be employed, as well as for facilitating novel dynamic difficulty adjustment approaches that adapt to individual strengths, weaknesses and progresses of players over time.

**CONCLUSION**

Since unintentional, as well as deliberate disconnects, drop offs or client terminations are unlikely to disappear with conventional, stability-improving hardware and software methods, we demonstrated an alternative approach that bridges (temporary) player absence by substituting them with Deep Player Behavior Models (DPBM). An ecologically valid online field study (n = 312) with a duration of four weeks simulated the replacement of a human player in the online multi-player fighting platformer *Korona:Nemesis*, assessing the remaining players’ awareness, the believability of the substitution, and the performance-related representativeness. We conclude that players were not able to distinguish between DPBM bots and original human players, but notably managed to detect bots based on heuristic behavior. Perceived performance and predictability changes did not differ between players who did detect DPBM bots and players who indicated that they thought that they had been playing against other human players only. All together, we implemented and evaluated a novel approach to tackling online match disruptions and lay ground for further evaluations spanning additional games, genres and integrations.

According to the guidelines of transparent statistics, the collected data of this approach, as well as its implementation, will be made openly available upon publication, using an open-access repository.

**ACKNOWLEDGMENTS**

We would like to thank all participants. This work was funded by the German Research Foundation (DFG) as part of Collaborative Research Center (SFB) 1320 EASE - Everyday Activity Science and Engineering, University of Bremen (http://www.easecrc.org/), subproject H2.

**REFERENCES**


[56] Huan Yan, Chunfeng Yang, Donghan Yu, Yong Li, Depeng Jin, and Dah-Ming Chiu. 2019. Multi-site user behavior modeling and its application in video recommendation. IEEE Transactions on Knowledge and Data Engineering (2019).