

# Merlynne: Motivating Peer-to-Peer Cognitive Behavioral Therapy with a Serious Game

TINA CHAN, University of Waterloo, Canada  
 ROBERT P. GAUTHIER, University of Waterloo, Canada  
 ALLY SUAREZ, University of Waterloo, Canada  
 NICHOLAS F. SIA, University of Waterloo, Canada  
 JAMES R. WALLACE, University of Waterloo, Canada

Human-Computer Interaction researchers have explored how online communities can be leveraged for peer support, but general disinterest and a lack of engagement have emerged as substantial barriers to their use in practice. To address this gap, we designed *Merlynne*, a serious game that seeks to motivate individuals to support peers through Cognitive Behavioural Therapy (CBT). Our game explored use of the Proteus Effect – a phenomenon where players adopt characteristics of their in-game avatar – to motivate peer support through stereotyped ‘helpful’ and ‘unhelpful’ avatars. We then conducted a mixed-methods, exploratory study to investigate its design. We found that our game successfully motivated players to offer peer support, despite the substantial emotional labour required by CBT. However, we were not able to replicate the Proteus Effect, and did not find differences in that support based on a player’s avatar. In reflecting on our findings, we discuss design challenges and considerations for the use of serious games to motivate participation in mental health support, including: fatigue, a player’s need for self-expression and to relate to those they are supporting, and ludonarrative dissonance.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: mental health; peer-to-peer; serious game; cognitive behavioural therapy; proteus effect;

## ACM Reference Format:

Tina Chan, Robert P. Gauthier, Ally Suarez, Nicholas F. Sia, and James R. Wallace. 2021. Merlynne: Motivating Peer-to-Peer Cognitive Behavioral Therapy with a Serious Game. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 250 (September 2021), 23 pages. <https://doi.org/10.1145/3474677>

This is a test.

## 1 INTRODUCTION

Rising rates of mood disorders such as anxiety and depression have placed a substantial burden on mental health services [14]. To relieve this burden, research has explored a combination of informal symptom management techniques like online peer-to-peer support [6], and internet-based Cognitive Behavioural Therapy (CBT), offered through self-help programs [2]. These approaches

---

Authors’ addresses: Tina Chan, University of Waterloo, Waterloo, Ontario, Canada, [lt2chan@uwaterloo.ca](mailto:lt2chan@uwaterloo.ca); Robert P. Gauthier, University of Waterloo, Waterloo, Ontario, Canada, [rpgauthier@uwaterloo.ca](mailto:rpgauthier@uwaterloo.ca); Ally Suarez, University of Waterloo, Waterloo, Ontario, Canada, [ally.suarez@uwaterloo.ca](mailto:ally.suarez@uwaterloo.ca); Nicholas F. Sia, University of Waterloo, Waterloo, Ontario, Canada, [nicholasfsia@gmail.com](mailto:nicholasfsia@gmail.com); James R. Wallace, University of Waterloo, Waterloo, Ontario, Canada, [james.wallace@uwaterloo.ca](mailto:james.wallace@uwaterloo.ca).

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2021/9-ART250 \$15.00

<https://doi.org/10.1145/3474677>



Fig. 1. In *Merlynne*, players help non-player characters form CBT-based replies to irrational negative thoughts from real-world Reddit posts. They take on the role of either a cleric (left) or monster (right). The game takes place in an 8-bit fantasy role-playing setting (centre)

have demonstrated the therapeutic value in connecting strangers who can then support each other through the CBT technique of cognitive reappraisal [48]. However, online peer support has also been shown to have substantial limitations: it fails to attract diverse sets of participants [47], is found unengaging [11, 22], and has high dropout rates [2].

Games have been explored by the HCI community as an effective motivator for health behaviour change and may be particularly effective in overcoming these barriers. For instance, gamification – the use of game elements in non-gameful contexts [24] – can elicit higher adherence to exercise programs [69] and higher remission rates in clinical depression [64]. Serious games have also been shown to be effective in motivating use of self-help applications like *SPARX* [64], *Superbetter* [36], and *Habitica* [35]. However, there is currently little research investigating the use of gamification to engage *supporters* in therapeutic activities.

To investigate how serious games can motivate supporters to engage in CBT, we created *Merlynne* (Figure 1), a serious game which asks players to reply to stress-related content from *Reddit*, within a fantasy role-playing game (RPG) setting. *Merlynne*'s design builds on existing guidelines for internet-based CBT and gamification (e.g., [2, 18, 26, 78]) and was developed using a research through design approach. This artifact-driven approach enabled us to combine art, design, science, and engineering to develop insights into the ‘wicked problem’ [59, 80] of motivating peer-to-peer support through gameplay, where “it can be extremely difficult to understand and predict how a game’s design choices will play out dynamically in the presence of its players’ creative inputs” [31, pp.194].

We then conducted an exploratory, mixed-methods user study to investigate *Merlynne*'s potential for motivating participation in peer-to-peer CBT. Our study included quantitative measures of engagement, attitudes towards helping others, and connectedness to game content, and then used semi-structured interviews to investigate players’ motivations and reflections on the game. Our results show that players felt motivated to perform emotional labour in *Merlynne*: they purposefully provided encouraging support, empathized with those reaching out for help, and used their own experience in relating to those individuals and in shaping their responses. However, we also identified a number of challenges in motivating players to participate in CBT, including fatigue related to emotional labour, players’ overconfidence and need for self-expression, and dissonance between the game content and its subject matter.

In our study, we also tried to replicate the Proteus Effect [76] – a phenomenon where players adopt characteristics of their in-game avatar – to understand if it could be used as a design tool to motivate participation in CBT. We found that despite carefully designing two avatars to evoke ‘helpful’ and ‘unhelpful’ behaviours, a player’s avatar had a limited effect on their in-game activities. Thus, in discussing our results, we also reflect on our inability to evoke the Proteus Effect, and its suitability in motivating peer-to-peer support, despite its effectiveness in other studies (e.g., [41, 78]).

In summary, the contributions of our research are threefold:

- (1) The design and development of *Merlynnne*, a serious game that motivates peer-to-peer support through play;
- (2) A mixed-methods study that shows how serious games can motivate players to perform emotional labour, but which also identifies barriers including: fatigue, player needs for self-expression and to relatedness to those they are supporting, and ludonarrative dissonance;
- (3) A critical discussion of the avatar design choices we made in *Merlynnne*, and their ineffectiveness in evoking the Proteus Effect in our study.

## 2 RELATED LITERATURE

CBT is a short-term psychotherapy based on the premise that thoughts and feelings play a fundamental role in our behaviour and that purposeful reframing of cognitive distortions can decrease an individual's anxiety, distress, and problematic behaviours [33]. A typical CBT activity asks individuals to perform cognitive reappraisals, cognitively restructuring or "reframing" their cognitive distortions with rational and positive perspectives, to mitigate undesired emotions and behaviours [29]. For example, one might reframe "I am late, I am a bad friend!", in terms of the positive perspective "I am usually on time, my friends will understand." Cognitive reappraisals are important to cognitive therapy (CT), but also to CBT which uses behavioural strategies like problem-solving and response prevention alongside cognitive framework for symptom management [7]. CT and CBT are sometimes used interchangeably in modern resources [7].

CBT was pioneered in the late 1970s [8], and since then has been established as an effective treatment for an array of disorders, such as anxiety and depression, somatic symptom disorders, and eating disorders such as bulimia [33]. Based on this effectiveness, researchers have explored ways of making CBT more accessible, and in particular, have focused on establishing peer-to-peer models of CBT that eliminate the need for one-on-one meetings with health professionals. Mead et al. [47, pp.6] defines peer-to-peer support as "a system of giving and receiving help founded on key principles of respect, shared responsibility, and mutual agreement of what is helpful" and empathy for each other's emotional experiences. In line with these benefits, the health care community has explored online communities of practice [66] as a means of matching those with common needs, lived experiences, and interests (e.g. [3, 57]).

But, there are also substantial benefits to including persons with no mental health service experience (or 'non-service users') in peer-to-peer programs. Diversity in support groups can help ensure the ability to evaluate subjective experiences, and to avoid normalizing abnormal behavior due to shared biased experiences [47]. Engaging non-service users in peer-to-peer support benefits users by simulating offline interactions and developing necessary skills for real world interactions. Davison et al. [22] found that online support users also often experience social isolation, and online mediums may be one of the few ways to reach those who have yet to seek the help they need. Moreover, redirecting the mental health narrative away from separating the diagnosed from the undiagnosed [47] aligns with the World Health Organization's holistic approach to health [54] and destigmatizes the treatment seeking process.

However, a common problem with these online peer-to-peer supports is that they fail to retain their members' interest, with members choosing not to participate when unsure of their own experiences' relevance or impact [68]. For instance, Morris developed *Panoply* [48], a website which crowdsourced strangers to cognitively reappraise each others' cognitive distortions with a guided CBT format. *Panoply*'s two-way engagement outperformed a one-way control; participants preferred peer responses over trained volunteers, and adopted CBT principles post-study into their daily life, specifically, the cognitive reappraisal technique used to restructure cognitive

distortions. Panoply showed the benefits of crowdsourced cognitive reappraisals for non-severe stress management, but Panoply's successor, *Kokobot* [49] has received criticism from industry reviewers for its lack of appeal and incentive for engagement [11]. Similarly, *Reach out Central* struggled with low user retention, despite high uptake from advertising on gaming media sites [16]. These criticisms are common for online CBT tools without face-to-face guidance [2], such as *MoodGYM*.

In this work, we address this common criticism of peer-to-peer CBT by exploring how peer supporters can be motivated to participate through the design of a serious game. We propose that serious games can organize interconnections between someone seeking advice and a peer's lived experiences, akin to "boundary objects" in communities of practice [75]. Our work parallels commercial games, such as Popcannibal's *Kind Words (lo fi chill beats to write to)* that asks players to write supportive letters to each other. A key difference with our work, however, is that we explicitly seek to motivate players to conform to the CBT format, and those interactions are embedded within an immersive fantasy RPG setting.

## 2.1 The Elaboration Likelihood Model and Serious Games

The Elaboration Likelihood Model (ELM) is a model of attitude change used in the public health field for marketing healthful behaviors to consumers [52], as well as to healthcare workers [9]. Using a person-centric approach, the ELM considers someone's motivation and ability to process a 'message' as predictors of attitude change [56]. The model consists of two routes, 1) central processing, for those with motivation and ability to consider a message based on its merits, and 2) peripheral processing, for those without the motivation or ability, and who therefore judge the message based on its association with peripheral cues (e.g., wants or fears) [56].

The ELM is also widely used in gaming research (e.g., [58]) as a behavioural change theory that justifies tailored narratives and directing attention with feedback [70] as a means of persuading players via motivational cues [58]. Games – through audio visual design, game mechanics, and narrative – offer peripheral cues to disinterested persons by providing opportunities for entertainment, expression, exploration, and achievement [26], that might then accompany a central message like participating in peer support activities. The ELM has been particularly helpful in designing gameful interventions for mental health. For instance, *SPARX*, a CBT self-help fantasy game yielded higher remission rates in clinical depression than a traditional face-to-face control [64]. *SuperBetter*, a self-help app which uses in-game rewards to motivate players to complete mindfulness exercises [36], also reduced depression symptoms more than a wait-list control [60].

However, current games research suggests that reward-based motivations in apps like *Kokobot* and *SuperBetter* can only fulfil short term goals [5, 50]. That is, they work primarily via peripheral processing in the ELM. Similarly, Barik et al. [5] argues that gameful design should provide experiences beyond Badges, Leaderboards, Awards, and Prizes (BLAPs), which target central processing routes. In this work, we used the ELM to inform the design of our serious game, and show that gameful design elements can act as peripheral cues which are then associated with central processing, enduring attitude change, and the potential for long-term behaviour change.

## 2.2 Gender and The Proteus Effect

Gender also plays a role in mental health support; men are less likely to seek out support than women [28]. When they do seek it out, men prefer peer-to-peer over professional support due to the stigma of help-seeking [42], and the requested support tends to be informational rather than emotional [38, 61]. Research also suggests that different messaging may be needed to appeal to such groups [19]. In light of these challenges, public health organizations have stressed the need for support systems for men, and which consider diverse and complex masculinities [62].

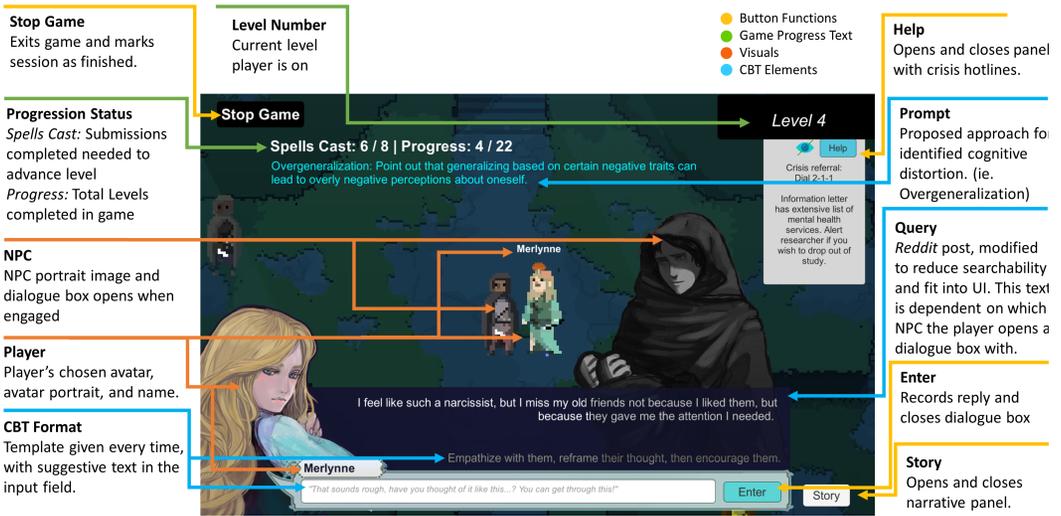


Fig. 2. We created *Merlynnne* using a research through design process, where we focused on integrating CBT into the game, motivational game elements like progress text and levels, and avatars to evoke the Proteus Effect.

The Proteus Effect [76], a phenomenon where players adopt characteristics based on avatar stereotypes such as gender role, offers one potential route to address these disparities. Research has found that in-game avatars can influence helping behaviours [79], and that assigning female avatars to male players can motivate them to take on behaviours that are traditionally considered female. When describing the Proteus Effect, Yee et al. [78] found that men using female avatars offer more help to teammates, including interpersonal interactions between players than with men using male avatars. Further, Lehdonvirta et al. [41] observed fewer gender differences in support-giving in massively multiplayer online role-playing games (MMORPGs). Male players with female avatars provided more overall help in the form of materials, labour, and information than emotional help [41]. The Proteus Effect has been incorporated in prosocial game designs for understanding gender [63, 78], ethnicity [4], and nationality politics [34], but not for mental health support.

We therefore recruited *only* male participants to consider male-patterned support giving by design in our exploratory study. We attempted to replicate the Proteus Effect within the context of mental health support, and specifically sought to show that female avatars increased male players' willingness to offer peer support over time. Our study did not replicate the Proteus Effect in this context, and we reflect on the design decisions we made that may have limited its efficacy in *Merlynnne*.

### 3 GAME DESIGN

*Merlynnne's* interface and environment are inspired by North American RPGs like *Celeste* and Japanese RPGs like *Fire Emblem*, where players navigate an 8-bit fantasy world and interact with non-player characters (NPCs) through dialogue (Figure 2). Players take on the role of either a healer (CLERIC) or barbarian (MONSTER) on a quest to investigate a plague of negativity in the fantasy world Khamelot, in reference to Camelot of *King Arthur*. It was important for *Merlynnne* as a serious

game to resemble current commercialized games and popular literature to establish familiarity in controls and cultural context [67], and be seen as a game, instead of a mental health tool.

Merlynnne follows a hero's journey narrative, used in popular fiction such as *Lord of the Rings*, *Star Wars*, and *Don Quixote*, featuring a central character called to a challenge, overcoming obstacles, then bringing back knowledge to help others. Throughout the story, the player aids villagers in taming irrational negative thoughts so they can resume daily life, while being tasked with searching for the supernatural cause of their negativity. "Taming," instead of "killing," which is common phrasing in RPGs, because taming aligns with CBT principles that thoughts are to be accepted and reframed rather than rejected or suppressed [8]. The heroic narrative is noted by Lehdonvirta et al. [41] to be a motivator for male players to offer support to others in MMORPGs, due to media depictions of masculinity as heroic.

In each game level, players encounter three NPCs who each present the player with a cognitive distortion. Players are then asked to cognitively reappraise the thought through a dialogue box. The cognitive distortions were paraphrased posts about mental health from local university subreddits, chosen to create game content relevant and engaging for our target audience. Each of the 22 levels presented negative thoughts requiring different cognitive distortion categories, except for every sixth map, termed 'special levels', which feature a *rational* negative thought where reframing was unnecessary.

After each submission, the NPC would change appearance, signifying the negativity was addressed by the player. After responding to two NPC's, the player advanced to the next level featuring a new map, avatar cosmetic upgrades, and more narrative. We wished to avoid the forcibly linear progression between queries that Panoply was criticized for [48]. There were no penalties for submitting blank responses or responses with meaningless text to advance if players desired.

When developing *Merlynnne*, we followed a research through design process [80] that included opportunities for iterative design via feedback through play testing and expert review. Early prototypes were evaluated heuristically with 12 in-house HCI experts and a local mental health group in round table discussions. We then built a working version of our game and play tested *Merlynnne* with five members of our lab group to ensure that visuals, content, and interface design aligned with our goals (Appendix A). Lastly, we refined our game design using their feedback before moving to our exploratory study.

We now describe three aspects of *Merlynnne's* design: how we incorporated CBT into a serious game, our use of gameful design elements to motivate players, and how we designed avatars to evoke the Proteus Effect. Additional details about narrative, NPC interactions, gameflow, and game mechanics are available in Appendices B–E.

### 3.1 Incorporating CBT

We chose to focus on the cognitive reappraisal technique seen in Panoply [48] and MOODGYM [2]. Cognitive distortions provided by NPC interactions were grounded in real-world issues from two local university subreddits, using posts paraphrased by the first author who was familiar with the subreddit and the affiliated universities' student life to remove identifying factors or searchability. We manually searched the subreddits for applicable content, collected examples in a local database, and paraphrased each entry to represent common themes of the subreddit on the topic of mental health. Situations include stress associated with academic performance, conflicts with friends and family, and other examples of anxiety. Mentions of substance abuse or self-harm was omitted to reduce discomfort to players who may have previous experiences with these issues and queries were shortened to fit within space constraint of *Merlynnne's* dialogue box. Our final database included 48 entries, all of which were used for game content.

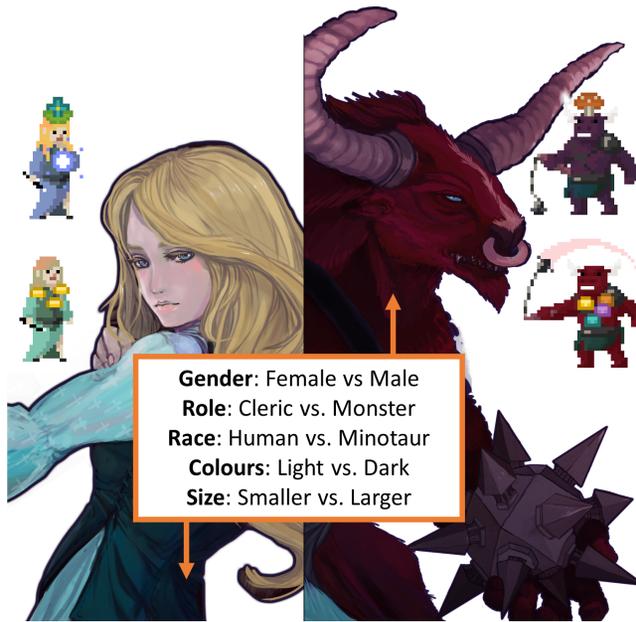


Fig. 3. We designed two 8-bit avatars and detailed portraits to appear either ‘helpful’ (Cleric, left) or ‘unhelpful’ (Monster, right) based on archetypes in popular media. Players chose from one of two colours for their avatar and earned new accessories with each level, ranging from novelty items to more prestigious items such as medals and jewels.

We used *Panoply*’s in-game CBT reply structure, which includes the following steps 1) identification of cognitive distortion [1]; 2) empathize; 3) reframe; and 4) encourage. Identification of the cognitive distortion (Step 1) was completed for the participant, as a prompt to address each cognitive distortion. We initially designed the game interface to mirror *Panoply*’s, with NPC dialogue boxes prompting players to empathize, reframe, and encourage in three separate fields. However, pilot testing pointed towards a single prompt for simplicity and usability. The literature also suggests that free-form templates encourage more personal responses [53].

Play testers also found the CBT format difficult to learn without examples, thus the first three game levels were converted into tutorial levels to ensure that players understood the game’s controls and had adequate practice with the format. The first level featured example replies, the second level pointed out progression mechanics, and the third offered a supervised chance to practice replying. Prompts were also added to assist players for each cognitive distortion type. The tutorial levels did not use queries paraphrased from reddit, but instead used clearly structured examples of cognitive distortions, and the NPCs were non-humanoid “blobs” instead of villagers to signify the difference.

Finally, following Andersson et al.’s [2] clinical guidelines for CBT, Merlynnne opens with clear statements that it is not a substitution for therapeutic support training. Since the game is not able to address urgent mental health concerns, and is based on real-world content that has the potential to instigate concerns among players, regional hotlines are listed on every game screen. Our design received full approval from our institution’s research ethics board (REB), and is consistent with HCI community guidelines for use of pseudonymous content (e.g. [15, 44]).

### 3.2 Gameful Design Elements

To understand how we could best motivate game play and facilitate mastery of CBT within the serious game [77] we reviewed potential game mechanics from the HCI community (e.g., [43, 72]). A summary of that review is available and the design elements we identified is provided in Appendix E. We ultimately decided to focus on three broad categories of in-game goals: achievement, immersion, and social. To ensure that mastery of in-game CBT was also conveyed to the player, we also integrated ability cues throughout the game. We summarize each of these gameful design elements below:

*Achievement goals* [77] were motivated through progression, feedback, and fixed reward schedules [43], through audio-visual feedback upon reappraisal submissions, and a progression bar indicating the number of levels completed and number of reappraisals sent. Novel content including new stories and maps were awarded with each level. Cosmetic rewards such as hats and necklaces (Figure 3) were also given to avatars as an indicator of achievement to motivate continual play [10].

*Immersion goals* [77] were implemented by limiting the information available to players at and offering opportunities for emotional investment [43]. Narrative remained unknown to players prior to unlocking and ‘special levels’ offered interactions different to regular levels throughout *Merlynnne*. A familiar game narrative emphasizing the need for the player’s engagement and relatable real-world queries persuaded players’ emotional investment. Avatar customization was also intended to increase player-avatar identification and immersion [10].

*Social goals* [77] could be achieved through opportunities for sharing knowledge and purpose [43], such as when NPCs asked players for emotional support using their knowledge and experiences. But, *Merlynnne* also made players aware of the real-world impact of their actions.

Finally, *Ability cues* [56] for CBT mastery were provided through learning supports and anonymity [43]. Gameplay and the CBT format were taught to players through a tutorial and subsequent prompts at each level to aid reappraisals. For player confidence, pseudonyms for anonymity [6] were encouraged with no penalty for gibberish or blank answers. We also anticipated that participants might provide poor/fast responses to speed up advancement in the game. While *Merlynnne*’s code permitted any response, copy and paste was disabled.

### 3.3 Avatar Design

We based our avatar designs on the fantasy tabletop gaming archetypes (e.g., [30]) of a woman cleric and a monster barbarian (Figure 3). Consistent with the stereotypes typically associated with the Proteus Effect [78], the player avatar was intended to evoke one or two extremes: either an intelligent and supportive (Cleric), or an aggressive and antagonistic (Monster). NPC avatars were designed to have a similar social identity to the cleric to promote helping behaviours by those playing as the cleric [73]. On the other hand, the barbarian monster avatar was designed to be much larger, and to appear inhuman, potentially de-motivating those behaviours. We also allowed players to customize their avatar to foster stronger identification [10].

During play testing we found that players quickly recognized the stereotypes of the chosen sprites, linking the cleric to the supportive character of Galadriel in *Lord of the Rings*, and the barbarian with ‘demonic’ creatures. Feedback from an in-house game linguistics expert suggested the term ‘cleric’ instead of ‘wizard’ which was originally used, as wizards are considered damage-dealing classes in fantasy games. We also found that the player’s focus was not on the avatar but on dialogue text, breaking the intended immersion of RPGs. We therefore added portraits of the NPCs and player avatar to enhance immersion in the game (e.g. Figure 2).

### 3.4 Implementation Details and Supplementary Materials

*Merlynnne* was built on the *Unity 5* engine. Artwork was created by Calciumtrice from [OpenGameArt.com](https://www.unity.com) and music composed by Monika Ziska. All appendices and *Merlynnne's* source code are available at: <https://osf.io/8ck6r/>.

## 4 EXPLORATORY STUDY

Having iterated on our game design through heuristic expert review, we next wanted to collect initial feedback from potential users, measure their engagement with the game, and explore the feasibility of different design choices in motivating engagement (e.g., different avatars). We decided on a convergent mixed-methods design [21], as this would allow us to collect quantitative data surrounding engagement and attitude change, while also enabling us to triangulate findings and more deeply investigate players' experiences through qualitative interviews.

Our experimental design included AVATAR as a between-subjects independent variable, with participants assigned by a research assistant through block randomization of 4 units [39] to play the game using either the CLERIC (or helpful) or MONSTER (or unhelpful) avatars. Dependent variables included their attitudes towards helping, connectedness to their avatar, and the degree to which they participated in CBT.

### 4.1 Ethical Considerations

We initially aimed to replicate Morris's Panoply study [48] which crowdsourced peer supporters to provide two-way CBT. However, our inability to control for malicious or inappropriate responses was identified as a potential harm to participants when developing the study protocol through consultation with our Research Ethics Board (REB). As a compromise, we decided on a laboratory study where participants were instructed that their answers would instead be stored into a response bank for potential future service improvement for the University. As we paraphrased real-world Reddit content for in-game use, we also protected original posters' identity in accordance with ACM guidance for research with public data from online communities [15, 44].

Participant recruitment and awareness of experimental conditions was also weighed against the risk of potential harm, in consultation with our REB. In the interest of participant safety, they needed to be aware of the study details upon recruitment, including exposure to stressful content and how their responses would be used. Recruitment materials emphasized the study as a games study rather than mental health study to avoid only attracting persons with a mental health interest (Appendix O). We also anticipated that observation may influence their behaviour (i.e., the Hawthorne Effect [45]), and included a hypothesis awareness test in our post-study interviews to assess whether this impacted our results.

Finally, to mitigate potential harms arising from deception we monitored participants at all times throughout the study, provided appropriate resources in case of distress, and assured them that their time was not being misused.

### 4.2 Participants

To determine an appropriate number of participants for our mixed-methods study we considered both local standards in HCI (i.e., [17, 74]) and a-priori power analysis. Local standards suggested 12-24 participants would be appropriate for our qualitative analyses. For our quantitative analyses, we performed a power analysis in G\*Power, using an effect size of  $f = .527$  [10], and Cohen's recommendation of a power of 80.75% [20], that suggested 36 participants were sufficient to identify differences in the HAS and Avatar Identification Scale data, corresponding to the Proteus Effect. We thus chose the higher of the two estimates for our study.

We then recruited 36 male participants through on-campus posters between June and October 2019. Inclusion criteria included being male, able to converse in English, and being familiar with using a keyboard and mouse at a computer. Participants were remunerated \$15 CDN at the end of the study.

All 36 participants were between 18 and 38 years of age ( $\bar{x}=25.11$ ,  $\sigma=4.47$ ,  $M = 25.5$ ), and none revealed transgender statuses. All had obtained at least a high school diploma, with the most common response being a terminal Bachelor's Degree (18/36), and all but two participants had at least some post-secondary education. All participants reported playing games between once a month and once every four months, with the most common response being "every day" (10/36). The majority of participants reported no mental health diagnoses which affected their stress (86.11%,  $n=31$ ), not having accessed mental health professionals in the last 6 months (75.00%,  $n=27$ ), and not being on any mental health treatment plans or medications (88.89%,  $n=32$ ). Only two participants (P21, P31) reported having worked or volunteered in the mental health field. There were no significant demographic differences between AVATAR groups.

### 4.3 Procedure & Apparatus

Our study was conducted at a local university where participants completed an introductory survey, then played *Merlynn* on a 2560 × 1600 LG 21" LED monitor at full screen, on a high-end gaming machine (Windows 10, Intel i7 3.6 GHz, 16 GB RAM, NVIDIA GeForce GTX 980) while seated at a desk with a Razer gaming mechanical keyboard and mouse. The computer's volume was set to 33 and adjusted at participants' request. Tutorials were completed with the researcher's guidance, followed by unsupervised play up to a 30 minute cutoff. An exit questionnaire was completed afterwards, followed by a semi-structured interview and remuneration. We told participants that their responses would be stored into a 'bank' for potential mental health service improvement.

### 4.4 Data Collection & Analysis

Introductory surveys collected participants' demographic information (Appendices I and J), experience with mental health services, and familiarity with the fantasy genre and games. Quantitative analyses were conducted on data logged to computer files, and pre- and post-study questionnaire data (Appendix K), including: the number of in-game responses and the time spent per session, attitude change as assessed by the Helping Attitudes Scale (HAS) [51] (Appendix G), avatar identification questions adapted from Dunn and Guadagno [27], and feelings toward the narrative (Appendix H). Mann Whitney U tests were used to compare usage data between the two avatar groups, and to identify pre- and post-study differences. Non-parametric tests were used because we could not assume the normality of our data. Analysis was conducted with  $\alpha = .05$  for all tests.

Collected in-game responses were assessed as binary variables ('Yes' or 'No') by three raters for whether the response 1) expressed empathy; 2) reframed a thought by offering a positive and rational perspective of the problem described in the query; 3) provided encouragement, 4) provided a solution, and 5) was not damaging such as not discouraging seeking support and encouraging harm to self and others (including use of non-prescribed drugs). The first three authors assessed a set of 24 responses of one participant together to reach consensus on rating guidelines (Appendix L), then the remainder sets were scored individually with scores tested for inter-rater reliability [46]. Raters shared student status (at the Undergraduate, Masters, or Doctorate level) with participants and did not employ a formal mental health framework in coding, to capture benefits of peer-to-peer systems [47]. If a Kappa score was less than 0.5 with  $p < 0.05$ , the category was re-assessed together for consensus until an acceptable Kappa score was reached. Final Kappa scores were: Empathetic (all the same), Reframing ( $\kappa = .83$ ,  $p < 0.01$ ), Encouraging ( $\kappa = .70$ ,  $p < 0.01$ ), Solution ( $\kappa = .864$ ,  $p < 0.01$ ), and Not Damaging (all the same).

Since participant responses were entered over the course of their sessions, we wanted to explore whether the number of queries completed impacted their response. We performed binary logistic regression modelling for the assessment variables Empathy, Reframe, Encourage, Solution, and Not Damaging in SAS [32]. For each assessment variable, we started with unconditional means models and introduced the following components one at a time: number of query, avatar types, and interaction. For each new model we tested the significance of main effects, for any components not involved in an interaction, and for interaction effects.

Finally, in-person interviews (Appendix F) were conducted post-play with all participants, recorded, and manually transcribed within two weeks of study completion. The main aim of the semi-structured interviews was to explore players' experiences and motivations and to reveal impacting factors such as pre-existing habits and the Proteus Effect. To explore these questions, the interview guide comprised: 1) post-play opinions and general game experience, 2) motivations for and impact of in-game actions, 3) relationship with the avatar and narrative, 4) expectations of themselves in a helping role, 5) feelings towards CBT format, 6) game playing habits and preferences, and 7) a hypothesis awareness test. The hypothesis awareness check consisted of asking the participant what they thought the hypothesis of the study was to screen for any biases in performance or response if they had correctly anticipated the hypothesis of the Proteus Effect.

We then performed a latent thematic analysis [13] using *NVivo 12*. Since our goal was to explain players' experiences in *Merlynnne*, we approached our thematic analysis from a constructivist epistemology [12]. The first, third, and fourth authors transcribed audio-recorded interviews verbatim. The first author then read all transcripts multiple times, segmented those responses into meaningful units, and performed initial open coding. The first author developed code sets based on the interview guide and transcripts, identifying a total of 414 nodes. The most frequently re-occurring nodes were then reduced under axial coding into four themes that were selected based on their relevancy to our research questions: dismissing fatigue in favour of play, conflict between self-expression and CBT, increased helpfulness post-play, and acceptance of avatar stereotypes. A reflexive journal [55] documented the first author's thoughts during the coding process (Appendix M). A summary of our codes, axial codes, and themes is provided in Appendix N.

#### 4.5 Reflexivity Statement

In examining the value of mental health solutions and gamification, reflexively understanding the author's mental health and gaming experiences may bring certain affinities into perspective [55]. The first author has volunteered in peer to peer communities online, plays RPG games regularly, and has previously engaged in work using CBT principles.

### 5 RESULTS

All participants completed at least 8 levels on their own, and the highest number of levels completed by any participant was 21 ( $M = 12$ ,  $\bar{x} = 13.11$ ,  $\sigma = 3.584$ ). Unsupervised play sessions lasted a median of 33 minutes and 17 seconds ( $\bar{x} = 33:17$ ,  $\sigma = 3:10$ ), with participants responding to a median of 18.5 NPC requests ( $\bar{x} = 19.81$ ,  $\sigma = 6.663$ ). The median response contained 28 words ( $\bar{x} = 30.76$ ,  $\sigma = 21.23$ ). No participants indicated providing gibberish responses during their interviews, and we did not identify any in our own post-study assessment. However, 36% ( $n=13$ ) of participants submitted blanks, with a median of 0.21 blanks per level ( $\bar{x} = 0.32$ ,  $\sigma = 0.31$ ). Blank responses were evenly distributed; that is, we did not see blank responses increase towards the end of the play session.

We first report on the results of our quantitative tests, before presenting our thematic analysis of participant interviews.

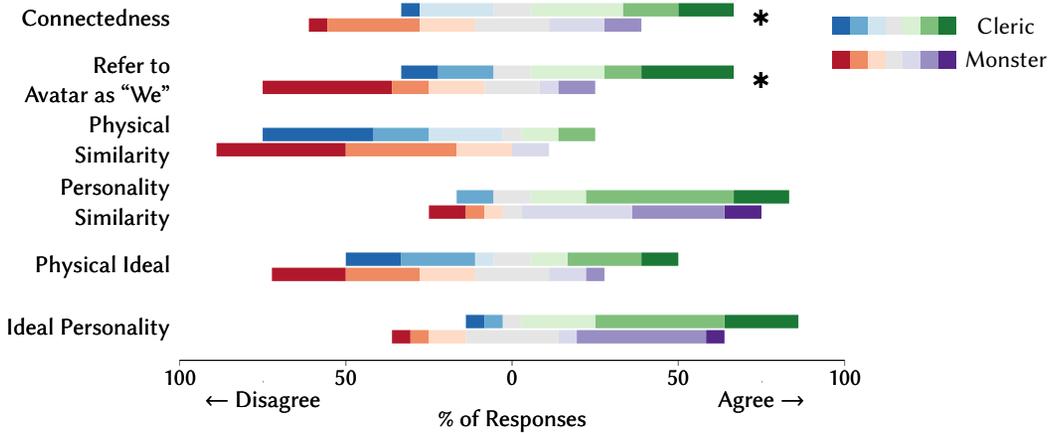


Fig. 4. Summary of avatar identification responses, based on 7-point Likert scales. Mann Whitney U tests revealed significant differences between AVATAR groups for ‘Connectness’ and ‘Refer to Avatar as We’.

## 5.1 Quantitative Results

**5.1.1 Helping Attitudes Scale.** Participants’ pre-study HAS score was ‘high’ ( $M = 82.5$ ,  $\bar{x} = 81.93$ ,  $\sigma = 10.42$ ) and post-study remained ‘high’ ( $M = 84$ ,  $\bar{x} = 84.10$ ,  $\sigma = 8.42$ ). When assessing change in score for the entire participant pool, Wilcoxon signed-rank tests showed a significant increase ( $Z = 154.5$ ,  $p = 0.02375$ ). However, Mann-Whitney  $U$  tests revealed no significant differences between the AVATAR groups ( $U = 145.5$ ,  $p = 0.4081$ ).

## 5.2 Avatar Identification

Avatar Identification Scale data (Figure 4) were evaluated with Mann Whitney  $U$  tests: Participants from the CLERIC group scored significantly higher on connectedness ( $U = 94.0$ ,  $p = 0.01514$ ) and on ‘refer to avatar as we’ ( $U = 80.0$ ,  $p = 0.004464$ ). There was no significant difference between AVATAR groups for physical similarity ( $U = 129.0$ ,  $p = 0.1430$ ), personality similarity ( $U = 127.0$ ,  $p = 0.1294$ ), physical ideal ( $U = 120.5$ ,  $p = 0.09415$ ), or ideal personality ( $U = 113.0$ ,  $p = 0.05592$ ). No participants were aware of AVATAR controls when asked post-study in interviews.

**5.2.1 Response Quality.** Nearly three quarters of responses successfully Reframed the irrational negative thought ( $n = 509$ ,  $\hat{p} = 71.39\%$ ), with more than half showing Empathy ( $n = 371$ ,  $\hat{p} = 52.03\%$ ) and Encouragement ( $n = 372$ ,  $\hat{p} = 52.17\%$ ) from the CBT criteria. Additionally, over ninety percent of responses were considered Not Damaging ( $n = 652$ ,  $\hat{p} = 91.44\%$ ) and just over thirty percent of responses described solutions ( $n = 235$ ,  $\hat{p} = 32.96\%$ ) which advised problem-solving or response prevention seen in behavioural approaches in CBT [7].

Prompted by a visual inspection of the response criteria data (Figure 5), we performed post-hoc binary logistic regression modelling to evaluate potential negative relationships over time. We found that scores for Empathy ( $p = 0.0421$ ,  $OR = 0.955$ ), Reframe ( $p = 0.0322$ ,  $OR = 0.965$ ), Encourage ( $p = 0.0095$ ,  $OR = 0.947$ ), and Not Damaging ( $p = 0.0017$ ,  $OR = 0.896$ ) decreased over time (NPC Requests). Despite expecting a decrease over time for Solution we found no significant effect. Our logistic regression included checking for impact of AVATAR Type and interactions between AVATAR and NPC Requests, however no significant effects were observed.

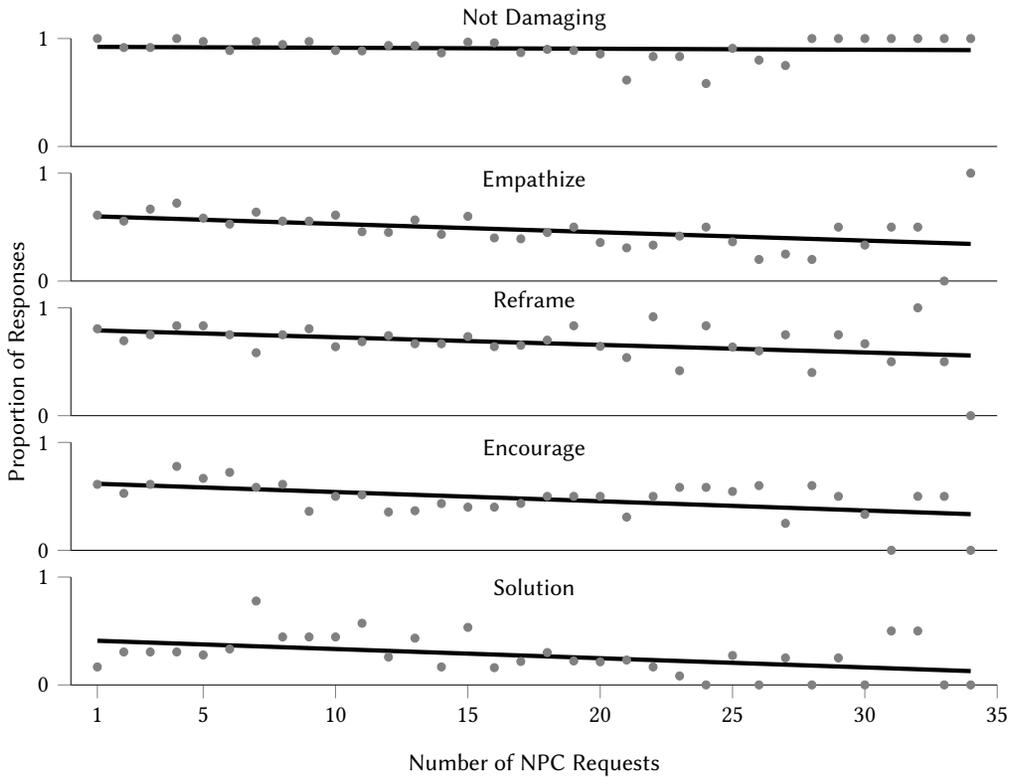


Fig. 5. The proportion of responses that were classified as meeting five criteria: Empathize, Reframe, Encourage, Solution, and Not Damaging. The three categories related to CBT, Empathize, Reframe, and Encourage, along with Not Damaging were found to decrease over time ( $p = 0.05$ ). However, we did not find these changes for Solution.

### 5.3 Thematic Analysis of Interview Data

Overall, participants reported feeling that they successfully used *Merlynnne* as a peer support tool; 27/36 expressed that their main motivation was to help others, and 21/36 used personal experiences to drive answers. When asked in the post-study hypothesis check, none of the participants correctly guessed our hypothesis that avatars would influence engagement. We identified four themes in our analysis of interview data: dismissing fatigue in favour of play, conflict between self-expression and CBT, increased helpfulness post-play, and acceptance of avatar stereotypes.

**5.3.1 Motivating Play, Dismissing Fatigue.** Participants felt that their avatars interacting with NPCs simulated their taking initiative to do so ( $n=16$ ). When reflecting on each of *Merlynnne*'s gameful elements, participants cited narrative ( $n=6$ ), anticipation of later special levels (P23), and cosmetic upgrades of their avatar ( $n=10$ ) as being particularly motivating. However, most participants saw level progression ( $n=16$ ) and a general sense of achievement as their primary motivators (19/36). P23 elaborated on the role that in-game achievements played on their own play:

“Because I’m more of a ‘completionist,’ I feel if there’s an achievement, I will pour my blood, sweat, and mostly my tears into achieving it.” (P23)

Some participants also acknowledged the strategy of submitting blanks or gibberish for the sole goal of advancement in the game ( $n=7$ ). However, many viewed these submissions as ‘cheating’ ( $n=15$ ), P23 explains:

“I don’t feel rewarded by skipping something ... I don’t feel like hitting enter or typing out some random ass [answer] so just so I could see what the last level is.” (P23)

However, some participants also reported feeling fatigued post-play ( $n=7$ ), yet were motivated to push themselves as much as possible either for game goals or personal responsibilities. Despite this fatigue, all 36 participants played for 30 minutes, the maximum time allowed in our study. When describing the source of their fatigue, participants cited a range of factors, including the relentlessness of negative requests, and being reminded of stressful situations in their own lives:

“I feel in some scenarios I feel the same thing so I feel a little bit stressful because I think I am the person who asked for the help.” (P13)

P16 thought the game was deliberately designed to induce fatigue for a study objective, yet emphasized that did not change their personal motivations to help.

*5.3.2 Self-Expression vs. CBT.* Many participants disliked *Merlynn*’s guided CBT format due to its structure, which they felt was not natural (11/36), and their inability to accurately convey their relatedness to those they were helping ( $n=5$ ). Participants also reported feeling more effective at cognitive reappraisals when they could draw from their personal experience ( $n=22$ ) than when they adhered strictly to the CBT format. In line with these perceptions, participants deliberately sought out opportunities to relate to NPCs in the game:

“I only ever answered 2 questions per level. But I would kinda go and walk over the third one and just kinda read what their problem was just out of interest.” (P21)

More than two thirds of participants said they forwent the CBT format (19/36), with almost half of participants responding that they specifically disliked CBT’s empathy component (16/36) since they did not know enough about the character they were helping to genuinely care, feel empathetic towards their negative thoughts, or to want to normalize overreactions. In these cases, participants reported adopting a number of different strategies to help them give advice. For instance, some preferred action-oriented advice and to deliver opinions more directly ( $n=8$ ), like P2 who prioritized situational factors over individual feelings:

“I am not that kind of person who’ll cheer you or something, I am that kind of person, I force you to face this reality and that then you save the reality and the you see the problem you can, when you calm down and then wipe out all the negative thoughts” (P2)

Conversely, some participants preferred emotional support over logical reframing ( $n=5$ ), and one participant felt that instructions for CBT were unnecessary and helping should be intuitive (P19). While acknowledged for its usefulness, participants felt the CBT format did not support personal relatedness, since no prior relationship or contextual understanding with the help seekers existed.

Finally, we found that although the game context put participants into a serious mindset, there were still players who gave harmful advice. Only one participant admitted to submitting hurtful advice on purpose, but justified it as “Generally it is good to give bad advice,” (P10). They had recommended self-harm to someone who was grieving a dying pet, and justified it by alluding to their own experiences with sarcasm:

“I told her that you can kill yourself as well to go with and to stay with her forever, it was a lie, but it just you know it was on purpose, it was to just to say to her that you know the situation can be worse.” (P10)

Other participants generally felt that it was not in their nature ( $n=5$ ) or it was unethical ( $n=7$ ) to hurt others with bad advice, especially since they were instructed to be helpful.

**5.3.3 Increased Willingness to Help.** Almost half of participants (16/36) felt positive post-play, and more than a third reported being more motivated to help others (15/36), due to increased perceived competence ( $n=10$ ) and increased willingness to perform emotional labour ( $n=4$ ). P21 described feeling compelled to help others in real life:

“if they say something like – maybe just like, sarcastically or like self-deprecatingly humorous, like ‘oh I haven’t studied for econ at all this week, I’m probably gonna fail the midterm’ or something – maybe not something like that but something similar where usually I’d just be like ‘yeah, me too’ or something. I might actually wanna be more supportive or helpful and be like ‘hey, like if you want we can meet up and study for it’ or like ‘I can give you some notes.’” (P21)

Interestingly, P15 said the game experience persuaded him to prioritize helping others over his own comfort:

“...when I had to choose between being myself or predicting myself, or choosing to be more helpful, I chose to help, be helpful, I wanted to come out of my comfort zone and challenge myself more, I didn’t really feel a shift but I wanted to stretch myself more and help a little more.”(P15)

Merlynnne offered opportunities to practice applying CBT in a short, engaging, manner to build perceived competence in helping others (P2) and most participants said they would adopt using cognitive reappraisals (29/36). P2 reported that the skills they developed in-game might translate into self-help:

“It’s tough to know when you’re being overly negative of yourself, right? But I, maybe I would now.” (P2)

**5.3.4 Selective Rejection of Stereotypes.** The majority of players did not feel their avatar influenced the way they responded ( $n=22$ ), but a third of participants felt their avatar appearance affected their mindset and consequential behavior ( $n=10$ ). Players acknowledged intended stereotypes of the CLERIC ( $n=13$ ) and MONSTER ( $n=17$ ), but several in the MONSTER group rejected negative stereotypes citing visual conflicts with the positive task of helping others ( $n=7$ ). P17 explains that since the CLERIC avatar’s image is seen as more “confident and caring” (P17), they should also provide good advice ( $n=2$ ):

“... if it had instead been like a clown or something like that I might have been tempted to say like silly things or be more joking but she seems sort of like, she seems like the kind of person who would give good advice so then I thought I should give good advice.” (P17, CLERIC)

Most participants who played as the CLERIC described imagining themselves as the character (14/18), adopting their perspective in the game. The MONSTER on the other hand was described as an interaction tool and not an embodiment of the player (9/18). One explanation for this difference was the misfit between the MONSTER’s appearance which implies aggression and violence, and the supportive tasks the player was performing. P20 explained how they consciously separated themselves from the MONSTER avatar:

“I kind of separated it. I want to beat the game so the minotaur will get me there, then as myself I want to help these people so it felt like me talking a real people person to person, not a minotaur talking to a villager.” (P20, MONSTER)

This separation aligns with descriptions of helping behaviours in commercial games, where assisting other players with real life problems was separate from the game. P17 describes player information as different from a game:

“...confined to the game world then I’d go with whatever makes sense for the game world, but if I get more information and some real world thing happening, then I would shift my perspective.” (P17, CLERIC)

Although many players in the MONSTER group rejected their character’s stereotype (10/18), few reported the avatar would have negatively influenced their responses ( $n=2$ ). P4 chose to reject the MONSTER avatar expecting its ‘rough’ visual influence to be nonconstructive:

“I was like hey maybe I should play like this minotaur and get more like I don’t know roughish and strong and, like uh you know, helpful answers, but I was like oh that’s not a – that’s not very helpful.” (P4, MONSTER)

Some players instead re-imagined the MONSTER avatar into a positive archetype to fit the task ( $n=2$ ):

“I am kind of just imagining him as like a sports coach that is like yelling out, and encouraging people, when people are like working out, or like goading things, like, ‘yeah you can do it yeah! Keep going!’” (P5, MONSTER)

These findings are suggestive that the game-task disconnect might have dampened the Proteus Effect when players have personal relevance to the task. This disconnect was also repeatedly mentioned to break immersion ( $n=10$ ). This phenomenon may be due to ludonarrative dissonance where the players’ in-game activity is not in alignment with the game’s narrative [71].

## 6 DISCUSSION & IMPLICATIONS

Previous work in promoting peer-to-peer mental health support focused on persuading individuals to seek help and subsequent therapeutic success. *Merlynnne* focuses on persuading emotional helping behaviour from those who do not usually help others online. In this section, we reflect on our game design, challenges and successes in motivating peer-to-peer support for mental health, and the challenges we faced in replicating the Proteus Effect in this context.

### 6.1 Motivating Emotional Labour

The results of our exploratory study suggest that *Merlynnne*’s design incorporated serious game elements like progression feedback, emotional investment, and meaningful interactions as peripheral cues within the ELM to successfully motivate peer-to-peer CBT. More than half of participants specifically mentioned in interviews being motivated by game elements and presentation such as ability to move around, music, and art (19/36), and more than a third of participants (16/36) said their main motivation was to achieve in the game; to advance the narrative, obtain cosmetic items, or simply progress to complete it.

Moreover, these serious game design elements did not impair peer-to-peer CBT activity. The majority of participants (27/36) reported prioritizing helpfulness of their responses due to personal relevance or desired positive impact on others, suggesting that central processing within the ELM occurred as players thought deeply about the messaging. Most responses successfully reframed the INT using empathy and encouragement. Overall, these findings point to *Merlynnne* facilitating player understanding of mental health cultures, and motivating genuine intentions to help via peer-to-peer CBT.

An emerging theme from interview data was an increased willingness to engage with queries in *Merlynnne* than on online forums. HAS scores also significantly increased post-play for all

participants, as players reported a higher willingness to help others after playing *Merlynnne*. Further, participants described an expectation for sensitivity and honesty in *Merlynnne* absent on Reddit, meaning that while satire was culturally expected on some subreddits, it was impermissible in *Merlynnne* — despite content coming from the same source. Participants also expressed a desire to keep up this activity, even after the game had stopped, with nearly all participants saying they will use the format in daily life. This result was an indicator of attitude change in Morris’s Panoply [48], suggesting *Merlynnne* achieved the same.

## 6.2 Design Challenges

While the results of our study are largely positive, it also revealed several potential pitfalls that warrant consideration. In particular, the higher motivation to contribute peer support without adequate ability presents risks for both supporter and seeker, such as potential burnout and harmful advice.

**6.2.1 Fatigue.** Our results suggest a potential decline in participants’ response effectiveness over time, even though their attitudes towards helping did not. That is, scores for empathy, proposed solutions, encouragement, and not damaging — which play an important role in therapeutic outcomes [8] — decreased over time, and participants reported that the emotional labour of CBT felt ‘draining’ in post-study interviews. But, participants’ attitudes towards helping (HAS scores) did not decrease. Together, these findings suggest that designers need to consider a player’s ability to engage with CBT content over extended periods of time. Potential solutions include designing for shorter sessions [48], or including more varied gameplay such as map exploration or more avatar customization opportunities as a form of breaks.

**6.2.2 Need for Self-Expression.** Peer supporters’ willingness to conform to the CBT format is also an important consideration, as 19/36 of our participants resisted using the CBT format, despite acknowledging its merit for mental health support, and reported feeling that helping others is ‘intuitive’ and that they could provide effective support without the need for guided responses. Eight participants (8/36) also felt that action-oriented and direct advice was more helpful than elements of the CBT format like empathy. While most responses were overwhelmingly positive, and more than 95% ( $n = 426$ ) were rated harmless, we did see potential for harm. For instance, one response from P10 sarcastically promoted suicide, an attempt at humour, which propositions harm in practice. Given the potential for harm, human moderation could remain necessary in peer-to-peer support systems [25, 48]. However this moderation needs to be considered within the context of trade-offs; research has found that limiting player expression and over-structured formats can limit reliability [47, 53], and that humour can facilitate patient community building [23].

**6.2.3 Need for Relatedness.** Related to the need for self-expression was a need for participants to relate to those they were supporting and to help them through shared experiences. Participants reported going out of their way to find NPCs with queries that they could relate to and had difficulty responding to those with whom they shared little common ground. Players also reported in post-study interviews that NPC appearance did not influence their predisposition to give help, 22/36 of our participants included their own personal anecdotes in their responses, and a third of responses contained unprompted practical solutions based on their personal experience ( $n = 149$ ,  $M = 33.33\%$ ). These results suggest that matching players with peers with similar experiences will improve engagement and provide opportunities for genuine empathy, which echoes the social needs of students in mental health technologies [40]. However, care needs to be taken in such a matchmaking process, since segregation based on mental health experience is discouraged and a diversity is considered a strength of peer support communities [47].

**6.2.4 Ludonarrative Dissonance.** Players were motivated by both narrative and the prospect of helping others, yet they described the two motivations as existing in separate dimensions. Although the fantasy narrative is used by well-researched platforms like *SPARX*, *Habitica*, and *SuperBetter*, it may not be conducive to tasks that require applying personal relevance for interpersonal interactions, as immersion can be broken when the game world is interrupted with real world elements. Thus, designers should consider the use of narratives that are personally relevant for their audience [65], such as a school simulation for student players.

### 6.3 Design Implications: The Proteus Effect

We were not able to replicate the Proteus Effect in the context of peer-to-peer mental health support, despite carefully designing our game and study to elicit the effect. We exclusively recruited male participants, building on work that has found an increased effect on male players [41, 78]. We iteratively designed the two avatars, play tested their designs, and consulted with linguistic experts to ensure they were perceived differently by participants. In post-study interviews, participants described avatars using gendered archetypes that align with previous work (e.g., [76–78]). Despite encouraging differences in avatar identification reports, we found no significant differences in HAS scores, responses between AVATAR groups. We interpret these results as indicative that our CLERIC and MINOTAUR avatar designs were perceived differently by participants, but that these differences did not translate into different in-game behaviours.

However, our qualitative results point to some differences in how the two AVATAR groups engaged with the in-game content. Our interview data suggests that participants in the MONSTER group were playing as themselves independent of avatar influence, whereas those in the CLERIC group assumed the avatar’s persona. CLERICS also identified more strongly with their avatar, evidenced by a higher connectedness and likelihood of referring to their avatar as ‘we’ than those in the MONSTER group. Furthermore, participants reported actively rejecting the antagonistic stereotype; some viewing the MONSTER avatar as a helpful sports coach, rather than a monster.

Our results also point to several limitations of our game design that might have influenced the Proteus Effect. Ludonarrative dissonance is a potential mitigating factor; despite the game itself being engaging, the mismatch between modern student issues and the high fantasy genre may have limited immersion. Similarly, *Merlynn*’s 8-bit graphics and limited avatar customizability may have impacted the degree to which participants related to their avatars. Only one participant had an initial low HAS score, thus the demographics of our participant pool could have been a limitation of our methods.

Overall, our results show that a serious game can be an effective way to engage peers, but that influencing emotional labour through the Proteus Effect may be difficult to accomplish. These findings support the complexity and ‘wickedness’ of motivating peer-to-peer CBT, and of the design space for our serious game, and help to contextualize the need for more research through design [37, 80].

## 7 LIMITATIONS

As with any study that seeks to explore a complex design space like serious games for mental health, ours has limitations. Two in particular should be acknowledged. First, our study included only male participants. This decision was driven by our objective to improve male participation in peer-to-peer support with CBT. Thus, our game design, and in particular choice of avatars, may not generalize to other genders or underrepresented masculinities [62]. The feminine CLERIC avatar may have also conflicted with our participants’ male gender identity, limiting the connectedness between player and avatar needed to elicit the Proteus Effect [19]. Future work should consider

alternative archetypes of helpfulness, like animals, to determine if similar effects translate beyond our participant group.

Second, our exploratory study examined *Merlynnne* through an in-lab simulation of peer-to-peer support, where two-way support was deliberately not included. We considered experiments with two-way CBT, but ultimately decided on one-way interactions to safeguard potential end-users from harmful advice. These decisions were made after extensive consultation with our institution's REB, reflection on previous work, and careful consideration of how we could explore our research questions.

Given these constraints, our study provides a useful proxy for understanding how serious games might motivate participation in peer-to-peer CBT. During post-study interviews, our participants commented that they felt that since their responses were being used for service improvements, they were making important contributions. We suggest that this level of enthusiasm would likely translate to a peer-to-peer setting. However, it's also possible that the nature of their responses could change when directed to a person rather than a (inhuman) response bank. For instance, they may be more likely to empathize, encourage, or propose a real solution if they felt it would directly benefit somebody. They might also be even more willing to perform emotional labour in these settings.

We envision that, in practice, a final CBT system would function like *Panopoly* [48] or its successor *Kokobot* [49], where help seekers submit posts and peers respond directly to them. In building such a system, future work must first address the above limitations by better understanding how the Proteus Effect might play out with a diverse set of peer supporters, and establishing how two-way CBT might influence peers' motivations. Finally, future work should study the long-term retention of peer-support habits with gamification and the integration of applications like *Merlynnne* into existing health systems, communities of practice [75], or industry markets [37].

## 8 CONCLUSIONS

We developed *Merlynnne*, a serious game that motivates peer-to-peer CBT. In particular, this work explores how game elements borrowed from fantasy RPGs can be leveraged to motivate persons without mental health service experience to offer cognitive reappraisals to others. The results of our user study revealed high engagement by players and a significantly higher willingness to help others post-play, but also identified barriers to facilitating emotional labour within a game experience. We also tried to replicate the Proteus Effect in the context of motivating mental health peer support, but were unable to do so our study.

Overall, our results show that serious games can be effective in promoting participation in online peer-to-peer support platforms, and in motivating emotional labour and empathetic behavior. We also identified barriers to their use, and challenges for future work in this space. Our work is an important first step in establishing benefits of gameful design in the mental health space with a focus on motivating peer *supporter* participation, where previous work had focused on the help seeker's experience.

## 9 ACKNOWLEDGEMENTS

We thank our local mental health experts and school boards for engagement in the conceptual design of study. We also thank Monica Ziska and Calciumtrice providing open source art. We also thank the University of Waterloo Games Institute, and its members for their constructive feedback throughout *Merlynnne*'s development. Finally, this work was made possible by NSERC Discovery Grant 2015-06585.

## REFERENCES

- [1] William and Irwin and Gregory Bassham. 2003. Depression, Informal Fallacies, and Cognitive Therapy. *Inquiry: Critical Thinking Across the Disciplines* 21, 3 (2003), 15–21. <https://doi.org/10.5840/inquiryctnews200321335>
- [2] Gerhard Andersson. 2014. *The internet and CBT: a clinical guide*. CRC Press, Linköping, Sweden. 154 pages.
- [3] Susan J Andrae, Lynn J Andrae, Joshua S Richman, Andrea L Cherrington, and Monika M Safford. 2020. Peer-delivered cognitive behavioral training to improve functioning in patients with diabetes: a cluster-randomized trial. *The Annals of Family Medicine* 18, 1 (2020), 15–23.
- [4] Erin Ash. 2015. Priming or Proteus Effect?: Examining the Effects of Avatar Race on In-Game Behavior and Post-Play Aggressive Cognition and Affect in Video Games. *Games and Culture* 11 (02 2015). <https://doi.org/10.1177/1555412014568870>
- [5] Titus Barik, Emerson Murphy-Hill, and Thomas Zimmermann. 2016. A perspective on blending programming environments and games: Beyond points, badges, and leaderboards. In *2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, Cambridge, UK, 134–142. <https://doi.org/10.1109/VLHCC.2016.7739676>
- [6] Daniel Robert Bateman, Erin Brady, David Wilkerson, Eun-Hye Yi, Yamini Karanam, and Christopher M Callahan. 2017. Comparing Crowdsourcing and Friendsourcing: A Social Media-Based Feasibility Study to Support Alzheimer Disease Caregivers. *JMIR Research Protocols* 6, 4 (2017), e56. <https://doi.org/10.2196/resprot.6904>
- [7] Aaron Beck. 2007. <https://beckinstitute.org/does-cognitive-therapy-cognitive-behavior-therapy/>
- [8] A. T. Beck, A. J. Rush, B. F. Shaw, and G. Emery. 1979. *Cognitive Theory of Depression*. The Guilford Press, New York, NY.
- [9] Melissa F. Bernstein, Samantha E. Cinnick, Emily Franzosa, Marita K. Murrman, and Nicholas Freudenberg. 2019. Rationale and Design of Distance-Based Training to Persuade Local Health Department Employees that Addressing Social Determinants of Health Is Their Job, Too. *Journal of Public Health Management and Practice* 25, 5 (2019), 448–453. <https://doi.org/10.1097/phh.0000000000000827>
- [10] Max V. Birk and Regan L. Mandryk. 2018. Combating Attrition in Digital Self-Improvement Programs Using Avatar Customization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI '18*). ACM, New York, NY, USA, Article 660, 15 pages. <https://doi.org/10.1145/3173574.3174234>
- [11] BotHub. 2016. KokoBot. <https://bot-hub.com/bots/KokoBot>. Accessed: 2019-08-30.
- [12] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (jan 2006), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [13] Virginia Braun and Victoria Clarke. 2017. Thematic analysis - The University of Auckland. <https://www.psych.auckland.ac.nz/en/about/our-research/research-groups/thematic-analysis.html>
- [14] June S.L. Brown, Sara Evans-Lacko, Lisa Aschan, Max J. Henderson, Stephani L. Hatch, and Matthew Hotopf. 2014. Seeking informal and formal help for mental health problems in the community: A secondary analysis from a psychiatric morbidity survey in South London. *BMC Psychiatry* 14, Article 275 (2014), 15 pages. Issue 1. <https://doi.org/10.1186/s12888-014-0275-y>
- [15] Amy Bruckman. 2002. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology* 4, 3 (2002), 217–231. <https://doi.org/10.1023/A:1021316409277>
- [16] Jane Burns, Marianne Webb, Lauren A Durkin, and Ian Hickie. 2010. Reach Out Central: a serious game designed to engage young men to improve mental health and wellbeing. *The Medical journal of Australia* 192 (06 2010), S27–30.
- [17] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 981â–š992. <https://doi.org/10.1145/2858036.2858498>
- [18] Long Ting Chan and James R. Wallace. 2018. Changing Peer Support Attitudes with Avatar-based Gamification. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI EA '18*). ACM, New York, NY, USA, Article LBW566, 5 pages. <https://doi.org/10.1145/3170427.3188497>
- [19] Colleen Clark, Blake Barrett, Autumn Frei, and Annette Christy. 2016. What makes a peer a peer. *Psychiatric Rehabilitation Journal* 39, 1 (2016), 74–76. <https://doi.org/10.1037/prj0000147>
- [20] J Cohen. 1988. *Statistical power analysis for the behavioral sciences, 2nd edn*. Erbaum Press, Hillsdale, NJ, USA.
- [21] John W. Creswell and Vicki L. Plano Clark. 2018. *Designing and Conducting Mixed Methods Research* (4th ed.). SAGE, 2455 Teller Road, Thousand Oaks, California.
- [22] Kathryn P Davison, James W Pennebaker, and Sally S Dickerson. 2000. Who Talks? The Social Psychology of Illness Support Groups. *American Psychologist* 55, 2 (2000), 205–217. <https://doi.org/10.1037/0003-066X.55.2.205>
- [23] Zsófia Demjén. 2018. Complexity theory and conversational humour: Tracing the birth and decline of a running joke in an online cancer support community. *Journal of Pragmatics* 133 (2018), 93–104. <https://doi.org/10.1016/j.pragma.2018.06.001>
- [24] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. 2011. Gamification. Using Game-Design Elements in Non-Gaming Contexts. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*

- (Vancouver, BC, Canada) (*CHI EA '11*). Association for Computing Machinery, New York, NY, USA, 2425a–2428. <https://doi.org/10.1145/1979742.1979575>
- [25] Bryan Dosono and Bryan Semaan. 2019. Moderation Practices As Emotional Labor in Sustaining Online Communities: The Case of AAPI Identity Work on Reddit. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (*CHI '19*). ACM, New York, NY, USA, Article 142, 13 pages. <https://doi.org/10.1145/3290605.3300372>
- [26] Anders Drachen, Pejman Mirza-Babaei, and Lennart E. Nacke. 2018. *Games user research*. Oxford University Press, New York, NY. 1–524 pages. <https://doi.org/10.1093/oso/9780198794844.001.0001>
- [27] R. A. Dunn and Rosanna E. Guadagno. 2010. Playing Me: A Study of Gender, Gameplay, and Game Preference Effects on Avatar Selection. In *ICA 2010 Singapore*. International Communication Association, Washington, DC, USA, 1–33.
- [28] Men's Health Forum. 2014. *How to make mental health services work for men*. Technical Report. Centre for Men's Health at Leed's Beckett University, England, UK. [https://www.menshealthforum.org.uk/sites/default/files/pdf/how\\_to\\_mh\\_v4.1\\_lrweb\\_0.pdf](https://www.menshealthforum.org.uk/sites/default/files/pdf/how_to_mh_v4.1_lrweb_0.pdf)
- [29] Itzhak Gilat and Eyal Reshef. 2015. The perceived helpfulness of rendering emotional first aid via email. *British Journal of Guidance & Counselling* 43, 1 (2015), 94–104. <https://doi.org/10.1080/03069885.2014.909006> arXiv:<https://doi.org/10.1080/03069885.2014.909006>
- [30] Greg Gillespie and Darren Crouse. 2012. There and back again: Nostalgia, art, and ideology in old-school dungeons and dragons. *Games and Culture* 7, 6 (2012), 441–470. <https://doi.org/10.1177/1555412012465004>
- [31] John Harris and Mark Hancock. 2018. Beam Me 'Round, Scotty! II: Reflections on Transforming Research Goals into Gameplay Mechanics. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (Melbourne, VIC, Australia) (*CHI PLAY '18 Extended Abstracts*). ACM, New York, NY, USA, 193–204. <https://doi.org/10.1145/3270316.3273039>
- [32] Donald Hedeker and Robert D Gibbons. 2006. *Longitudinal data analysis*. Vol. 451. John Wiley & Sons, Hoboken, New Jersey.
- [33] Stefan G Hofmann, Anu Asnaani, Imke J J Vonk, Alice T Sawyer, and Angela Fang. 2012. The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognitive therapy and research* 36, 5 (oct 2012), 427–440. <https://doi.org/10.1007/s10608-012-9476-1>
- [34] Anne Holohan, Justyna Pyz, and Kamila Trochowska. 2018. Gaming for peace: Gender awareness training and the polish military. *Security and Defence Quarterly* 21, 4 (2018), 41–57. <https://doi.org/10.5604/01.3001.0012.5179>
- [35] Marvin Alexander Kalde. 2015. Behavioural changes through serious health games : a viable alternative to traditional treatments / techniques ?. In *6th IBA Bachelor Thesis Conference, November 5th, 2015*. University of Twente, Enschede, The Netherlands, 12.
- [36] Eva Kaltenthaler, Glenys Parry, Catherine Beverley, and Michael Ferriter. 2008. Computerised cognitive-behavioural therapy for depression: Systematic review. *British Journal of Psychiatry* 193, 3 (2008), 181–184. <https://doi.org/10.1192/bjp.bp.106.025981>
- [37] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to Evaluate Technologies for Health Behavior Change in HCI Research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). Association for Computing Machinery, New York, NY, USA, 3063a–3072. <https://doi.org/10.1145/1978942.1979396>
- [38] Wendy Kliewer, Stephen J. Lepore, Alex Broquet, and Linda Zuba. 1990. Developmental and gender differences in anonymous support-seeking: Analysis of data from a community help line for children. *American Journal of Community Psychology* 18, 2 (1990), 333–339. <https://doi.org/10.1007/BF00931308>
- [39] Suresh Kp. 2011. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of human reproductive sciences* 4 (03 2011), 8–11. <https://doi.org/10.4103/0974-1208.82352>
- [40] Emily G. Lattie, Rachel Kornfield, Kathryn E. Ringland, Renwen Zhang, Nathan Winquist, and Madhu Reddy. 2020. Designing Mental Health Technologies That Support the Social Ecosystem of College Students. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3313831.3376362>
- [41] Mika Lehdonvirta, Yosuke Nagashima, Vili Lehdonvirta, and Akira Baba. 2012. The stoic male: How avatar gender affects help-seeking behavior in an online game. *Games and Culture* 7 (2012), 29–47. Issue 1. <https://doi.org/10.1177/1555412012440307>
- [42] Louise Lynch, Maggie Long, and Anne Moorhead. 2018. Young Men, Help-Seeking, and Mental Health Services: Exploring Barriers and Solutions. *American journal of men's health* 12, 1 (jan 2018), 138–149. <https://doi.org/10.1177/1557988315619469>
- [43] Andrzej Marczewski. 2015. Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Desig. In *Gamified UK*. CreateSpace Independent Publishing Platform, Scotts Valley, California, 65–80. <http://www.amazon.co.uk/Even-Ninja-Monkeys-Like-Play/dp/1514745666/>

- [44] Annette Markham. 2012. Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication & Society* 15, 3 (2012), 334–353. <https://doi.org/10.1080/1369118X.2011.641993>
- [45] Jim McCambridge, John Witton, and Diana R. Elbourne. 2013. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology* 67 (11 2013). <https://doi.org/10.1016/j.jclinepi.2013.08.015>
- [46] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. In *Proc. ACM Hum.-Comput. Interact.* 39. ACM, New York, NY, USA, 23.
- [47] Shery Mead, David Hilton, and Laurie Curtis. 2001. Peer support: A theoretical perspective. *Psychiatric Rehabilitation Journal* 25, 2 (2001), 134–141. <https://doi.org/10.1037/h0095032>
- [48] Robert R Morris. 2015. *Crowdsourcing Mental Health and Emotional Well-Being*. Ph.D. Dissertation. Massachusetts Institute of Technology.
- [49] Robert R. Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M. Schueller. 2018. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. *Journal of medical Internet research* 20, 6 (2018), e10148. <https://doi.org/10.2196/10148>
- [50] Scott Nicholson. 2015. A RECIPE for Meaningful Gamification. In *Gamification in Education and Business*. Springer International Publishing, Cham, 1–20. [https://doi.org/10.1007/978-3-319-10208-5\\_1](https://doi.org/10.1007/978-3-319-10208-5_1)
- [51] Gary S Nickell. 1998. The Helping Attitude Scale Gary. *Annual Convention of the American Psychological Association* 56563, 218 (1998), 1–10.
- [52] Tunu Nkunya, Thinavan Periyayya, and Santhidran Sinappan. 2017. Enhancing the effectiveness of tobacco package warning labels: An elaboration likelihood perspective. *SEARCH (Malaysia)* 9, 2 (2017), 29–54.
- [53] Kathleen O’Leary, Stephen M. Schueller, Jacob O. Wobbrock, and Wanda Pratt. 2018. “Suddenly, We Got to Become Therapists for Each Other”: Designing Peer Support Chats for Mental Health. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI ’18). ACM, New York, NY, USA, Article 331, 14 pages. <https://doi.org/10.1145/3173574.3173905>
- [54] World Health Organization et al. 2013. *Mental Health Action Plan 2013-2020*. Technical Report. World Health Organization. 1–44 pages. arXiv:NIHMS150003 [http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021\\_jeng.pdf?ua=1%5Cnhttp://apps.who.int/iris/bitstream/10665/89966/1/9789241506021\\_jeng.pdf](http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_jeng.pdf?ua=1%5Cnhttp://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_jeng.pdf)
- [55] Michelle Ortlipp. 2008. Keeping and Using Reflective Journals in the Qualitative Research Process. *Qualitative Report* 13, 4 (2008), 695–705. [http://csaweb110v.csa.com/ids70/view{}\\_record.php?id=3{&}recnum=83{&}log=from{ }res{&}SID=7nds10lqpq9t0qbm6m905dier3{&}mark{ }\\_jid=search{ }3A3{ }3A83{ }2C80{ }2C90](http://csaweb110v.csa.com/ids70/view{}_record.php?id=3{&}recnum=83{&}log=from{ }res{&}SID=7nds10lqpq9t0qbm6m905dier3{&}mark{ }_jid=search{ }3A3{ }3A83{ }2C80{ }2C90)
- [56] Richard E. Petty and John T. Cacioppo. 1986. The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology* 19, C (1986), 123–205. [https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)
- [57] Paul N Pfeiffer, Brooke Pope, Marc Houck, Wendy Benn-Burton, Kara Zivin, Dara Ganoczy, H Myra Kim, Heather Walters, Lauren Emerson, C Beau Nelson, et al. 2020. Effectiveness of peer-supported computer-based CBT for depression among veterans in primary care. *Psychiatric services* 71, 3 (2020), 256–262.
- [58] L.-M. Putz and H. Treiblmaier. 2015. Creating a theory-based research agenda for gamification. In *2015 Americas Conference on Information Systems, AMCIS 2015*, Vol. 4. Association for Information Systems, Puerto Rico, Article 12, 13 pages.
- [59] Horst W.J. Rittel and Melvin M. Webber. 1973. Dilemmas in a general theory of planning. *Policy Sciences* 4, 2 (1973), 155–169. <https://doi.org/10.1007/BF01405730>
- [60] A M Roepke, J McGonigal, and R Broome. 2013. Getting SuperBetter: Using gameful technology to improve well-being. In *Third World Congress on Positive Psychology, June 27-30, 2013, Los Angeles, CA*. International Postive Psychology Association, Apple Valley, MN, USA, 110. <https://doi.org/10.1037/e574802013-268>
- [61] Sigalit Ronen and Ayala Malach Pines. 2008. Gender differences in engineers’ burnout. *Equal Opportunities International* 27, 8 (2008), 677–691. <https://doi.org/10.1108/02610150810916749>
- [62] Zac E. Seidler, Simon M. Rice, Jo River, John L. Oliffe, and Haryana M. Dhillon. 2018. Men’s Mental Health Services: The Case for a Masculinities Model. *Journal of Men’s Studies* 26, 1 (2018), 92–104. <https://doi.org/10.1177/1060826517729406>
- [63] Brett Sherrick, Jennifer Hoewe, and T. Franklin Waddell. 2014. The role of stereotypical beliefs in gender-based activation of the Proteus effect. *Computers in Human Behavior* 38 (sep 2014), 17–24. <https://doi.org/10.1016/J.CHB.2014.05.010>
- [64] Merry S.N., Stasiak K., Shepherd M., Frampton C., Fleming T., and Lucassen M.F.G. 2012. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: Randomised controlled non-inferiority trial. [http://www.bmj.com/highwire/filestream/579896/filestream{}\\_highwire{}\\_article{}\\_pdf/0.pdf?%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS{&}PAGE=reference{&}D=emed10{&}NEWS=N{&}AN=2012292499](http://www.bmj.com/highwire/filestream/579896/filestream{}_highwire{}_article{}_pdf/0.pdf?%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS{&}PAGE=reference{&}D=emed10{&}NEWS=N{&}AN=2012292499)
- [65] Jane Speedy. 2000. The ‘storied’ helper: narrative ideas and practices in counselling and psychotherapy. *European Journal of Psychotherapy & Counselling* 3, 3 (2000), 361–374. <https://doi.org/10.1080/13642530010012011>

- [66] Bruce Struminger, Sanjeev Arora, Sarah Zalud-Cerrato, David Lowrance, and Tedd Ellerbrock. 2017. Building virtual communities of practice for health. *The lancet* 390, 10095 (2017), 632–634.
- [67] Abhay Sukumaran, Stephanie Vezich, Melanie McHugh, and Clifford Nass. 2011. Normative Influences on Thoughtful Online Participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). Association for Computing Machinery, New York, NY, USA, 3401–3410. <https://doi.org/10.1145/1978942.1979450>
- [68] Na Sun, Patrick Pei Luen Rau, and Liang Ma. 2014. Understanding lurkers in online communities: A literature review. *Computers in Human Behavior* 38 (2014), 110–117. <https://doi.org/10.1016/j.chb.2014.05.022>
- [69] Alexander Tallner, Klaus Pfeifer, and Mathias Maurer. 2016. Web-based interventions in multiple sclerosis: The potential of tele-rehabilitation. , 327–335 pages. <https://doi.org/10.1177/1756285616640684>
- [70] Debbe Thompson, Tom Baranowski, Richard Buday, Janice Baranowski, Victoria Thompson, Russell Jago, and Melissa Juliano Griffith. 2010. Serious video games for health: How behavioral science guided the development of a serious video game. *Simulation and Gaming* 41, 4 (2010), 587–606. <https://doi.org/10.1177/1046878108328087>
- [71] Weimin Toh and Weimin Toh. 2018. Ludonarrative Dissonance. In *A Multimodal Approach to Video Games and the Player Experience*. Routledge, New York, 49–74. <https://doi.org/10.4324/9781351184779-4>
- [72] Gustavo F. Tondello and Lennart E. Nacke. 2018. Gamification: Tools and Techniques for Motivating Users. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI EA '18*). ACM, New York, NY, USA, Article C25, 4 pages. <https://doi.org/10.1145/3170427.3170662>
- [73] Sarah F. Van Der Land, Alexander P. Schouten, Frans Feldberg, Marleen Huysman, and Bart van den Hooff. 2015. Does avatar appearance matter? How team visual similarity and member-avatar similarity influence virtual team performance. *Human Communication Research* 41, 1 (2015), 128–153. <https://doi.org/10.1111/hcre.12044>
- [74] Jan B. Vornhagen, April Tyack, and Elisa D. Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Ottawa, Ontario, Canada) (*CHI PLAY '20*). Association for Computing Machinery, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3410404.3414229>
- [75] Etienne Wenger, Richard Arnold McDermott, and William Snyder. 2002. *Cultivating communities of practice: A guide to managing knowledge*. Harvard business press, Boston, Massachusetts, USA.
- [76] Nick Yee and Jeremy Bailenson. 2007. The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research* 33, 3 (2007), 271–290. <https://doi.org/10.1111/j.1468-2958.2007.00299.x>
- [77] Nick Yee, Nicolas Ducheneaut, and Les Nelson. 2012. Online Gaming Motivations Scale: Development and Validation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (*CHI '12*). ACM, New York, NY, USA, 2803–2806. <https://doi.org/10.1145/2207676.2208681>
- [78] Nick Yee, Nicolas Ducheneaut, Mike Yao, and Les Nelson. 2011. Do Men Heal More when in Drag?: Conflicting Identity Cues Between User and Avatar. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). ACM, New York, NY, USA, 773–776. <https://doi.org/10.1145/1978942.1979054>
- [79] Gunwoo Yoon and Patrick T. Vargas. 2014. Know Thy Avatar: The Unintended Effect of Virtual-Self Representation on Behavior. *Psychological Science* 25, 4 (2014), 1043–1045. <https://doi.org/10.1177/0956797613519271>
- [80] John Zimmerman and Jodi Forlizzi. 2014. Research through design in HCI. In *Ways of Knowing in HCI*. Springer New York, New York, NY, 167–189. [https://doi.org/10.1007/978-1-4939-0378-8\\_8](https://doi.org/10.1007/978-1-4939-0378-8_8)

Received February 2021 ; revised September 2021 ; accepted July 2021