This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3270725

Measuring Embodiment: Movement Complexity and the Impact of Personal Characteristics

Tabitha C. Peck, *Senior Member, IEEE,* and Jessica J. Good

Abstract—A user's personal experiences and characteristics may impact the strength of an embodiment illusion and affect resulting behavioral changes in unknown ways. This paper presents a novel re-analysis of two fully-immersive embodiment user-studies (n=189 and n=99) using structural equation modeling, to test the effects of personal characteristics on subjective embodiment. Results demonstrate that individual characteristics (gender, participation in science, technology, engineering or math – Experiment 1, age, video gaming experience – Experiment 2) predicted differing self-reported experiences of embodiment Results also indicate that increased self-reported embodiment predicts environmental response, in this case faster and more accurate responses within the virtual environment. Importantly, head-tracking data is shown to be an effective objective measure for predicting embodiment, without requiring researchers to utilize additional equipment.

Index Terms—Embodiment, Gender, Age, STEM, Structural Equation Modeling, Avatars, Behavior.

1 Introduction

The positive impact of self-avatars in immersive virtual environments (VEs) has been demonstrated in numerous ways including users having improved distance estimation [1], [2], reduced cognitive load [3], [4], and an increased sense of presence [5]. This positive impact of self-avatars may be due to virtual sense of embodiment–when the user feels that the self-avatar has effectively replaced their own-body at a physical and functional level [6], [7]. When users embody their self-avatar they even take on stereotypical behaviors based on the avatar's appearance such as negotiating more aggressively when in a taller body [8] or having improved cognitive function when looking like Einstein [9].

In the examples listed above, behavior change may be impacted by the user's level of embodiment such that higher embodiment in the self-avatar leads to a greater adoption of stereotypical self-avatar behavior. However, measuring an individual's level of embodiment is challenging and the subjective level of embodiment is often high regardless of avatar appearance when wearing a co-located avatar that moves synchronously with the user [10]. For example, self-reported embodiment levels, measured with condensed embodiment questionnaires, are similar regardless of wearing an avatar of a different race [11] or age [12] when compared to the user's identity. The relationship between behavior change and embodiment is not well understood and may become more apparent when using a longer embodiment questionnaire that provides more nuanced data to better discern higher or lower levels of embodiment [13]. In fact, a few studies have found a relationship between the subjective level of embodiment and increased performance including studies on brain-computer interfaces [14], improved distance estimation [15], and a greater connection to nature, [16]. However, the differing ways (mechanisms) in which sense of embodiment may impact users' responses within virtual environments is not yet fully understood.

Further, users have different individual characteristics that may affect how embodied they become in different avatars within different situations. For example, an individual's locus of control—how much they believe they are in control of situations and experiences—may affect their level of embodiment [17]. There are many individual characteristics that may affect embodiment including, but not limited to, gender, age, or gaming experience. Since embodiment may lead to improved virtual experiences, both cognitively and experientially, it is important to understand factors that may impact an individual's level of embodiment. Understanding how individual characteristics affect embodiment can aid researchers in designing optimal virtual scenarios that foster embodiment for each user. Demonstrating that individual characteristics affect embodiment supports the necessity of studying diverse populations [18].

In order to understand how individual characteristics affect embodiment, effective measurement of embodiment is imperative. Numerous subjective and objective tools have been proposed including questionnaires [13], electroencephalography (EEG) [19], skin-conductance [20], and heart-rate monitors [21]. Questionnaires are the most common of the measurement tools since they are easy to administer, do not require users to wear additional cumbersome equipment, and do not require environment specific responses to threatening situations [6]. However, subjective responses in questionnaires are prone to user bias [22] and it is important to support subjective responses with objective measures. Problematically, there are currently no objective, easy-to-collect embodiment measures that are environment independent and do not require users to wear additional hardware.

- *T. C. Peck and J. J. Good are with Davidson College, Davidson, NC, 28035. E-mail: tapeck@davidson.edu*

Manuscript received May 19, 2022; revised August 26, 2015.

In this work we revisit two embodiment experiments with data provided by the authors of those works [23], [24], [25]. We analyze these data using structural equation modeling (SEM), a statistical technique commonly used in the behavioral sciences. The previous multivariate analysis was limited to analysis of responses within the VE and only reported standard descriptive statistics on demographic data. Our new analysis 1) investigates the relationship between embodiment level and environment influence—a user's attitudinal and/or behavioral change caused by the VE scenario—to determine if increased embodiment leads to greater influence, 2) investigates the effect of user characteristics (gender, age, video gaming experience, and participation in Science, Technology, Engineering and Math (STEM)) on subjective embodiment, and 3) explores the efficacy of using head-tracking data as an objective measure to predict embodiment. We use the term environment influence in a broader sense to mean the user's behavior or actions compared to "Proteus Effect" [8] that refers to users demonstrating behaviors consistent with stereotypes associated with an avatar's social identity group. In Experiment 1, environmental influence does include the Proteus Effect; the outcome of interest is how participants take on characteristics stereotypically associated with the gender of their avatar. In Experiment 2 however, the Proteus Effect is not relevant as all participants are represented with race and gender matched avatars. The outcome of interest is their shooting behavior depending on target race and weapon status. Therefore, we use the term environment influence to broadly represent the user's behavior or actions.

We demonstrate that higher levels of measured embodiment led to higher levels of environment influence. Further, we identify individual characteristics associated with differing levels and submeasures of subjective embodiment and present proposed theoretical explanations for those associations. Finally, we present a new, easy to collect, objective measure that predicted subjective embodiment levels and user response to the environment.

The paper is organized as follows: Section 2 presents relevant background information. Section 3 presents the SEM methods used to evaluate the data from Experiment 1 (Section 4) and Experiment 2 (Section 5). Each experiment section includes a Results section that explains how the SEM was built, followed by a Discussion section that works through the model and discusses the results in relation to hypotheses and previous research. Finally, Section 6 provides concluding remarks.

2 Background

2.1 Embodiment

The plasticity of the human mind to accept an other-body or body-part as if it were an own-body was demonstrated in the real world where people accepted a rubber hand as if it were their own hand [26]. Acceptance over the hand can be induced through different forms of stimulation include visuotactile as well as active and passive movements [27]. The illusion was then extended to full-body illusions which were eventually demonstrated in VEs [28]. Levels of embodiment may be affected by the appearance of the avatar, the user's level of control over the avatar, and the user's point of view [29]. In addition to people accepting an other-body as if it were an own-body, the appearance of the self-avatar appears to impact user behavior [8] to provide the user with numerous cognitive and experiential benefits within the VE. For example, people with taller avatars act more confidently [8], people in dark-skinned avatars demonstrate reduced implicit bias [11], people embodying Einstein have higher cognitive performance [9], and self-avatars buffer people from stereotype threatening situations based on salient stereotypes [23], [24]. Self-avatars have been shown to increase trust and cooperation in collaborative tasks [30] and enable an innovative counseling system to give one's-self advice [31]. Self-avatars have been shown to improve spatial awareness and distance perception, [32], [33] and may dissuade users from walking through virtual walls such that the more present the user is in the VE the less likely they are to walk through a virtual wall [34].

Most research with self-avatars demonstrates the benefits or changes to participant behavior of having a self-avatar without directly investigating level of embodiment beyond high or low levels of embodiment. Fewer experiments have investigated the relationship between embodiment level and change in behavior, an exception being Gonzalez-Franco et al. [15] who found that higher levels of embodiment in participants directly predicted better distance estimation in the VE. Though it is generally believed that higher levels of embodiment lead to increased behavior change, the relationship between embodiment and behavior change is not well understood. For example, no minimum threshold has been defined for an embodiment illusion and it is not clear if there is a linear relationship between embodiment and behavior change. In particular, research is needed to understand how different submeasures of embodiment may impact different types of behavior.

2.2 Measuring Embodiment

Embodiment illusions are often measured with both objective and subjective assessments. Objective measures include physiometric measures such as skin conductance [35], heart rate [36], and Electroencephalogram (EEG) [19]. Additional objective measures look at participant response to a threat in the environment such as a saw moving toward the virtual body [37]. Response to an attack situation is an effective measure, however not all experiments lend themselves to attack scenarios and additional environment-independent objective measures may be useful.

Objective measures that do not require the participant to wear additional equipment are beneficial for participant usability and to reduce the financial burden on the experimenter. Head-tracking data is a common component of VR systems, is an easy to collect metric that is utilized in many VR experiments, and has been linked to user behavior. Kilteni et al. [38] found that participants in a casually dressed dark-skinned avatar had more complicated movements compared to participants in a formally-dressed light-skinned avatar. Additionally, Peck et al. [25] found that head and hand tracker data could predict implicit racial bias.

Embodied-cognition literature further supports the potential of using head-tracker data as an objective measure

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3270725

of virtual embodiment. Body movement has been shown to influence one's later choices and judgment, outside of one's awareness [39]. For example, when considering experimenter-directed movements, how people move their head can influence their acceptance of the environment. Wells and Petty [40] and Tom et al. [41] found that participants who were instructed to shake their heads vertically up and down were more favorable to a verbal message or objects in the scene compared to participants who shook their heads horizontally side-to-side. Further, participants nodded their heads vertically more quickly when they heard agreeable messages compared to hearing non-agreeable messages, and shook their heads horizontally more quickly when hearing a non-agreeable message [40]. A similar result was demonstrated in VR where participants who rotated their heads side-to-side were more likely to sign a petition [42]. Considering non-experimenter-directed movements, additional research on playing video games suggests that natural movements within a video game increase engagement [43] and that encouraging movement vs. prohibiting it further increases engagement [44]. This suggests that participant movement, both experimenter-directed and self-directed, may influence or signify embodiment levels.

In addition to objective measures, subjective measures are used to evaluate embodiment. Even though subjective questionnaires may not produce the most accurate results [22] they are easy to administer, do not require the participant to wear extra equipment, and do not require shocking or threatening environment situations. Because questionnaires are commonly used they enable comparison of results across different experiments. Gonzalez-Franco and Peck [6] collected the most used embodiment questions and proposed a first draft of a standardized embodiment questionnaire. This questionnaire was later refined to 16 questions and validated with nine experiments and over 400 participants [13].

However, social-desirability and personal identity can influence how people answer questionnaires [45]. Since embodiment questionnaires are commonly used it is important to understand how additional factors, such as participant gender or age, may affect subjective virtual embodiment. Comparing subjective ratings on questionnaires to VE-driven behavioral change can shed further insight into whether self-report measures are affected by social-desirability versus actual differences in embodiment levels.

2.3 User Characteristics and Embodiment

The embodiment literature shows that when users step into a self-avatar within a VE they become embodied within the identity of that self-avatar. However, users do not uniformly cast off their real-life identities and personal characteristics. Indeed, user characteristics such as personal and social identities or previous experiences may influence the extent to which they become embodied in a self-avatar.

Psychological research shows that social categories such as race, gender, and age are instantaneously perceived by others (as quickly as 200ms for racial categorization [46]) and carry important value for the self (as early as 3 months of age for gender [47]). A social identity is an aspect of the self that is defined in relation to others in a group. For example, being older or younger has no inherent meaning, but carries social significance because of the comparison to, or connection with, others. Participants in VR research do not enter the VE as a blank slate; they carry with them their social identities, which may influence the way in which they experience their new self-avatar. In other words, women may feel more or less embodied in a self-avatar compared to men, depending on the characteristics of the avatar and the particular VE. It is important to note that social identity categories are socially constructed such that the boundaries between groups are arbitrary and fluid. For example, although sex, gender, and gender identity exist on a continuum [48], within the United States people often self-identify within one of two binary categories (women and men). Racial categories are defined differently across the world and are revised throughout history. Despite the artificial nature of social identity categories, they are powerful lenses through which people experience the self, and thus are important to study with regard to embodiment in self-avatars.

In addition to social identities, individual experiences may also impact embodiment. For example, research participants who more regularly engage with immersive VEs may be more or less embodied within a novel self-avatar. Most research on embodiment, however, does not consider user characteristics when assessing overall embodiment within the sample. Notably, Dewez et al. [17] investigate the influence of gender and locus of control on body ownership and found some evidence of correlation. We argue that it is imperative to assess user characteristics and test, possibly as covariates or moderators in analysis, the extent to which they may predict embodiment in various types of self-avatars and VEs.

2.4 SEM and Virtual Reality

Structural Equation Modeling (SEM) is a commonly used statistical technique in behavioral sciences that uses factor analysis and regression to predict the relationship between theoretical constructs and factors. SEM corrects for measurement error and therefore produces more generalizable estimates of the population parameters. In comparison, traditional multivariate statistical techniques do not correct for measurement error leading to biased parameter estimates, lower power, and reduced generalizability. Due to the advantages of SEM techniques it is recommended to reanalyze data that were previously analyzed using multivariate statistics [49].

Although SEM has clear advantages over traditional multivariate statistics it is generally agreed that relatively large sample sizes of approximately $n = 100$ for a single group model and $n \geq 150$ to compare between groups [50]. The necessary large sample sizes may explain why SEM is not regularly used within the VR community where the average experiment includes $n = 20$ participants [18].

Even though its use is not widespread, SEM has been applied in VR research to evaluate results. Makrandky et al. [51] analyzed learning characteristics with desktop VR and Cheng et al. [52] used SEM to investigate how the virtual experience, environment characteristics, and when the user enters the environment affected participant mood. Recently

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3270725

Venkatakrishnan et al. [53] applied SEM techniques and identified that motion control in the environment was an influential factor that affected both presence and cybersickness.

2.5 Present Research

It is not clear which user characteristics may impact embodiment. One way to investigate the impact of these characteristics on embodiment is through SEM, however VR user studies are often limited by participant size. In this work we revisit data collected from two embodiment experiments [23], [24], [25] with sample sizes large enough to perform SEM. In both experiments participants wore HTC Vive head-mounted displays and had full-body avatars. Data are reanalyzed using SEM to investigate the previously unexplored associations of personal characteristics or social identities including gender, age, participation in STEM, and video game experience on subjective levels of embodiment as measured by the Peck and Gonzalez-Franco embodiment questionnaire [13]. The effect of level of embodiment on participant behavior is further explored using SEM to determine if increased embodiment levels correlate with environment influence. Finally, subjective embodiment measures are limited but current objective measures are either challenging to collect or environment specific. We investigate the efficacy, through SEM, of using head-tracking data as an objective measure of embodiment.

3 Structural Equation Modeling Procedure

The following SEM analysis procedure was performed for both Experiments 1 and 2 following standard SEM methodology [49]. First, a confirmatory factor analysis (CFA) was built using latent factors constructed from observed variables based on theoretical insight and/or statistical procedures such as exploratory factor analysis. The CFA included latent factors—theoretical constructs that cannot be directly measured such as embodiment. The latent factors are constructed from observed variables—measured variables in the experiment such as each individual question in an embodiment questionnaire. The SEM analysis computes the covariance between each observed measure (i.e., individual question), using fewer latent factors (i.e., submeasures of a questionnaire) where the latent factors are assumed to cause the variance and covariance between the observed variables.

Control variables, such as gender, were added to the model at this early stage since they were predicted confounds. The CFA was determined acceptable if all goodness-of-fit measures were in the acceptable or better range. Goodness-of-fit measures included a χ^2-test, the Tucker-Lewis Index (TLI), a Comparative Fit Index (CFI), and a Root Mean Square Error of Approximation (RMSEA). A χ^2-test that is significant suggests that the data does not fit the model. However, since the χ^2-test is affected by population, additional goodness-of fit measures are recommended. Goodness-of-fit cutoffs are defined according to Little [49]. A TLI and a CFI greater than .9 is considered acceptable and greater than .95 is considered a good fit. A TLI of 1 indicates a perfect fit. The RMSEA is an additional approximation of fit where a value less than .08 is acceptable and less than .05 is considered good.

Fig. 1. A first-person perspective of the VE from Experiment 1. A White male avatar is sitting at a desk with a computer monitor and mirror.

After an acceptable CFA was constructed, directional linear regressions were incrementally added into the model to maintain acceptable fit where the estimated coefficient β is the regression slope. The added regressions were based on previous hypotheses. Nonsignificant regressions were trimmed if they did not negatively affect model-fit to produce a final parsimonious model.

The model for Experiment 1 investigated if there was a psychometric equivalence across groups. For this reason, prior to adding regressions, invariance of the CFA model was incrementally tested to determine if the model had weak or strong invariance following the guidelines proposed by Cheung and Rensvold [54]. As recommended by Gonzalez and Griffin [55] the χ^2 difference test was used to determine differences between models at an adjusted p value of .01 due to repeated tests. Significantly different models are considered unacceptable and were rejected. A weak invariant model can test the equivalence across groups between variance and covariance while a model with strong invariance can also test equivalence across means.

Specific details for each SEM procedure can be found in Sections 4.4 and 5.4.

4 Experiment 1

Experiment 1 investigated the effects of same or cross-gendered, race-matched avatars and stereotype-threatening situations on working memory impairment. The full detailed experiment procedure is defined in [23], [24]. The experiment was a 2 (avatar appearance: male, female) × 2 (threat: threat, no threat) between-participant experiment. Participants came to the lab, completed a participant check list to confirm eligibility and signed an informed-consent form. Participants were then seated at a table and donned the head-mounted display (HMD).

Upon entering the VE (Figure 1) participants found themselves in a room and seated at a desk that was co-located with a real-world table. The environment contained two virtual mirrors so that participants could see their self-avatar reflection. Self-avatars were race-matched and either male or female appearing based on the assigned condition. The upper-body, arms, and head of the avatar moved synchronously with the user's movements.

Within the VE, participants completed an approximately three-minute embodiment phase where they looked around the room, into the mirrors, and down at their body. Participants then completed two blocks. For each block, participants in the threat condition heard the instructions, "The test you are about to take has been shown to have gender

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3270725

differences" while participants in the non-threat condition heard the instructions, "The test you are about to take has not been shown to have gender differences". After hearing the instructions participants completed a spatial N-back test interleaved with a saccade test. The experiment lasted for approximately 20 minutes.

After exiting the VE participants completed a demographics questionnaire including a brief four-question embodiment questionnaire, identified if they majored or worked in a STEM field, and on a 7-point Likert scale reported how much they enjoyed mathematics and how good they thought they were at mathematics. Note that this experiment was run before the development of the Peck and Gonzalez-Franco embodiment questionnaires and therefor only used four standard questions to measure embodiment.

Participants were thanked for their time and compensated with a $10 US gift card.

4.1 Participants

189 participants (67.7% women and 32.3% men), (age 22 ± 8 years) were recruited to participate in the IRB approved study. Participants were distributed evenly into the four conditions based on self-identified gender: female avatar $\times$ no threat (F: 31, M:16), female avatar $\times$ threat (F: 31, M:15), male avatar $\times$ no threat (F: 32, M:16), and male avatar $\times$ threat (F: 32, M:16). 71% of participants identified as White, 7% as Black or African American, 7% as East Asian, and 3% as Hispanic or Latinx. The remaining 11% of participants identified as mixed of other races.

4.2 Measures

Latent constructs and control measures added to the SEM are italicized throughout the paper to aid readability.

Measurements included two latent constructs, *embodiment* and *math confidence*, both of which were measured after exiting the VE. Between the experiments run in [23] and [24] the stimulation time was modified. This directly impacted both accuracy and latency making these measures not comparable between the groups of participants.

The observed measures used to construct *embodiment* were four standard embodiment questions [6] adapted from [11] answered using a 7-point Likert scale ranging from strongly disagree to strongly agree. The questions included: [Q1]- I felt as if the body I saw in the virtual world might be my body. [Q2]- I felt like I controlled the virtual body as if it were my own body. [Q3]- I felt like the body in the virtual world was not me. [Q4]- I liked being able to control the movements of the virtual body.

Q3 was negatively scored, and responses to Q1 and Q4 were parceled following Little's [49] parceling recommendations of combining the highest and lowest latent estimates.

Math confidence was parceled from each participant's response to the questions, "How much do you enjoy math?" and "Do you consider yourself to be good at math?" also evaluated on a 7-point Likert scale with 7 being the highest. Within the SEM, responses to all Likert questions were normalized to a range of 0-1 [49].

Three control measures were also considered, 1) participant self-identified *gender*, 2) if the participant was in the *threat* or no threat condition, and 3) if the participant worked or was majoring in a *STEM* discipline.

4.3 Hypotheses

The model was built to investigate the effects of the control variables on *embodiment* and *math confidence*. Additionally, the model investigated if *embodiment* in different gendered avatars predicted *math confidence*. We hypothesized that **[H1]** higher levels of *embodiment* would predict higher environment response, i.e. predict *math confidence*. Based on this experimental design, higher *embodiment* in a female avatar would have lower *math confidence* while higher *embodiment* in a male avatar would have higher *math confidence*.

Since previous literature suggests that users have high embodiment regardless of having self-avatars that are visually different [10] we hypothesised that **[H2]** participants in opposite gendered avatars from their self-identified gender would not have different *embodiment* scores compared to participants in same-gendered avatars. In response to the environment, based on stereotype threat literature [56], we predicted that **[H3]** *math confidence* would further be modulated based on *threat* where existence of a *threat* would lower *math confidence* for participants in female avatars and raise *math confidence* for participants in male avatars.

When considering confidence, having high math confidence is directly linked to majoring in a STEM discipline regardless of race/ethnicity, and men are more likely to major in STEM disciplines compared to women [57]. Based on this information we predicted that **[H4a]** there would be a positive correlation between being in *STEM* and identifying as male, and **[H4b]** there would be a positive correlation between being in *STEM* and *math confidence*. Investigating the effects of being in *STEM* on *embodiment* was viewed as exploratory and no hypotheses were made.

Hypotheses **[H1]** and **[H3]** are related to the gender appearance of the avatar while hypotheses **[H2]** and **[H4a]** are related to the gender identity of the participant.

4.4 Analyses

Analysis was performed using R version 4.0.0 with the lavaan package version 0.6-8. A base model with two latent variables, *embodiment* and *math confidence*, were added to the model using the fixed-factor scaling method.

Building the base model included adding control variable regressions of participant *gender* identification, participation in *STEM*, and the *threat* condition. Control variables were added using a full-partial approach to investigate the affect of each control variable on each latent variable. See Supplemental Material Table 1. For both the male and female avatar groups *threat* did not significantly affect *embodiment*, and participant *gender* did not affect *math confidence*. These control paths were therefore trimmed from the model.

Modification indices suggested adding the correlation between Q2 of the embodiment questionnaire and participant *gender*. Since participant *gender* was a predicted control this correlation was added to the model. *Embodiment* with the correlation between Q2 and participant *gender* can be seen in Figure 2. The models were finally grouped based on avatar gender. The final configural model had excellent fit, ($\chi^2(22) = 18.09$, $p = .70$, CFI=1.00, TLI=1.04, RMSEA=.00, CI=$(.00 - .07)$).

Full descriptive statistics of the model fit procedure can be found in Supplemental Material Table 2. The model was

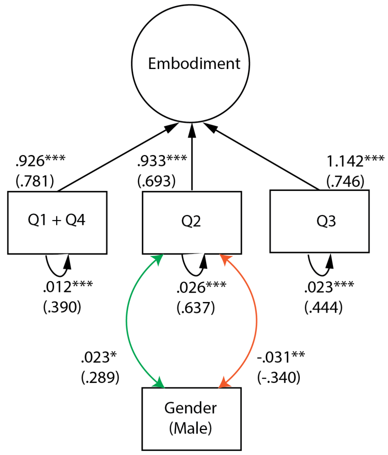

Fig. 2. The embodiment latent construct built from observed questions Q1-Q4. The control variable, *gender* significantly correlated with Q2 based on avatar gender condition (female: green, left, male: orange, right). Data reported is the estimated coefficient (standardized coefficient) with labeled significance level: * $p < .05$, ** $p < .01$, *** $p < .001$.

determined to be strongly invariant by incrementally testing the models from configural to weak to strong invariance. For each new model, the change in CFI was less than .01 and the RMSEA was within the 90% confidence interval of the previous model [54]. Producing a strongly invariant model indicates that observed differences between embodying male or female avatars are "true construct differences" that are not caused from question bias or measurement noise [49].

Differences between avatar groups of variance, covariance, and means were then compared against the strong invariant model. Neither constraining the variance ($\Delta\chi^2 = 1.12$, $p = .57$) or covariance ($\Delta\chi^2 = 0.39$, $p = .53$) significantly changed the model. This suggests that variance and covariance were not significantly different between avatar groups. However, the model worsened when constraining *embodiment* and *math confidence*, ($\Delta\chi^2 = 10.33$, $p = .006$). Further comparison of constraining individual means identified a significant difference between groups on *math confidence* ($\Delta\chi^2 = 10.25$, $p = .001$) but not for *embodiment* ($\Delta\chi^2 = .06$, $p = .81$).

The final model, see Figure 3, constrained variance, covariance, and *embodiment* to be equal between avatar groups. The final model included the *embodiment* to *math confidence* regression and the correlation between *gender* and *STEM* participation. The final model had excellent fit, ($\chi^2(29) = 21.292$, $p = .85$, CFI=1.00, TLI=1.06, RMSEA=.00, CI=$(.00 - .05)$).

4.5 Results and Discussion

The regression from *embodiment* level to *math confidence* investigates whether there is a relationship between embodiment level and response to the VE. Contrary to **[H1]**, *embodiment* did not significantly predict *math confidence*. Even though the regressions were not significant, the direction of both regressions did match the directional hypothesis. A negative regression from *embodiment* to *math confidence* was found for female avatars ($\beta = -.10$, $z = -.93$, $p = .35$). Conversely for male avatars a positive regression from *embodiment* to *math confidence* was found ($\beta = .07$, $z = .55$, $p = .58$).

Embodiment was measured using four commonly used embodiment questions instead of the 16 questions proposed by Peck and Gonzalez-Franco [13]. The condensed questionnaire may be limited in its ability to detect more nuanced levels of embodiment that may have enabled prediction of behavioral response to the environment. This leaves an open question about embodiment questionnaires' prediction of behavioral response and is further investigated in Section 5.

Additionally, the subjectivity of the questionnaire may be directly influenced by participant *gender*. This can be seen with responses to Q2 significantly correlating with participant *gender*. See Figures 2 and 4. Q2 asked participants if they felt like they "controlled the virtual body" as if it were their own body. Identifying as male positively correlated with Q2 in the female self-avatar condition and negatively correlated with Q2 in the male self-avatar condition. In other words, men self-reported that they felt like they controlled the female body more than they controlled the male body. Previous research has found that participants in avatars that differ across a visually salient diversity-dimension from the self have similar levels of embodiment [10], [11], [58] suggesting that participants, regardless of avatar gender would answer similarly. Moreover, if a difference is detected it would be assumed that participants would have higher levels of embodiment in avatars that are more similar to the user, i.e., the same gender as the user. However, the responses for Q2 found the opposite. In this case, men felt greater control over the female avatar compared to the male avatar. The difference in response may relate to the perceived power differences between men and women regarding the feeling of "control" over the "virtual body". That is, women's bodies may be viewed as more controllable than men's bodies. For example, Stets [59] found that gender identity was a significant predictor of perceived control over a partner. Specifically, men who identify as more stereotypically masculine felt more control over women compared to men who identify as less stereotypically masculine. For reference, hyper-masculine men identify with a need for power and dominance, and hyper-masculinity has been linked to excessive aggression and the use of force against others including against women [60].

Further evidence that gender identity affected responses to the subjective embodiment questionnaire can be seen investigating **[H2]**, that participants in avatars with the opposite gender from their self-identified gender would not have different *embodiment* scores compared to participants in same-gendered avatars. In support of **[H2]** and replicating previous work, embodiment was not found to be significantly different between embodying a male or female avatar. This was demonstrated by not finding a significant difference in models when constraining latent means for *embodiment*. However, participant *gender* did significantly predict *embodiment* but only in the female avatar condition. Specifically, identifying as male negatively correlated with *embodiment* ($\beta = -.84$, $z = -2.93$, $p = .003$). Conversely, no significant effect of *gender* on *embodiment* was found in the male avatar condition ($\beta = .22$, $z = .71$, $p = .48$). These results support **[H2]** for female but not male participants.

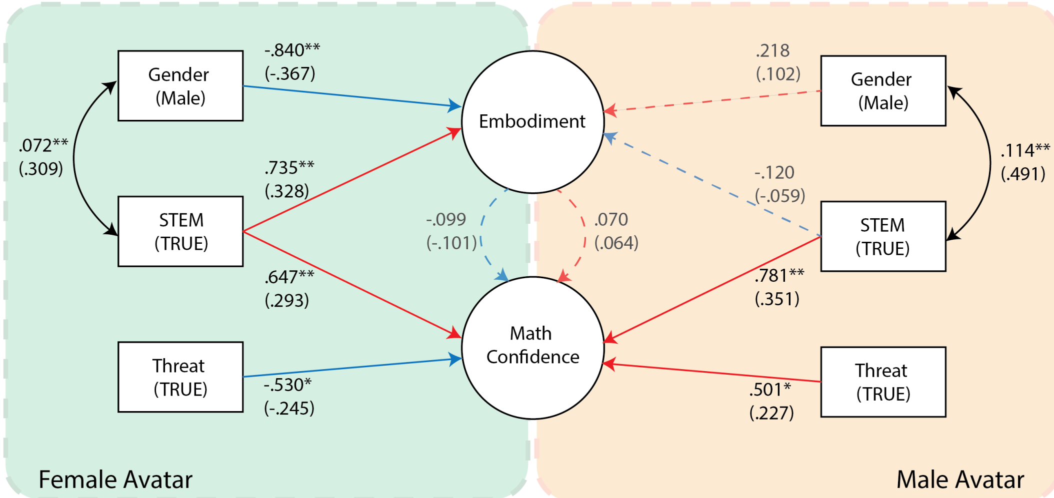

Fig. 3. The final structural equation model for Experiment 1 ($\chi^2(29) = 21.292$, $p = .85$, CFI=1.00, TLI=1.06, RMSEA=.00, CI=$(.00 - .05)$). The model is grouped by avatar condition (female: left, green, male: right, orange). Regressions are one-sided arrows (positive: red, negative: blue). Correlations are double-sided arrows (black). Solid lines are significant; dotted lines are non-significant. Data reported is β, the estimated coefficient (standardized coefficient) with labeled significance level: * $p < .05$, ** $p < .01$, *** $p < .001$.

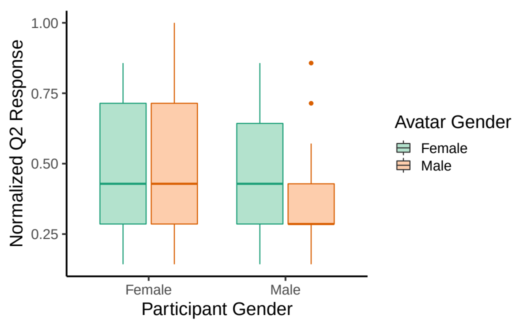

Fig. 4. The embodiment latent construct built from observed questions Q1-Q4. The control variable, gender significantly correlated with Q2 based on avatar gender condition (female: green, left, male: orange, right).

The lack of support for **[H2]** for male participants could be related to the assessment methodology. *Embodiment* was measured using a subjective questionnaire and questionnaires are known to be problematic [22]. Specifically, social desirability responding is known to affect responses to explicit measurement tools [45] such that participants respond with answers based on what they think are socially acceptable answers. Social desirability may have impacted men's responses to the questionnaire providing further support that masculinity may have affected responses to Q2. From childhood boys are encouraged to avoid emotional intimacy and self-disclosure by wearing a "gender straitjacket" that discourages any behavior that may be seen as feminine [61]. Societal pressure for boys to be manly and to not "be a girl" are ingrained in relationship behaviors between young boys that continue into adulthood [62]. The social desirability for men to not be women may have directly impacted how participants responded to the questionnaire by encouraging men to not claim that a female appearing body "might be my body" in the VE while still claiming "control" over the body. Alternatively, these years of training for boys to suppress femininity may cause men to be less embodied in female avatars which would support a basic explanation that men experience lower *embodiment* in a cross gendered avatar compared to women.

However, regardless of *embodiment* level for men, the effect of *threat* and the avatar condition on *math confidence* supports **[H3]**. Participant *gender* did not affect *math confidence* and that regression was trimmed from the model. However, regardless of participant gender, being in the *threat* condition significantly affected *math confidence* in the predicted directions. *Threat* lowered *math confidence* in the female avatar condition ($\beta = -.53$, $z = -2.49$, $p = .01$) and increased *math confidence* in the male avatar condition ($\beta = .50$, $z = 2.39$, $p = .02$). The effect on *math confidence* from *threat* between avatar conditions, regardless of participant *gender*, suggests that the avatar gender affected results and that participants likely experienced some effect of their avatar gender regardless of their self-identified gender or subjective embodiment scores.

These results are supported by previous research that found that the avatar gender, regardless of participant gender, affects behavior. For example, in online forums avatar gender determined whether people sought help such that female avatars were more likely to ask for help [63] and men were more likely to perform healing actions when playing as female characters [64].

Finally, in support of **[H4a]** there was a significant correlation between identifying as male and being in *STEM* ((F)emale Avatar: $\beta = .07$, $z = 2.70$, $p = .01$, (M)ale Avatar: $\beta = .11$, $z = 4.17$, $p < .001$). Additionally, supporting **[H4b]**, *math confidence* was higher for participants in *STEM* disciplines (F: $\beta = .65$, $z = 2.81$, $p = .005$, M: $\beta = .78$, $z = 3.40$, $p = .001$).

Interestingly, being in *STEM* significantly increased *embodiment* within the female avatar condition ($\beta = .74$, $z = 2.96$, $p = .003$) but not within the male avatar condition ($\beta = -.12$, $z = -.44$, $p = .66$). Masculinity may again help explain this unexpected result. Men are more likely to be in

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3270725

STEM majors as seen by the significant correlation between *gender* and *STEM*. According to Simon et al. [65] men who identify as less stereotypically masculine are more likely to major in STEM fields compared to men who identify as more masculine. Because men in STEM are more likely to identify as less stereotypically masculine (ex. aggressive, dominant, controlling) compared to non-STEM men, they may be more likely to report higher subjective *embodiment* in female avatars. This suggests that the extent to which people identify with their gender-typed characteristics may impact responses on embodiment questionnaires, especially for gender-swapping situations. For future embodiment studies, collecting and analyzing data about the extent to which people identify with their gender identity may further the understanding of the effect of gender identity on embodiment in same and differently gendered avatars.

5 Experiment 2

Experiment 2 investigated the existence of shooter-bias in VR. The detailed experiment procedure is defined in [25]. Participants came to the lab where they confirmed eligibility and completed an informed consent form. Participants received verbal instructions about the experimental procedure and were asked to repeat the instructions to the experimenter to confirm understanding before entering the VE.

Participants were instructed that they would be carrying a virtual gun and should shoot at avatars carrying guns and not shoot at avatars carrying cell phones. A shooting decision was made by pulling the trigger button on a hand-held tracker. Each trial would end when the participant holstered their virtual gun by placing it at their hip.

Fig. 5. A first-person perspective of the VE from Experiment 2. A White male self-avatar is pointing a gun at an armed avatar. The self-avatar is reflected in the mirror on the left of the image.

The VE (Figure 5) was a darkened alley with a mirror and reflective window. Participants were given a self-identified gender and race matched avatar that moved with full-body synchrony. Participants then completed 160 trials where they either shot a virtual gun at armed avatars holding guns or inhibited shooting at unarmed avatars holding cell phones. The avatars were either Black or White males and wore clothing chosen to indicate high or low socioeconomic status. At the end of the experiment participants completed a race implicit association test and a demographics survey including gender, age, and video game playing experience. They were debriefed, thanked for their time, and compensated $10 US.

5.1 Participants

99 participants (55.5% women, 42.5% men, and 2.0% non-binary), (age 23±10 years ranging from 18 to 76) were recruited to participate in this IRB-approved study. 81% of participants identified as White, 9% as Asian, 8% as Hispanic or Latinx, 3% as Black or African American, and 1% as multi-racial. 56% of participants identified as non-gamers, 24% as casual gamers, 17% as core gamers, and 3% as hard-core gamers.

5.2 Measures

The measures included six latent constructs. Four latent constructs were the submeasures of embodiment as defined by Peck and Gonzalez-Franco [13]: *appearance*, *response*, *ownership*, and *multi-sensory*. More information about the final models for the latent constructs can be found in the supplemental material. The observed measures used to build each embodiment submeasure construct were the individual questions from the embodiment questionnaire [13]. An embodiment latent construct was not built from the four submeasures and was not added to the CFA due to an inadequate sample size for a multi-level model [49]. The current model enabled interpretation at the individual embodiment submeasure level.

The additional two latent variables were built from objective measures collected during the virtual experience. The first latent construct was *environment behavior* defined as the average latency and accuracy of shooting at armed avatars, regardless of race or clothing. Accuracy was defined as making the correct decision to shoot at armed avatars where 100% accuracy was making the correct shooting decision for all armed avatars. The accuracy of making the decision to not shoot at unarmed avatars was not included in analysis since participants had near perfect accuracy in these conditions. To be directly comparable to accuracy the latency measure was defined as *normalized_latency*, where normalized_latency was each participant's averaged latency normalized across all participants. A high *environment behavior* measure indicated high accuracy and low latency. This metric was designed to capture how people responded to the environment task to accurately and quickly shoot at armed avatars, regardless of avatar appearance. Averaging over race and clothing was performed to minimize personal bias, such as racial bias, from the latent measure. Participants experiencing a higher sense of embodiment would be more likely to respond to the VE and therefore have higher environment behavior.

The final latent measure, *movement complexity* was derived from the p95 head tracker scores as defined in [25]. Participants were instructed to stand in one location and to move their right-arm to raise and holster a virtual gun. Position and orientation head-tracker data was collected. The p95 score is defined as "the sum of the variance accounted for by the first principal components that on average accounted for at least 95% of the variance" in the (x, y, z) position and $(yaw,\ pitch,\ roll)$ orientation head-tracker data. The p95 scores were also averaged over race and clothing to minimize personal bias. A high *movement complexity* score indicates more complicated head-tracker movements.

Finally, three control measures were added to the model, 1) participant *gender*, 2) *age*, and 3) *gaming* experience.

5.3 Hypotheses

The model was built to investigate the relationship between both conscious and unconscious environment behavior and embodiment. The model was further built to investigate the affects of control measures on embodiment and environment behavior. The following directional hypotheses were made.

Previous research supports that having a virtual self-avatar versus not having an avatar improves VE performance [4]. Additional evidence suggests that higher levels of embodiment should more strongly affect environment behavior [14]. We again hypothesize that **[H1]** higher levels of embodiment will positively predict virtual environment behavior. That is, people with higher embodiment will be more accurate and make faster shooting decisions at armed avatars.

Since head movement has been shown to indicate environment acceptance [40], [66] we hypothesized that **[H6]** *movement complexity* calculated from head-tracker data would predict embodiment such that greater *movement complexity*, i.e. more complicated head movements, would predict higher embodiment.

The experiment was a within-participant design and the self-avatar was gender and race matched for all participants. Since the experiment scenario did not manipulate the avatar, effects of the control variables relate directly to subjective embodiment levels. We hypothesised that **[H7]** gender would not affect embodiment since previous research has not found differences in embodiment levels due to swaps in differently-appearing bodies [11], [12]. Based on reductions in fluid intelligence–a term used to describe reasoning and comprehension— in older adults, increased age should lower embodiment [67]. No prediction about the effect of gaming on embodiment was made.

5.4 Analyses

A base model with six latent variables was specified using the fixed-factor scaling method. The latent variables included *movement complexity*, *environment behavior*, and the four submeasures of embodiment as defined by Peck and Gonzalez-Franco [13]: *appearance*, *response*, *ownership*, and *multi-sensory*.

Control variables included participant *gender* identification, *age*, and video *game* playing experience. Control variables were added using a full-partial approach to investigate the effect of each control variable on each latent variable. Non-significant control variables were incrementally removed from the model while maintaining acceptable fit. The included control variables in the CFA can be seen in Supplemental Material Table 3 while the remaining control variables were removed. The effect of gender on both *movement complexity* and *environment behavior* was not significant, however when removed reduced model fit to unacceptable levels. This suggests that ~~even though minimal,~~ gender affected these latent constructs and they should remain in the model. The final configural model had acceptable to good fit, ($\chi^2(204) = 240.80$, $p = .04$, CFI=.95, TLI=.93, RMSEA=.04, CI=($.01 - .06$)).

The planned regressions were added after establishing the configural model. The planned regressions were from *movement complexity* to each embodiment submeasure and then from each embodiment submeasure to *environment behavior*. Non-significant regressions were incrementally trimmed from the model while maintaining acceptable fit to build a final parsimonious model. The removed regressions were from *multi-sensor*, *ownership*, and *appearance* to *environment behavior*.

The final model, see Figure 6, had good fit ($\chi^2(203) = 230.99$, $p = .09$, CFI=.96, TLI=.95, RMSEA=.04, CI=($.00 - .06$)).

5.5 Results and Discussion

The final model provides evidence in partial support of **[H1]** that higher levels of embodiment in a self-avatar influence *environment behavior*. The *response* submeasure significantly predicted *environment behavior* ($\beta = .34$, $z = 2.30$, $p = .02$). Participants with higher subjective measures of *response* shot more quickly and accurately at armed avatars. The *response* submeasure is the observed correlation between the question answers. *Response* questions inquired about the body changing into the avatar body such as, "It felt as if my real body were turning into an avatar body" and "I felt as if my body had changed", as well as questions involving response to the virtual gun such as, "I felt that my own body could be affected by the gun" and "It seemed as if my body was touching the gun". This supports previous research demonstrating that having a self-avatar may improve cognitive function within the VE [4] and expands on this work by demonstrating that being more embodied in that self-avatar further improves performance as measured by speed and accuracy. Future research should investigate why *response* was able to predict *environment behavior* while other submeasures were not.

It is possible that the remaining submeasures do not influence participant behavior. Alternatively, the lack of significance in the three other embodiment submeasures may be limited by the *environment behavior* metric only looking at speed and accuracy. Future research should investigate if alternate measures, such as behavior change or working memory, may be predicted by other submeasures. It is currently an open research question as to what behavior change, if any, each submeasure may predict. Each submeasure may relate to different types of participant behavior in ways not yet known.

The final model supported **[H6]**, that *movement complexity* would predict embodiment. More complicated head-tracker data significantly predicted three out of four embodiment submeasures: *multi-sensory* ($\beta = .30$, $z = 2.01$, $p = .04$), *ownership* ($\beta = .33$, $z = 2.25$, $p = .02$), and *response* ($\beta = .25$, $z = 1.99$, $p < .05$). No significant result was found for the fourth submeasure: *appearance* ($\beta = .25$, $z = 1.88$, $p = .06$). More complicated head-tracker movement predicting higher embodiment is supported by the theory of embodied cognition— how you move your body affects your cognitive understanding of the world. Further, body movement directly affects your emotional state as well as your decision making capabilities [39]. Increased and natural movements have also been shown to increase engagement in video games [43], [44]. It follows that increased and natural movements would also increase

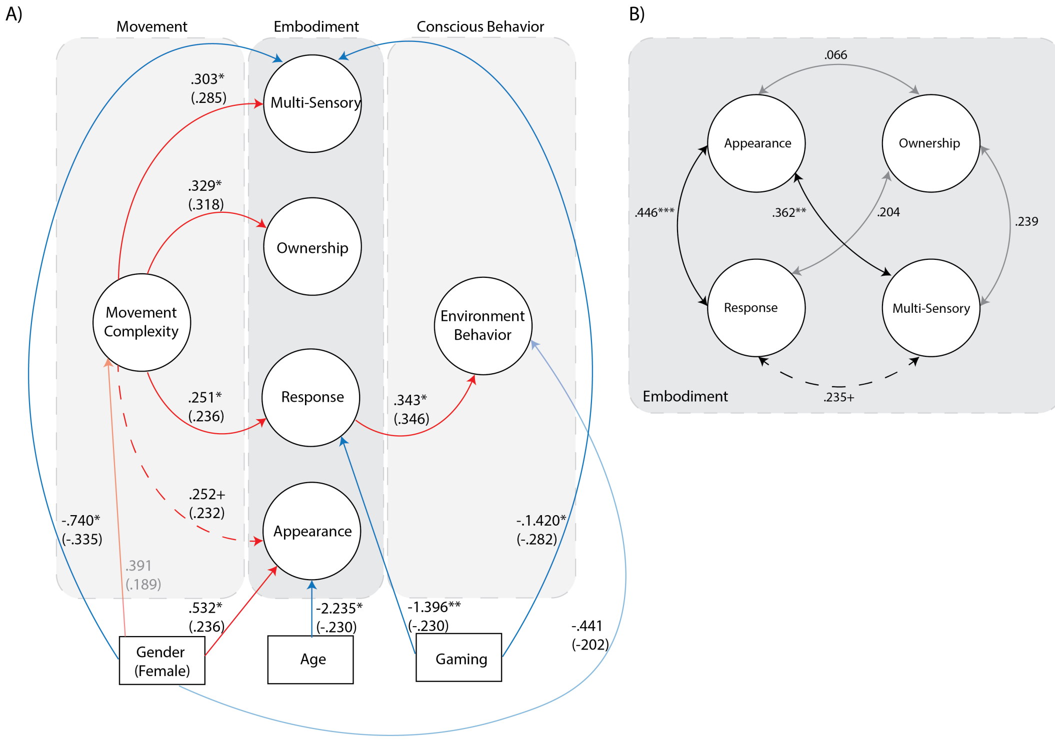

Fig. 6. The final structural equation model for Experiment 2 ($\chi^2(203) = 230.99$, $p = .09$, CFI=.96, TLI=.95, RMSEA=.04, CI=$(.00 - .06)$). Significant ($p < .05$) regressions are one-sided solid arrows (positive: red, negative: blue), ($p < .1$) are dotted lines, and ($p \geq .1$) regressions are faded lines. Correlations are double-sided arrows. Data reported is the estimated coefficient (standardized coefficient) with labeled significance level: + $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$. A) The main structure. For clarity, the correlations between the embodiment submeasures are represented in B).

engagement in VEs as demonstrated by *movement complexity* predicting higher subjective embodiment levels.

These higher levels of *response* then predicted *environment behavior*. This was previously demonstrated by Steed et al. [4] where prohibiting participant movement with a self-avatar limited cognitive performance while enabling movement with a self-avatar improved cognitive performance. That experiment did not measure embodiment, however it is possible that the increased movement increased embodiment which then improved cognitive performance.

Further, the existence of a self-avatar when movement is limited does not appear to increase cognitive performance [4], [68]. This further supports that movement and agency, as supported by the final model (see Figure 6), continue to be fundamental for embodiment [69]. Gonzalez-Franco and Peck's [6] initial embodiment questionnaire put agency into its own submeasure. Interestingly, Peck and Gonzalez-Franco's [13] updated embodiment questionnaire suggests that agency affects multiple submeasures and is not measurable as a single submeasure. Additionally, agency may be challenging to measure subjectively since the researchers found that two of the previously proposed agency questions were unreliable. They dropped the agency submeasure and added the remaining agency related questions into three of the four submeasures. The researchers further argue that embodiment experiments should "enable some amount of agency." In this work, *movement complexity* predicted embodiment submeasures. This suggests that agency likely influences embodiment. Even though agency may be challenging to measure with a subjective questionnaire, promisingly, movement and agency may be measurable through *movement complexity* as demonstrated here.

When considering individual characteristics, *gender*, *age*, and *gaming* experience each affected subjective embodiment differently. Identifying as female positively correlated with *appearance* ($\beta = .53$, $z = 2.10$, $p = .04$) and negatively correlated with *multi-sensory* ($\beta = -.74$, $z = -2.52$, $p = .01$). It is not surprising that gender differences were identified when measuring embodiment since numerous differences have already been identified between genders when considering avatars and embodiment. For example, men and women may accept different avatar hands such that women are less likely to accept male avatar hands and men are less likely to accept non-human hands [70]. Conversely, in this work, identifying as a woman predicted higher *appearance* submeasure scores when the self-avatar was gender and race matched with the participant. However, women's lack of acceptance of a male-appearing hand suggests that the specific appearance characteristics, or the cohesiveness of the avatar may be critical for women to become embodied. In the present study, women also had lower *multi-sensory* scores than men. *Multi-sensory* questions focused on feeling the "touch of the gun". Sex differences have been found for identifying objects haptically [71] and women may have been more sensitive to the inaccurate haptic feedback of holding an HTC Vive tracker versus an object shaped like a real gun. Our results suggest that when designing embodiment scenarios it may be important to create realistic haptic feedback to increase embodiment for women.

Similar to *gender*, *age* was also related to *appearance* ($\beta = -2.24$, $z = -2.18$, $p = .03$). We had predicted that age may impact embodiment levels. Cognitive function, including fluid intelligence declines in older adults [67] suggesting that embodiment illusions may not be as effective and that additional measures may need to be considered when working with this population. Previous work by Carrasco et al. [72] investigated avatar configuration in older adults and found that older adults may have different representational requirements including representations of the aging body. Additionally, Puri et al. [73] further found that older adults preferred photorealistic avatars to avatars that were less visually representative of the user. Our model further supports that the appearance of the avatar is important for older adults; *age* was negatively related to *appearance* leading to an overall lower embodiment score. Surprisingly, *age* did not predict any other submeasures of embodiment. This suggests that older adults may still experience embodiment, however the appearance of an age-matched avatar may be important. The relationship between age and lower levels of *appearance* may have been caused by the non-inclusive design of Experiment 2 where all self-avatars were younger appearing adults. The avatars used in the experiment were gender and race matched, but not age matched to the participant. This suggests that when working with older adults, age matching avatars may increase embodiment. Further work performing more comprehensive studies with older populations is needed to better understand the relations between age and embodiment.

Finally, *gaming* negatively affected both *response* ($\beta = -1.40$, $z = -2.73$, $p = .006$) and *multi-sensory* ($\beta = -1.42$, $z = -2.16$, $p = .03$). It may be that those with extensive gaming experience are accustomed to very high-resolution graphics, fine-grained movements, and sophisticated haptics. As such, the self-avatars in the present study may have felt somewhat less realistic, resulting in lower embodiment among more frequent gamers. Thus, gaming status should be measured and accounted for when sampling from various populations. In sum, the differences identified based on *gender*, *age*, and *gaming* highlight the necessity of investigating diverse populations when attempting to make general claims [18].

6 Conclusion and Future Work

Data from two embodiment experiments were reanalyzed using SEM. The reanalysis of the results identified new insights into virtual embodiment illusions and demonstrates the effectiveness of SEM to investigate the effects of individual characteristics on virtual embodiment. Regarding the first aim of our paper to test how embodiment influences behavior, results provide some small support that higher levels of embodiment in a self-avatar exhibit greater environment influence. This was demonstrated by participants being more accurate and having faster response times in Experiment 2. A similar pattern showing that greater embodiment has greater environment influence is shown by the stereotypical views of the self-avatar influencing the direction of the regression coefficients from embodiment to the post-experiment math confidence measure in Experiment 1. This should be replicated with the full 16-question embodiment questionnaire to see if the pattern replicates. In situations exhibiting stereotype threat greater embodiment in a female avatar may lower math confidence and greater embodiment in a male avatar may raise math confidence.

Regarding the second aim of out paper, we found that individual characteristics including gender, age, gaming experience, and majoring or working in STEM predicted embodiment in different ways. These results suggest that when designing embodiment illusions, user identity should be taken into account. For example, avatars should be age-matched to the participant as demonstrated by older participants having lower scores on the appearance submeasure when given non-age matched avatars. Further, gamers may have subjectively lower embodiment scores suggesting that follow-up research should investigate ways to increase embodiment levels for this group.

Additional results suggest that the effect of gender identity on embodiment is complicated. Results from an experiment with gender swapping suggest that men have lower embodiment scores in female avatars compared to male avatars. Conversely, women do not have lower embodiment scores in male avatars compared to female avatars. Further, men's response to the question "I felt like I controlled the virtual body as if it were my own body" was unexpected and was higher when men were in a female avatar compared to being in a male avatar. Future work should investigate questionnaire wording as well as the effects of gender identity on embodiment in avatars with visual identities different from the current self-identity. These results provide further evidence that individuals' unique personal and social identities may impact embodiment illusions in unknown and unexpected ways and that diverse participant populations should be studied to produce generalizable results [18].

Finally, regarding the third aim of out paper, we demonstrate that increased movement complexity, as measured through head-tracking data, predicted increased embodiment. We provide objective evidence that increasing participant movement may increase subjective embodiment. The more complicated the movements participants made in the experiment the more embodied they became in the avatar. Movement complexity provides a promising new objective measure for immersive head-tracked embodiment experiments. The measure is easy to collect, does not require additional hardware to be worn by the participant, and is environment independent. This new measure may add further insight into understanding and quantifying virtual embodiment illusions.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant 1942146. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] E. A. McManus, B. Bodenheimer, S. Streuber, S. De La Rosa, H. H. Bülthoff, and B. J. Mohler, "The influence of avatar (self and character) animations on distance estimation, object interaction and locomotion in immersive virtual environments," in *Proceedings of the ACM SIGGRAPH Symposium on applied perception in graphics and visualization*, 2011, pp. 37–44.

[2] L. Phillips, B. Ries, M. Kaeding, and V. Interrante, "Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments," in *2010 IEEE virtual reality conference (VR)*. IEEE, 2010, pp. 115–1148.

[3] Y. Pan and A. Steed, "Avatar type affects performance of cognitive tasks in virtual reality," in *25th ACM Symposium on Virtual Reality Software and Technology*, 2019, pp. 1–4.

[4] A. Steed, Y. Pan, F. Zisch, and W. Steptoe, "The impact of a self-avatar on cognitive load in immersive virtual reality," in *2016 IEEE virtual reality (VR)*. IEEE, 2016, pp. 67–76.

[5] M. Slater and M. Usoh, "The influence of a virtual body on presence in immersive virtual environments," in *VR 93, Virtual Reality International, Proceedings of the Third Annual Conference on Virtual Reality*, 1993, pp. 34–42.

[6] M. Gonzalez-Franco and T. C. Peck, "Avatar embodiment. towards a standardized questionnaire," *Frontiers in Robotics and AI*, vol. 5, p. 74, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2018.00074

[7] K. Kilteni, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence: Teleoperators and Virtual Environments*, vol. 21, no. 4, pp. 373–387, 2012.

[8] N. Yee and J. Bailenson, "The proteus effect: The effect of transformed self-representation on behavior," *Human communication research*, vol. 33, no. 3, pp. 271–290, 2007.

[9] D. Banakou, S. Kishore, and M. Slater, "Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias," *Frontiers in psychology*, vol. 9, p. 917, 2018.

[10] D. Banakou, A. Beacco, S. Neyret, M. Blasco-Oliver, S. Seinfeld, and M. Slater, "Virtual body ownership and its consequences for implicit racial bias are dependent on social context," *Royal Society open science*, vol. 7, no. 12, p. 201848, 2020.

[11] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater, "Putting yourself in the skin of a black avatar reduces implicit racial bias," *Consciousness and cognition*, vol. 22, no. 3, pp. 779–787, 2013.

[12] D. Banakou, R. Groten, and M. Slater, "Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes," *Proceedings of the National Academy of Sciences*, vol. 110, no. 31, pp. 12 846–12 851, 2013.

[13] T. C. Peck and M. Gonzalez-Franco, "Avatar embodiment. a standardized questionnaire," *Frontiers in Virtual Reality*, vol. 1, p. 44, 2021.

[14] J. M. Juliano, R. P. Spicer, A. Vourvopoulos, S. Lefebvre, K. Jann, T. Ard, E. Santarnecchi, D. M. Krum, and S.-L. Liew, "Embodiment is related to better performance on a brain–computer interface in immersive virtual reality: A pilot study," *Sensors*, vol. 20, no. 4, p. 1204, 2020.

[15] M. Gonzalez-Franco, P. Abtahi, and A. Steed, "Individual differences in embodied distance estimation in virtual reality," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 941–943.

[16] S. J. Ahn, J. Bostick, E. Ogle, K. L. Nowak, K. T. McGillicuddy, and J. N. Bailenson, "Experiencing nature: Embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature," *Journal of Computer-Mediated Communication*, vol. 21, no. 6, pp. 399–419, 2016.

[17] D. Dewez, R. Fribourg, F. Argelaguet, L. Hoyet, D. Mestre, M. Slater, and A. Lécuyer, "Influence of personality traits and body awareness on the sense of embodiment in virtual reality," in *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2019, pp. 123–134.

[18] T. C. Peck, L. E. Sockol, and S. M. Hancock, "Mind the gap: The underrepresentation of female participants and authors in virtual reality research," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 5, pp. 1945–1954, 2020.

[19] M. González-Franco, T. C. Peck, A. Rodríguez-Fornells, and M. Slater, "A threat to a virtual hand elicits motor cortex activation," *Experimental brain research*, vol. 232, no. 3, pp. 875–887, 2014.

[20] G. Tieri, E. Tidoni, E. F. Pavone, and S. M. Aglioti, "Body visual discontinuity affects feeling of ownership and skin conductance responses," *Scientific reports*, vol. 5, no. 1, pp. 1–8, 2015.

[21] C. Hamilton-Giachritsis, D. Banakou, M. G. Quiroga, C. Giachritsis, and M. Slater, "Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[22] M. Slater, "How colorful was your day? why questionnaires cannot assess presence in virtual environments," *Presence*, vol. 13, no. 4, pp. 484–493, 2004.

[23] T. C. Peck, M. Doan, K. A. Bourne, and J. J. Good, "The effect of gender body-swap illusions on working memory and stereotype threat," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1604–1612, 2018.

[24] T. C. Peck, J. J. Good, and K. A. Bourne, "Inducing and mitigating stereotype threat through gendered virtual body-swap illusions," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[25] T. C. Peck, J. J. Good, and K. Seitz, "Evidence of racial bias using immersive virtual reality: Analysis of head and hand motions during shooting decisions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2502–2512, 2021.

[26] M. Botvinick and J. Cohen, "Rubber hands 'feel'touch that eyes see," *Nature*, vol. 391, no. 6669, pp. 756–756, 1998.

[27] A. Kalckert and H. H. Ehrsson, "The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership," *Consciousness and cognition*, vol. 26, pp. 117–132, 2014.

[28] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. V. Sanchez-Vives, "Inducing illusory ownership of a virtual body," *Frontiers in neuroscience*, vol. 3, p. 29, 2009.

[29] R. Fribourg, F. Argelaguet, A. Lécuyer, and L. Hoyet, "Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 5, pp. 2062–2072, 2020.

[30] Y. Pan and A. Steed, "The impact of self-avatars on trust and collaboration in shared virtual environments," *PloS one*, vol. 12, no. 12, p. e0189078, 2017.

[31] M. Slater, S. Neyret, T. Johnston, G. Iruretagoyena, M. Á. de la Campa Crespo, M. Alabèrnia-Segura, B. Spanlang, and G. Feixas, "An experimental study of a virtual reality counselling paradigm using embodied self-dialogue," *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.

[32] B. J. Mohler, S. H. Creem-Regehr, W. B. Thompson, and H. H. Bülthoff, "The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment," *Presence*, vol. 19, no. 3, pp. 230–242, 2010.

[33] B. Ries, V. Interrante, M. Kaeding, and L. Phillips, "Analyzing the effect of a virtual avatar's geometric and motion fidelity on ego-centric spatial perception in immersive virtual environments," in *Proceedings of the 16th ACM symposium on virtual reality software and technology*, 2009, pp. 59–66.

[34] N. Ogawa, T. Narumi, H. Kuzuoka, and M. Hirose, "Do you feel like passing through walls?: Effect of self-avatar appearance on facilitating realistic behavior in virtual environments," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–14.

[35] F. Garbarini, L. Fornia, C. Fossataro, L. Pia, P. Gindri, and A. Berti, "Embodiment of others' hands elicits arousal responses similar to one's own hands," *Current Biology*, vol. 24, no. 16, pp. R738–R739, 2014.

[36] S. Y. S. Kim, N. Prestopnik, and F. A. Biocca, "Body in the interactive game: How interface embodiment affects physical activity and health behavior change," *Computers in Human Behavior*, vol. 36, pp. 376–384, 2014.

[37] K. Kilteni, J.-M. Normand, M. V. Sanchez-Vives, and M. Slater, "Extending body space in immersive virtual reality: a very long arm illusion," *PloS one*, vol. 7, no. 7, p. e40867, 2012.

[38] K. Kilteni, I. Bergstrom, and M. Slater, "Drumming in immersive virtual reality: the body shapes the way we play," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 4, pp. 597–605, 2013.

[39] M. Reimann, W. Feye, A. J. Malter, J. M. Ackerman, R. Castano, N. Garg, R. Kreuzbauer, A. A. Labroo, A. Y. Lee, M. Morrin *et al.*, "Embodiment in judgment and choice." *Journal of Neuroscience, Psychology, and Economics*, vol. 5, no. 2, p. 104, 2012.

[40] G. L. Wells and R. E. Petty, "The effects of over head movements on persuasion: Compatibility and incompatibility of responses," *Basic and applied social psychology*, vol. 1, no. 3, pp. 219–230, 1980.

[41] G. Tom, P. Pettersen, T. Lau, T. Burton, and J. Cook, "The role of overt head movement in the formation of affect," *Basic and applied social psychology*, vol. 12, no. 3, pp. 281–289, 1991.

[42] F. Herrera and J. N. Bailenson, "Virtual reality perspective-taking at scale: Effect of avatar representation, choice, and head movement on prosocial behaviors," *New Media & Society*,

vol. 23, no. 8, pp. 2189–2209, 2021. [Online]. Available: https://doi.org/10.1177/1461444821993121

[43] R. Weiss, A. Duda, and D. K. Gifford, "Composition and search with a video algebra," *IEEE multimedia*, vol. 2, no. 1, pp. 12–25, 1995.

[44] N. Bianchi-Berthouze, W. W. Kim, and D. Patel, "Does body movement engage you more in digital game play? and why?" in *International conference on affective computing and intelligent interaction*. Springer, 2007, pp. 102–113.

[45] R. H. Fazio and M. A. Olson, "Implicit measures in social cognition research: Their meaning and use," *Annual review of psychology*, vol. 54, no. 1, pp. 297–327, 2003.

[46] J. T. Kubota and T. Ito, "Rapid race perception despite individuation and accuracy goals," *Social neuroscience*, vol. 12, no. 4, pp. 468–478, 2017.

[47] P. C. Quinn, J. Yahr, A. Kuhn, A. M. Slater, and O. Pascalis, "Representation of the gender of human faces by infants: A preference for female," *Perception*, vol. 31, no. 9, pp. 1109–1121, 2002.

[48] A. Fausto-Sterling, *Sexing the body: Gender politics and the construction of sexuality*. Basic Books, 2000.

[49] T. D. Little, *Longitudinal structural equation modeling*. Guilford press, 2013.

[50] J. Wang and X. Wang, *Structural equation modeling: Applications using Mplus*. John Wiley & Sons, 2019.

[51] G. Makransky and G. B. Petersen, "Investigating the process of learning with desktop virtual reality: A structural equation modeling approach," *Computers & Education*, vol. 134, pp. 15–30, 2019.

[52] L.-K. Cheng, M.-H. Chieng, and W.-H. Chieng, "Measuring virtual experience in a three-dimensional virtual reality interactive simulator environment: a structural equation modeling approach," *Virtual Reality*, vol. 18, no. 3, pp. 173–188, 2014.

[53] R. Venkatakrishnan, R. Venkatakrishnan, R. G. Anaraky, M. Volonte, B. Knijnenburg, and S. V. Babu, "A structural equation modeling approach to understand the relationship between control, cybersickness and presence in virtual reality," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 682–691.

[54] G. W. Cheung and R. B. Rensvold, "Evaluating goodness-of-fit indexes for testing measurement invariance," *Structural equation modeling*, vol. 9, no. 2, pp. 233–255, 2002.

[55] R. Gonzalez and D. Griffin, "Testing parameters in structural equation modeling: every" one" matters." *Psychological Methods*, vol. 6, no. 3, p. 258, 2001.

[56] T. Schmader, "Gender identification moderates stereotype threat effects on women's math performance," *Journal of experimental social psychology*, vol. 38, no. 2, pp. 194–201, 2002.

[57] M. W. Moakler Jr and M. M. Kim, "College major choice in stem: Revisiting confidence and demographic factors," *The Career Development Quarterly*, vol. 62, no. 2, pp. 128–142, 2014.

[58] D. Banakou, P. D. Hanumanthu, and M. Slater, "Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias," *Frontiers in human neuroscience*, vol. 10, p. 601, 2016.

[59] J. E. Stets, "Role identities and person identities: Gender identity, mastery identity, and controlling one's partner," *Sociological perspectives*, vol. 38, no. 2, pp. 129–150, 1995.

[60] A. Cohn and A. Zeichner, "Effects of masculine identity and gender role stress on aggression in men." *Psychology of Men & Masculinity*, vol. 7, no. 4, p. 179, 2006.

[61] W. Pollack and T. Shuster, *Real boys' voices: Boys speak out about drugs, sex, violence, bullying, sports, school, parents, and so much more*. Random House, 2000.

[62] M. Oransky and J. Marecek, ""i'm not going to be a girl" masculinity and emotions in boys' friendships and peer groups," *Journal of adolescent research*, vol. 24, no. 2, pp. 218–241, 2009.

[63] M. Lehdonvirta, Y. Nagashima, V. Lehdonvirta, and A. Baba, "The stoic male: How avatar gender affects help-seeking behavior in an online game," *Games and culture*, vol. 7, no. 1, pp. 29–47, 2012.

[64] N. Yee, N. Ducheneaut, M. Yao, and L. Nelson, "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*, 2011, pp. 773–776.

[65] R. M. Simon, A. Wagner, and B. Killion, "Gender and choosing a stem major in college: Femininity, masculinity, chilly climate, and occupational values," *Journal of Research in Science Teaching*, vol. 54, no. 3, pp. 299–323, 2017.

[66] P. Briñol and R. E. Petty, "Overt head movements and persuasion: A self-validation analysis." *Journal of personality and social psychology*, vol. 84, no. 6, p. 1123, 2003.

[67] J. M. Bugg, N. A. Zook, E. L. DeLosh, D. B. Davalos, and H. P. Davis, "Age differences in fluid intelligence: Contributions of general slowing and frontal decline," *Brain and cognition*, vol. 62, no. 1, pp. 9–16, 2006.

[68] T. C. Peck and A. Tutar, "The impact of a self-avatar, hand collocation, and hand proximity on embodiment and stroop interference," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 5, pp. 1964–1971, 2020.

[69] A. Maselli and M. Slater, "The building blocks of the full body ownership illusion," *Frontiers in human neuroscience*, vol. 7, p. 83, 2013.

[70] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze, "" these are not my hands!" effect of gender on the perception of avatar hands in virtual reality," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1577–1582.

[71] H. Cohen and J. J. Levy, "Sex differences in categorization of tactile stimuli," *Perceptual and motor skills*, vol. 63, no. 1, pp. 83–86, 1986.

[72] R. Carrasco, S. Baker, J. Waycott, and F. Vetere, "Negotiating stereotypes of older adults through avatars," in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, 2017, pp. 218–227.

[73] A. Puri, S. Baker, T. N. Hoang, and R. C. Zuffi, "To be (me) or not to be? photorealistic avatars and older adults," in *Proceedings of the 29th Australian conference on computer-human interaction*, 2017, pp. 503–507.

Tabitha C. Peck is an Associate Professor of Mathematics and Computer Science at Davidson College. Dr. Peck received her PhD in Computer Science from The University of North Carolina at Chapel Hill in 2010. Dr. Peck's research involves developing and testing usable virtual reality systems. Her current research includes investigation of the psychological implications of embodiment in self-avatars with the goal of using avatars to reduce and mitigate bias. Dr. Peck has been both a Journal and Conference Paper Program Chair for IEEE VR and a Journal Paper Science and Technology Chair for IEEE ISMAR. She is a Review Editor for Frontiers in Virtual Reality, an Associate Editor for Presence, and an Associate Editor for TVCG. She has received numerous honorable mentions and nominations for best paper awards at IEEE VR, and received a NSF CAREER award in 2020. Dr. Peck is an IEEE Senior Member.

Jessica J. Good is Associate Professor and Chair of the Psychology Department at Davidson College. Dr. Good received her PhD in Social Psychology from Rutgers University in 2011. Dr. Good's research involves stereotyping and discrimination related to self-identities and self-evaluations as well experiences and perceptions of discrimination and prejudice. Dr. Good is a Fellow of the Society for Experimental Social Psychology and the Society for Personality and Social Psychology. She also serves as Co-Chair of the Diversity and Climate Committee within the Society for Personality and Social Psychology. Her research has been funded by the National Science Foundation and the Society for the Psychological Study of Social Issues.