Optimizing Performance through Stress and Induction Levels in Virtual Reality Using Autonomic Responses

Dan Archer*
Department of Computer Science
University College London

Anthony Steed†
Department of Computer Science
University College London

ABSTRACT

Virtual reality (VR) is now a consumer technology, but individuals' experience with systems or different applications varies enormously. This means that most consumer applications need to consider how to train naive users in the application's operation. We examine three different ways of imparting on-boarding instructions to users: first-person audio guidance, second-person non-player character (NPC) diegetic guidance or written instruction. Our primary hypothesis is that the second-person condition will induce a higher stress level on the user, given the perceived presence of a supervising NPC. Our secondary hypothesis is that there is a correlation between stress and performance, meaning that participants with elevated stress levels within a certain margin will complete their tasks faster and more successfully. By extension, participants whose stress levels are either above or below this optimal margin will under-perform on the same tasks. The tasks in question are an interaction test (IT), designed to test participants' abilities to pick up and manipulate virtual objects, and a mental rotation test (MRT), designed to place them under cognitive load. During these tasks we measure the users' level of stress from their bio signals via a mobile wearable device that tracks their heart rate (HR), galvanic skin response (GSR) and body temperature in real-time. Statistical significance was not found in the stress or performance levels between the instruction conditions, but the secondary hypothesis was supported and a correlation was found between stress and performance levels across the conditions in both HR and GSR.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality;

1 INTRODUCTION

The breadth of virtual reality (VR) experiences available to the general public is growing rapidly driven by the availability of cheap, capable devices. However, given that VR is a relatively new medium, there is a very wide range of levels of user familiarity and competence. Thus almost all consumer VR experiences need to impart instructions to users and this instruction might have to assume very little familiarity with the medium. Instructions could take the form of audio (either disembodied or from a diegetic non-player character (NPC)) or textual annotation. They could be adapted to the user in the form of a monologue (in the first person, "I need to...") or an instruction (the second person, "you need to..."). The method of instruction thus not only determines a user's subsequent ability to operate within a virtual space, but might also prove stressful in itself.

A key assumption of our research is that bio signals (heart rate, galvanic skin response, temperature) provide a suitable quantitative measure of stress, and that each user has an optimal range of these measure during a virtual reality experience (VRE). This is based on the well-established relationship between pressure and performance first proposed by Yerkes-Dodson in 1908, the so-called "Law of Arousal". The essential premise is a curvilinear relationship between motivation and performance, following a inverted U-model, in which optimal performance occurs under a moderate amount of arousal, but declines when that level of arousal is too high or too low. Broadhurst [3] improved on the original experiment by incorporating four motivation levels and three difficulty levels. Nixon's work in 1979 incorporated the Stress Response Curve [13] and Klein [9] related the efficiency of memory to arousal or stress.

Bringing this approach to performance in VR draws on the work of Claude et al. [4] who studied similar stress-related effects within the context of mental workload inside immersive training scenarios, Parsons et al. [16] who examined adaptive virtual environments for neuropsychological assessment in serious games using bio signals, and Luong et al. [10] who focused on real-time recognition of users' mental workload, adapting the Multi-Attribute Task Battery (MATB-II) test originally designed by NASA [5] into a VR cockpit.

Thus in this paper we explore the main hypothesis that the level of stress experienced by a user will directly impact their performance when instructed to perform a series of tasks in VR. We then seek to manipulate stress by controlling the method in which instruction is given. If instruction is given in a manner that induces a sufficient level of stress, their performance is hypothesised to be optimal, represented by a faster task completion time and greater variation in their bio signals from the baseline. In contrast, if the user's instructions are manipulated to not trigger sufficient stress, their task performance is hypothesised to be sub-optimal, resulting in slower task completion times and reduced absolute difference from the mean HR and GSR. Manipulating the instruction conditions might be an effective way to heighten or lessen the degree of stress, while also serving as a test of the physiological measurements. This should be of broad interest, because instruction and induction into new experiences is one area where designers have a lot of freedom.

The aim of the research is to build a baseline measurement that will determine which instructional method is the most effective, as well as laying the groundwork for a bespoke virtual reality system that gathers bio signal data from wearable sensors and adapts the virtual environment accordingly in real-time.

2 RELATED WORK

We are interested in the relationship between performance and stress in the virtual environment. The working assumption is that the participant will be embodied and feel present in the environment. Thus we review the core components of virtual embodiment and presence and we discuss how the induction style might impact on user experience.

The construction of a credible virtual reality environment is dependent on four main illusions: place illusion, plausibility illusion, body ownership and a sense of agency [22]. Place Illusion (PI) is a sense of being in the place depicted by VR, irrespective of where we are. Plausibility illusion (Psi) convinces the user that events in VR are actually occurring, a quality that is hard to maintain when characters don't react or interact with the user. These two are described as core components of "presence" [19]. Presence is a compound

---

*e-mail: daniel.archer.18@ucl.ac.uk
†e-mail: a.steed@ucl.ac.uk

cognitive state whereby the brain takes sensory stimuli and translates it into the sense of “being there” [25]. Witmer and Singer’s development of an immersive tendencies questionnaire (ITQ) measured differences in the way individuals tend to experience presence, which they cross-referenced with a presence questionnaire (PQ) for internal consistency and reliability. They noticed a weak but consistent positive relation between presence and task performance in VEs, and that individual tendencies as measured by the ITQ predict presence as measured by the PQ [26].

Body ownership and the sense that a virtual avatar is your body [27] has emerged as an important component of user experience in immersive systems. Botvinick and Cohen’s seminal paper [2] revealed the three-way interaction between vision, touch and proprioception (awareness of the position and movement of the body) through the Rubber Hand Illusion (RHI). Steed and Yuan, Zhang and Hommel as well as Sanchez-Vives and Slater [19,29,30] demonstrated that the RHI technique applied to a virtually embodied hand mediated by an HMD. That is, users demonstrated or expressed feelings of ownership over a virtual hand. Last of the aforementioned four qualities for a credible VRE, agency, is a key condition to ensuring that the experience does not feel too passive: the actions of the virtual body are attributed to our self [7].

Building on these fundamentals, how does self-representation as avatar moderate performance in VREs? Embodiment in the first-person perspective of an avatar moving in synchronicity with the user may help participants to overcome cognitive loads, increasing their memory performance after the VR experience [23]. In Steed’s experiment, participants embodied in an avatar whose virtual hands could be moved synchronously with their own hands had significantly higher recall of pairs of memorised letters after performing a spatial recognition task. Pan et al. later replicated the experiment using avatars with just virtual hands, corroborating the same findings [15]. Tutar and Peck [17] also used a VR version of the Stroop task (VRST) to compare user performance with or without collocated hands, finding that proximal hands produced a significant increase in accuracy compared to non-proximal hands. Interestingly, they found that Stroop interference was not mediated by the existence of a self-avatar or level of embodiment.

Next we consider external factors that effects stress and performance in VREs. Wu et al. embedded a VRST inside an immersive military simulation, with participant reaction time under varying degrees of stress as a performance measure [28]. They demonstrated that when reaction time is used as the performance measure, one of three stimuli presentations (moderate stress) elicited the optimal level of arousal for most (11 of 18) subjects. Furthermore, results suggested that high classification rates could be achieved when a support vector machine was used to classify the psycho-physiological responses (skin conductance level, respiration, ECG, and EEG) in these three stimuli presentations into three arousal levels.

Claude et al. [4] studied similar performance effects in relation to mental workload inside immersive training scenarios. Lecuyer et al. [11] highlighted improvements when comparing performance with VR feedback to 2D display feedback in the context of VR brain-computer interfaces. Luong et al. (2020) [10] also looked at real-time recognition of users’ mental workload, adapting the Multi-Attribute Task Battery (MATB-II) test designed by NASA into a Virtual Reality cockpit. Palmas et al. [14] made the comparison between a gamified (progress bar, points, sounds, visual feedback) and non-gamified vr training task - assembling a virtual drum set - using the metrics of completion time and number of errors made.

Lastly is the external impact on the user from the VRE induction protocol itself, most notably in the effect of avatar observation on user performance, as investigated by Hayes et al. [8], and within the context of social facilitation by Sterna et al. [24]. Also of note is Blascovich’s research around the biopsychosocial model of challenge and threat in relation to external observers, in which participants who performed a novel task in the presence of others had both increased cardiac response and increased vascular resistance from baseline [1].

3 EXPERIMENTAL CONDITIONS

In a between-subjects design, participants were randomly assigned one of three possible conditions at run time, which dictated the manner in which they received their instructions. Each condition had a different manner of presenting the instructions for the experiment:

1. First person instruction via audio narration: ”I need to pick up the sphere”

2. Second person diegetic instruction given with audio by a virtual NPC standing in the corner of the room: ”You need to pick up the sphere”

3. Written (diegetic) instructions that appear on-screen: ”Pick up the sphere”

Of the conditions listed, the second condition was hypothesised to produce the optimal performance, given its similarity to real-life instruction. See Figure 1 for screenshots from the different conditions.

3.1 Protocol

After arriving at the lab, reading the information sheet and being briefed, which provided them with sufficient time for their bio signals to return to normal, participants began the experience by donning an EVU TPS wireless Bluetooth wearable on the index finger of their non-dominant hand as well as the Quest 2 HMD. Their baseline bio signals were then recorded for two minutes, constituting an Ultra-Short-Term (UST) experimental norm for HR measurement [21]. They then started the VR experience and completed a virtual survey relating to their VR consumption habits and whether they own a headset (*survey scene*). This was to gauge their familiarity with navigating similar scenarios and obtain a pre-treatment baseline. They were then instructed to select the gender of their avatar.

Next they entered a scene where they seated at a table in front of a virtual mirror. Later we will refer to this as the *interactive scene*. All participants were given the same gender-selectable avatar corresponding to their own gender with a basic level of realistic animation applied (idle, with blinking eyes), synced to their hand movements via the Oculus Quest controllers. Their hands with articulated fingers were visible within short range and their entire bodies visible in the mirror directly in front of them.

In this scene participants were asked to complete a number of simple tasks: pick up a sphere from the table; pick up some keys from the table; use the keys to unlock a drawer in the table; take a battery out from the drawer and place it inside the alarm - adding the battery to the alarm turns it on; pick up a pen to draw a five-pointed star on a piece of paper, using the alarm timer as a countdown; and finally, to take a picture of the drawn star by picking up and aiming a virtual phone at it.

Once this scene was completed, they then moved into the last scene (*spatial scene*) to undertake a challenging mental rotation test, designed to place them under cognitive load (see Figure 2). The test used was the same as that in Pan and Steed’s 2016 study [23]. Participants were tasked with choosing a pair from four diagrams showing different illustrations of a rotated shape. They were shown 23 rounds in total (with the first 3 rounds constituting practice rounds) and each round was timed to last 25 seconds. The spatial scene was identical across all groups of participants, regardless of the instruction style in the previous scene. Once the spatial test was completed, participants were then asked to complete a questionnaire answering their perceived level of embodiment, presence and stress during the experiment. The whole experience lasted approximately 20 minutes.

Figure 1: First-person induction, second-person induction with NPC and written induction conditions

The study was approved by an anonymous ethics committee. Participants were recruited following a process (anonymised for review).

Answer as quickly as you can.

Please practice on 3 test questions.

Figure 2: The Mental Spatial Rotation Test

3.2 Measures

The study was comprised of three distinct phases: the survey scene, the interaction scene, and the spatial scene. In terms of performance measures, the completion times of the assorted tasks in the interactive scene were compared, as were the number of correct answers in the spatial scene. Participants received two points for every correct pair of answers they submitted, provided the pair was inputted within the 25 second time limit for each question. Points were then summed out of the total number of correct answers (3 practice pairs followed by a further 20, totalling a possible 40 points) to derive a percentage score. Bio signal data was recorded during all three scenes, showing Blood Volume Pulse (BVP) (from which HR data was extracted), Skin Conductivity (SC), and Temperature. A post-treatment, non-VR survey was also sent to participants to qualitatively measure their level of immersion and stress.

Stress measurements were obtained by calculating the absolute difference between the mean heart rate for a baseline measurement (taken during the survey scene, when the participant was instructed to stand idle) and for the instruction/spatial test scenes. The heart rate (HR) is the number of heartbeats per minute (bpm) and is derived from the raw Blood Volume Pulse (BVP) signal. The BVP signal is an optical detection of the pulsatile blood flow resulting from heart beats, sampled at a rate of 300 hz. The data was cleaned of anomalies such as improbably clustered repeated values and unsustained (i.e. 2 seconds or less) spikes or drops of more than 30 bpm, which indicated a temporary malfunction in the wearable, such as ambient light contamination on the PPG sensor. The same processes were repeated for the galvanic skin response (measured in micro-siemens) and skin temperature (measured in degrees Celsius) recordings. We hypothesise that the second condition will produce a high stress response in the form of a higher mean difference from baseline across all bio signals compared to the two other conditions in the interactive scene, as well as a higher score in the subsequent spatial scene.

3.3 Hypotheses

We hypothesise that a participant's autonomic responses provide a good indication of their level of embodiment and presence in the scenario, as well as their level of stress. Our main hypothesis is that the level of stress experienced by a user will directly impact their performance when instructed to perform a series of tasks in VR, and that we can manipulate said level by altering the way in which they are given task instructions. Improved task performance will be represented by a faster task completion time on the interactive test, higher score on the spatial test, and greater difference from their mean HR, GSR and temperature during both tests when compared to those taken during the baseline scene. By contrast, if the user's instructions are manipulated to not trigger sufficient stress, their task performance is hypothesised to be sub-optimal, resulting in slower task completion times and reduced absolute difference from the mean HR and GSR. If true, manipulating the instruction conditions would be an effective way to heighten or lessen the degree of stress, while also serving as a test of the physiological measurements.

4 RESULTS

Twenty-three participants completed the study - six were female and the average age was 24. Of these, three were removed due to inaccuracies in the wearable sensor readings, caused by movement or software failure. Of the final 20, five were female and the age was 23. The ages ranged from 18 to 51. As a result, six participants tried condition 1, seven condition 2 and seven condition 3. The mean age per condition was 19, 27 and 23 for conditions one, two and three respectively. Recruitment was largely focused within the engineering faculty, which explains the higher proportion of male participants.

In terms of VR experience, the mean score for prior VR usage was based on two factors: the number of VR experiences played and the type of VR headset ownership. The former was divided as follows on a 1-5 scale: zero; less than five; less than ten; less than thirty; over fifty. The latter was divided by ascending order in price of VR headsets, based on the rationale that the more expensive (and greater number of - since users could also select multiple options)

Condition	Baseline Mean HR (bpm)	Baseline Mean GSR (micro-siemens)	Baseline Mean Temp (degrees Celsius)	HR Diff Baseline-Interactive Scene	GSR Diff Baseline-Interactive Scene	Temp Diff Baseline-Interactive Scene	Interactive Completion Time (secs)	HR Diff Baseline-Spatial Scene	GSR Diff Baseline-Spatial Scene	Temp Diff Baseline-Spatial Scene	Spatial Score (%)
One	73.03	4.61	29.83	5.55	1.08	1.45	80	6.55	1.25	1.46	70
Two	70.6	3.8	33.58	14.43	1.07	0.79	76.86	18.09	1.22	0.66	72
Three	83.9	2.94	30.21	4.29	0.93	0.56	95.71	7.49	1	0.42	65

Table 1: Mean measurements across all scenes for all conditions

headsets owned, the more a participant had invested in the medium. This was divided into: none; Google Cardboard; Oculus Go/Gear VR; Quest/Pico; Valve, Index, Rift. The combined mean usage score was 2/5 for condition one, and 3/5 for conditions two and three, despite the randomisation of condition allocation. It is accepted that an increased sample size could improve the aforementioned discrepancies in gender distribution and experience across conditions. Sangier et al. [18] found that gender had an effect on participant self-assessment and the ability to act during an assembly task in VR, while prior experience had an effect on performance, pragmatic quality and hedonic quality stimulation.

As shown by the data in Table 1, condition 1 (first person instruction) during the interaction and spatial scenes had the highest average difference from the baseline mean in terms of GSR and temperature, although the amount was negligible compared to the second condition (0.01 and 0.03 microsiemens respectively) and the results were not statistically significant, producing a p value greater than 0.05 from a single factor ANOVA. Condition two showed the greatest average difference from baseline in heart rate (14.43 beats per minute), almost triple the second-highest difference of 5.55 from condition one and 4.29 from condition three, although the results were not statistically significant (p=0.257) in a single factor ANOVA. The change in absolute difference from the HR baseline was also visible in the spatial test scene, with the mean difference of condition two again almost triple that of the other two. By contrast, condition 3 (textual instruction) had the lowest level of absolute difference in HR, GSR and temperature from the baseline during the interaction scene, indicating the lowest level of stress. During that scene, both condition 1 and 2 had similar average completion times of 80 and 77 seconds respectively. But condition 3 had a slower completion time: almost 20 seconds slower than the fastest (condition 2), although this was not statistically significant ($p>0.05$). One explanation for this could be how vergence accommodation conflict can lead to decreased cognitive performance. This may partially explain the decrease in task performance seen in the written instruction task, which likely requires users to accommodate to the focal depth of the display while converging on the simulated distance of the text-based instructions [6].

This same level of increased stress in the second condition was reported in the self-reported questionnaire, which can be seen clearly in Fig 4, despite not being statistically significant (p=0.07). Although the written condition elicited the greatest variance in heart rate during the interactive test itself, it was the second-person instruction condition that produced the greatest difference from baseline, thereby manifesting the highest recorded increase in stress levels (3/5, compared to 2/5 for the other two).

The data from the spatial rotation test, as seen in Table 1, shows the highest mean score was produced by participants who were given the second condition in the previous scene, whose mean HR values differed the most from the baseline mean. This potentially implies a correlation between stress and performance, albeit only anecdotally. By contrast, further supporting this correlation, those given the third condition showed the least stressed bio signals and came lowest of the three conditions by 5 percent in the spatial score. They showed the smallest difference in mean HR compared to the baseline during the interactive scene, as well as the lowest difference in mean GSR from baseline in both the interactive and spatial scene.

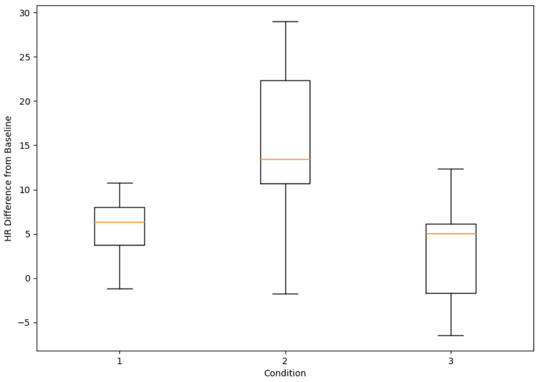

Figure 3: Heart Rate Difference from Baseline to Interaction Scene

In a two-tailed test, statistical significance was found between the average deviation from baseline between the interaction scene and spatial test scene for each condition ($p<0.0005$ for all conditions) confirming the experimental data from Pan and Steed's experiment in which the spatial rotation test was used to provoke a stress response in participants. With this in mind, a single factor ANOVA test was performed to plot stress levels against performance across all three conditions. Stress level was calculated by the absolute difference in HR from the baseline in both the interactive and mental spatial rotation tests. Statistical significance was found across all three conditions ($p<0.0005$ for each of conditions one, two and three), and is plotted using an average across both scenes and all conditions to show the trend line shown in Figure 5 below.

A similar correlation between the absolute difference in GSR levels during the baseline measurement to GSR levels during the interactive/mental rotation tests when compared to the final mental rotation score was also found to be statistically significant in a two-tailed ANOVA ($p<0.0005$ for each of conditions one, two and three). Similar to HR however, the manipulation between conditions was not sufficient to elicit a statistically significant difference in GSR. The same was true of the temperature measurement, which failed to produce a statistically significant difference between conditions when compared to the baseline (p=0.23).

5 CONCLUSIONS AND FUTURE WORK

Our primary hypothesis, that the manner of induction would affect the stress of the participant was unsupported. While promising results were found, the conditions were insufficiently distinct to provoke a quantifiable difference. Thus the results are a partial counter to the prior work that indicates that being observed doing a task causes stress. However due to the trend of the results, we would suggest that this type of manipulation deserves further study.

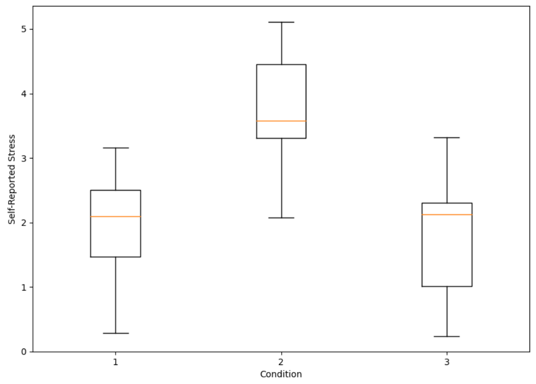

Figure 4: Self-reported Stress Levels on 5-point Likert Scale

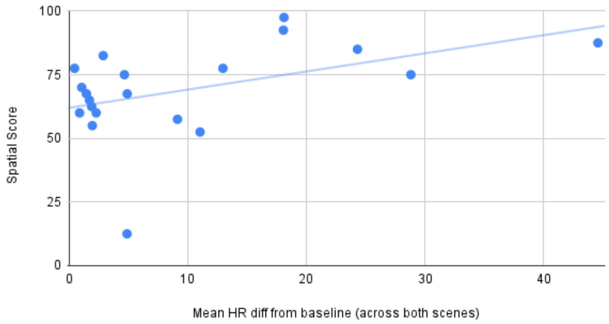

Figure 5: Stress versus performance in the spatial scene

One potential means of increasing stress during the interactive scene would be to make the NPC in the second-person induction condition more critical of the user's performance to provoke more negative affect, in addition to including a timer countdown for any of the interactions.

We found good evidence that the HR and GSR measures were able to capture the expected rise in stress levels during the spatial rotation test when compared to the baseline. This is useful support for the prior work that has investigated the manipulation of stress in VR. The spatial rotation test we used is very easy to embed in other experiments, and the images used are openly available. Further, we validated the effectiveness of a novel portable bio signal device.

We also were able to demonstrate good support for the secondary hypothesis: that stress level would affect performance. Elevated heart rate and skin conductivity were correlated to an improvement in the spatial test scores. Thus while we failed to manipulate the stress through the conditions, this gives us good evidence that if stress level can be manipulated, this can moderate performance.

As stated in the introduction, the broad hypothesis of this research was to correlate an optimal level of stress or arousal with an improvement in task performance. Our findings are analogous to Luong et al.'s work [10] on VR task performance under stressful conditions derived from the Multi-Attribute Task Battery (MATB-II) [5], since we were able to manipulate stress level through different tasks. Thus we can infer that the interaction tasks weren't difficult enough, or the level of instruction insufficiently distinct.

In future work, we aim to complicate the instruction tasks by situating them within a narrative-driven, reconstructed social scenario, while also adding additional stressors such as timers and more responsive NPCs. A future aim of this research is to measure whether task performance (such as cued recall) under increased stress can deepen embodiment and perspective-taking towards outgroup avatars. This is similar to the VR exposure therapy work around embodied perspectives of Neyret et al. [12] and Seinfeld et al. [20].

ACKNOWLEDGMENTS

The authors wish to thank Klara Brandstaetter, Harry Liversedge, and Tony Steffert. This work was supported by UCL's Doctoral Training Partnership grant from the UK Engineering and Physical Sciences Research Council (EPSRC) .

REFERENCES

[1] J. Blascovich, W. B. Mendes, S. B. Hunter, and K. Salomon. Social” facilitation” as challenge and threat. *Journal of personality and social psychology*, 77(1):68, 1999.

[2] M. Botvinick and J. Cohen. Rubber hands ‘feel’ touch that eyes see. *Nature*, 391(6669):756–756, Feb. 1998. doi: 10.1038/35784

[3] P. Broadhurst. The interaction of task difficulty and motivation: The yerkes dodson law revived. *Acta Psychologica, Amsterdam*, 1959.

[4] G. Claude, V. Gouranton, and B. Arnaldi. Versatile scenario guidance for collaborative virtual environments. *GRAPP*, 15(3):6, 2015.

[5] J. R. Comstock Jr and R. J. Arnegard. The multi-attribute task battery for human operator workload and strategic behavior research. Technical report, 1992.

[6] F. Daniel and Z. Kapoula. Induced vergence-accommodation conflict reduces cognitive performance in the stroop test. *Scientific reports*, 9(1):1–13, 2019.

[7] P. Haggard and V. Chambon. Sense of agency. *Current Biology*, 22(10):R390–R392, May 2012. doi: 10.1016/j.cub.2012.02.040

[8] A. L. Hayes, A. C. Ulinski, and L. F. Hodges. That avatar is looking at me! social inhibition in virtual worlds. In *International conference on intelligent virtual agents*, pp. 454–467. Springer, 2010.

[9] S. B. Klein et al. *Motivation: biosocial approaches*. McGraw-Hill College, 1982.

[10] T. Luong, F. Argelaguet, N. Martin, and A. Lécuyer. Introducing mental workload assessment for the design of virtual reality training scenarios. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 662–671. IEEE, 2020.

[11] A. Lécuyer, F. Lotte, R. B. Reilly, R. Leeb, M. Hirose, and M. Slater. Brain-computer interfaces, virtual reality, and videogames. *Computer*, 41(10):66–72, 2008. doi: 10.1109/MC.2008.410

[12] S. Neyret, X. Navarro, A. Beacco, R. Oliva, P. Bourdin, J. Valenzuela, I. Barberia, and M. Slater. An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later milgram obedience scenario. *Scientific reports*, 10(1):1–18, 2020.

[13] P. Nixon, R. Murray, and C. Bryant. Stress response curve. Technical report, Retrieved 8/6/17, from https://explorable. com/how-does-stress-affect . . . , 1979.

[14] F. Palmas, D. Labode, D. A. Plecher, and G. Klinker. Comparison of a gamified and non-gamified virtual reality training assembly task. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–8. IEEE, 2019.

[15] 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*, pp. 1–4, 2019.

[16] T. D. Parsons and J. L. Reinebold. Adaptive virtual environments for neuropsychological assessment in serious games. *IEEE Transactions on Consumer Electronics*, 58(2):197–204, 2012.

[17] 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*, 26(5):1964–1971, 2020. doi: 10.1109/TVCG.2020.2973061

[18] C. Sagnier, E. Loup-Escande, and G. Valléry. Effects of gender and prior experience in immersive user experience with virtual reality. In *International conference on applied human factors and ergonomics*, pp. 305–314. Springer, 2019.

[19] M. V. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332, Apr. 2005. doi: 10.1038/nrn1651

[20] S. Seinfeld, J. Arroyo-Palacios, G. Iruretagoyena, R. Hortensius, L. E. Zapata, D. Borland, B. de Gelder, M. Slater, and M. V. Sanchez-Vives. Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Scientific reports*, 8(1):1–11, 2018.

[21] F. Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, p. 258, 2017.

[22] M. Slater. How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 13(4):484–493, Aug. 2004. doi: 10.1162/1054746041944849

[23] 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)*, pp. 67–76. IEEE, Greenville, SC, USA, Mar. 2016. doi: 10.1109/VR.2016.7504689

[24] R. Sterna, P. Strojny, and K. Rkebilas. Can virtual observers affect our behavior? *Social Psychological Bulletin*, 14(3):1–18, 2019.

[25] M. Wells. Virtual reality: technology, experience, assumptions. *Human Factors Society Bulletin*, 35(9):1–3, 1992.

[26] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, June 1998. doi: 10.1162/105474698565686

[27] A. S. Won, J. Bailenson, J. Lee, and J. Lanier. Homuncular Flexibility in Virtual Reality. *Journal of Computer-Mediated Communication*, 20(3):241–259, May 2015. doi: 10.1111/jcc4.12107

[28] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie, and T. D. Parsons. Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task. *IEEE Transactions on Affective Computing*, 1(2):109–118, 2010.

[29] Y. Yuan and A. Steed. Is the rubber hand illusion induced by immersive virtual reality? In *2010 IEEE Virtual Reality Conference (VR)*, pp. 95–102, Mar. 2010. doi: 10.1109/VR.2010.5444807

[30] J. Zhang, K. Ma, and B. Hommel. The virtual hand illusion is moderated by context-induced spatial reference frames. *Frontiers in psychology*, 6:1659, 2015.