Modeling but NOT Measuring Engagement in Computer Games

Mark Chen, College of Education, University of Washington, 1100 NE 45th St, Suite 200, Seattle, WA 98105, markchen@uw.edu
Beth E. Kolko, Human Centered Design & Engineering, University of Washington, Box 352315, Seattle, WA 98195, bkolko@uw.edu
Elisabeth Cuddihy, Human Centered Design & Engineering, University of Washington, Box 352315, Seattle, WA 98195, ecuddihy@uw.edu
Eliana Medina, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA 98109, emedina@gmail.com

Abstract: This paper describes a study that was performed to define a model of "engagement" in digital games and the problems encountered with our testing methods. Drawing upon multiple disciplines, our working model of engagement was meant to help describe, predict, and analyze the conditions that create a high level of engagement in players. To refine the model and develop a methodology for studying engagement, an exploratory prototype study was performing in which participants were observed playing a pair of games (*The Curse of Monkey Island* and *The Oregon Trail 5th Edition*). This study used common usability testing methods along with a pre- and post-test modeled after Witmer and Singer's Presence Questionnaire and a flow test at timed intervals. Unfortunately, it became clear that our testing methods needed refinement, though we believe the engagement model may still be useful as a common artifact informed from multiple disciplines.

Introduction

In 2005, the Digital Games Research Group (DGRG) at the University of Washington presented a model of engagement in games (Chen et al., 2005) that was informed by diverse disciplines including game design theory, presence literature from virtual reality (VR) and simulations research, narrative immersion from literary theory, and motivation literature from psychology and cognitive science. Our theoretical model was comprehensive at the time, and we believe it is still a very useful model to think about how to measure engagement with games as a product of user interface, realistic or consistent simulation and systems modeling, and narrative and role-play.

To measure engagement using our model, we created a data collection toolkit for use in a lab setting. These included a pre- and post-game series of questions based on Witmer and Singer's presence questionnaire (1998), a mini-survey based on flow theory (Csikszentmihalyi, 1990), detailed forms for researchers to fill out while observing participants playing, and post-game interview questions. To validate the model, we conducted a few initial pilot tests where participants played a commercial game (*The Curse of Monkey Island* AKA *Curse*) that we knew was "good" via its average meta-review score on gamerankings.com. We compared this with an educational game (*The Oregon Trail 5th Edition* AKA *The Oregon Trail*), hypothesizing that the commercial game would score higher than the educational one and that our measurements for *Curse* would reflect its aggregate gamerankings score.

Unfortunately, the results of our pilot tests failed to give us measures that reflected the metascore for *Curse*, and, what's more, *The Oregon Trail* scored higher for our participants! Possible reasons for this include the fact that many game reviews are not written until the reviewer has finished the game, that many memorable and immersive elements to a game's story do not occur until hours into a game, and that we did not run enough participants in our initial tests to have anything statistically reliable. While our testing toolkit was well suited to uncover issues with usability, it was ill equipped to shed light on the affective measures of engagement with a game's full experience. We shared our model that year (Chen et al., 2005) but did not move forward with validating it and never produced a final research paper.

This paper will cover our model and its theoretical underpinnings, which we believe to be extremely timely and important, as evidenced by other scholars from around the world continuing to cite our work from 2005. Sharing our model and how we failed to measure it is also important because there seems to be a new push in games for learning research on measuring engagement that may be following in our footsteps by not including methods that are ecologically valid. Thus, this paper presents a case where data collection methods failed to provide a good way to validate a model of engagement. We will

also discuss how this helped shape our early careers as games scholars (e.g., pushing Chen into ethnography) and our current thoughts on how new research methods could be used to finally validate our model of engagement.

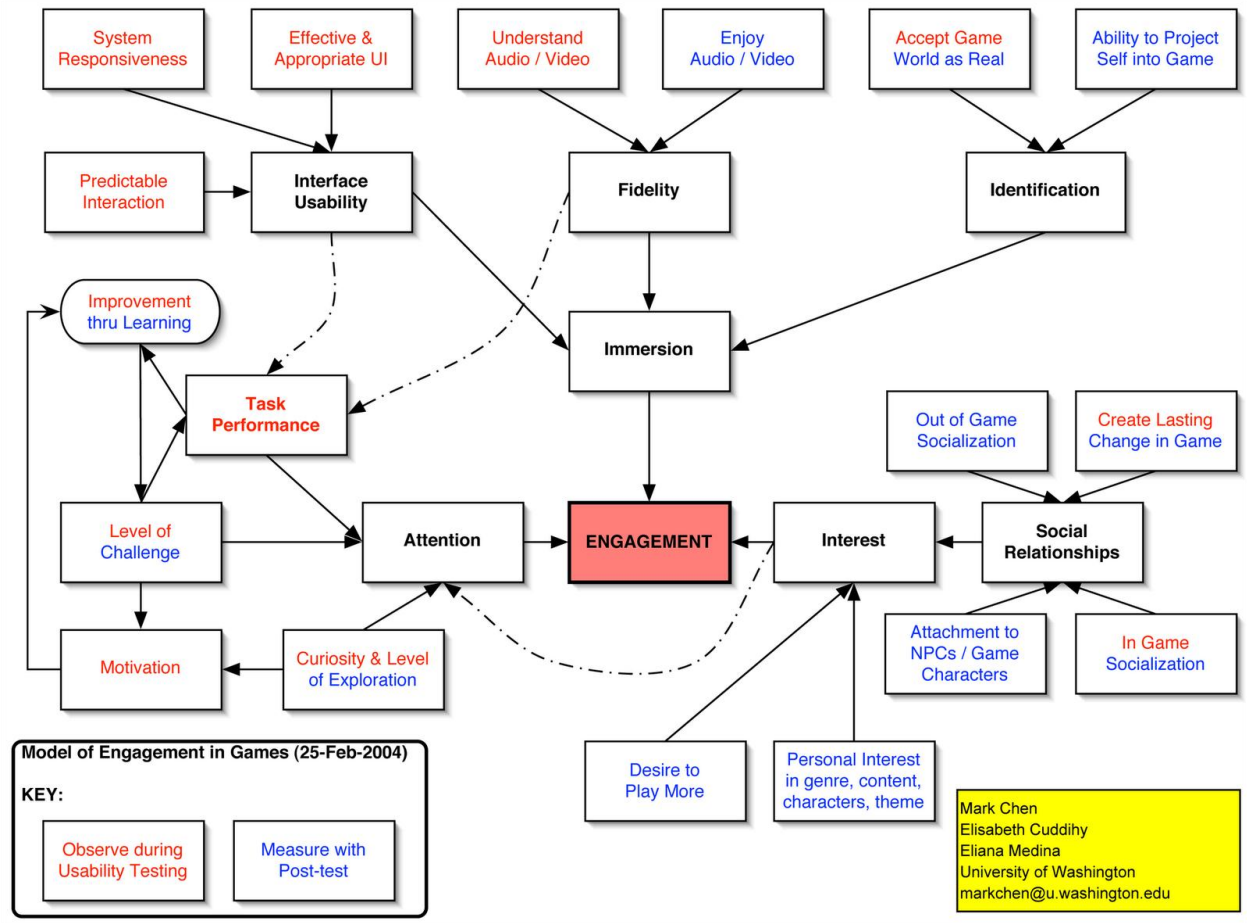

Figure 1. Digital Games Research Group's model of engagement in games circa 2004/2005.

Modeling Engagement

When DGRG first started this project, a number of us were just starting our graduate school careers, and we were taking a multidisciplinary approach to our studies. This let us see that different disciplines over the years have taken different strategies to measure how people interact with computers and software. We saw, for example, that while VR literature had been focusing on *presence* (Zeltzer, 1992; Heeter, 1992; Bystrom et al., 1999; Witmer & Singer, 1998), games people from communications and information studies were also trying to define "fun" (Heeter et al., 2003). While cognitive psychologists and educational technologists were focusing on (intrinsic and extrinsic) motivation (Malone, 1980; 1982; Malone & Lepper, 1987; Keller & Suziki, 1988; Alessi & Trollip, 2001), scholars from education thought about character identification and role-playing (Gee, 2003). The games industry was (and still is) interested in generating emotion (Lazzaro, 2004) and feedback loops (Prensky, 2000; Crawford, 1982), while Csikszentmihalyi (1990) came up with flow theory.

All of these literatures were attempting to define (immersion, presence, engagement, affect, motivation) in some sort of functional way that allowed future researchers to measure and compare different experiences, working on the assumption that higher engagement led to deeper learning, more meaningful experiences, longer sustained interaction, etc. We took what we could find that focused on digital games or educational software (without claiming it was an exhaustive list) and iterated through a couple of conceptual models for engagement, the last version can be found in Figure 1.

Mayes and Cotton (2001) define engagement with respect to computer games as how fun, involving, and motivating a task is. Regarding computer-based learning environments, Jones (1998) defines engagement as a combination of the knowledge, interest, and stimuli that promote initial interest and continued use of an environment. Building upon these definitions, we define engagement as a sustained level of involvement caused by capturing a person's interest, holding the majority of a person's attentional resources, and placing the person in an immersive state. These three factors are covered (in brief) next.

Interest

The first prerequisite for engagement is the level of interest that a person has in a game's content, presentation, characters, theme, and genre. Additionally, interest is reflected in a person's desire to continue playing a game. For multiplayer games and games that have out-of-game dedicated online communities, interest can also be measured by the level of interaction that a player has in communities devoted to game discussion or modification, design of game tutorials, provision of game tips, seeking out or creating game mods, and seeking out or creating fan art and fan fiction. These out-of-game experiences enhance the level of interest that a player already experiences in-game.

In a lab, interest can be investigated by inquiring about a player's level of personal interest in a particular kind of game, genre, and theme. During game-play, a player's desire to continue playing can be sampled at regular frequencies. If appropriate, a player can be asked about their involvement in out-of-game community activities, and within longitudinal multiplayer studies, the effects of social interaction on interest can be analyzed.

Attention

Holding the majority of a person's attentional resources is another requirement of engagement. When attending to a task, a person diminishes or blocks out stimuli that is outside of their locus of attention.

Attention can be observed during game-play to see how focused the player appears on the task at hand and how they focus their attention on in-game challenges. Overall decrease in task performance due to boredom or frustration from inappropriate challenge levels signals a lessening of attentional resources directed toward the game. Finally, a player's level of curiosity and desire to explore may reflect level of attention.

Immersions

Immersion has been defined both qualitatively and quantitatively. Witmer and Singer (1998) describe immersion as a psychological state "characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continual stream of stimuli and experience" (p. 227). Bystrom et al. (1999) and Slater (1999) describe immersion by the quantifiable features of a system, including its visual and audio fidelity and impact. For the purposes of this paper, we define engagement as 1) the psychological state of being enveloped by a system, which is mediated by the system's physical interface, logical interface, and output fidelity, and 2) the user's ability to identify themselves as being within the environment.

A user interface which produces unpredictable results, has sluggish response, is unnecessarily complex, or fails to provide appropriate levels of control will likely frustrate the player because the user interface will require a level of attention that detracts from game-play. In other words, there is too much cognitive load (Sweller, 1988) involved in understanding the interface rather than devoted to problem-solving the content of the game. If the user interface cannot be quickly learned, the player will be less immersed in the game and thus less engaged in the play of the game itself.

The fidelity and presentation of the game's graphics and audio can also affect the level of immersion and engagement one experiences while playing. If the player finds the graphics or audio difficult to understand, this will require greater attention for processing the meaning of the graphics or audio. Audio that is jittery, skips, or has poor sound quality, for example, is likely to distract the player.

The second aspect of immersion, strengthening or weakening the sense of immersion created by the interface and fidelity, is that players mentally project themselves into the game environment and accept the game world's rules as real. In a first-person shooter like *Quake*, this means that players think of themselves as the character holding the gun and the maze-like world that the character walks through

as the world that they are in. For character-driven, narrative-based games, such as Square Soft's *Final Fantasy* series or *Xenogears*, identification means personally identifying with the lead character, with supporting characters, and with their surroundings.

Immersion can be measured indirectly through observation and interviewing focused on how the player experiences the game's physical and local interface, the player's reactions to the game's fidelity and presentation, and the player's level of accepting the game world as real and projecting him or herself into the game. Difficulties in using the physical interface, misunderstanding of the logical interface, repeated errors, and expressions of surprise or frustration after an unpredicted outcome regarding the interface can be observed during game-play. Acceptance of the game world as real can sometimes be observed, for example, when a player physically ducks or crouches their own body at the same time as their avatar. Other aspects such as enjoyment of the audio and video can be explored after a game-play session.

(Not) Measuring Engagement

Within a usability lab, as a pilot study, a handful of participants were recruited to play a pair of games, complete pre-play and post-play questionnaires, and answer interview questions. We selected *The Curse of Monkey Island* and *The Oregon Trail 5th Edition* because they are similar in genre and fidelity yet one was a commercial game while the other was an educational title. We wanted to test our model for engagement against common wisdom that entertainment titles were inherently better than educational ones. The games were played back-to-back during two 45-minute play sessions, but first participants completed a shorter variant of the Immersive Tendencies Questionnaire (ITQ) based on Witmer and Singer (1998) and modified slightly for gaming (see Table 1 for sample questins). Selection of the first game to be played was randomly assigned. The player was told to play the game as if they were playing at home. They were told to think aloud whenever they wished, but that it was not required. At fifteen-minute intervals, play was briefly stopped to administer a short questionnaire addressing the level of flow participants were experiencing. After playing for 45 minutes, play was stopped and the participant given a Gaming Engagement Questionnaire (GEQ), again, based on Witmer and Singer's (1998) Presence Questionnaire (see Table 2 for sample questions). Follow-up interview questions were asked. These included open-ended questions about their game-play experience, such as what they enjoyed most and least about the game and how the sound and graphics affected their experience. Directed questions that clarified observations were also included in the interview as needed. These steps, minus the initial ITQ, were repeated for the second game.[1]

Our basic hypothesis was that *Curse*'s metascore (89.9%) from gamerankings.com would correlate to our measured level of engagement and that *Curse*'s score would be higher than that of *The Oregon Trail*. Yet this did not happen as expected with our initial participants.

Table 1: Immersive Tendencies Questionnaire sample questions

14. Do you ever become so involved in doing something that you lose all track of time?						
NEVER			OCCASIONALLY			OFTEN
1	2	3	4	5	6	7
15. Do you easily become deeply involved in computer games or video games?						
NEVER			OCCASIONALLY			OFTEN
1	2	3	4	5	6	7

Table 2: Gaming Engagement Questionnaire sample questions

14. Were you involved in the game to the extent that you lost track of time? NOT AT ALL SOMEWHAT COMPLETELY 1 2 3 4 5 6 7
15. How much did you feel like you were inside the game world? NOT AT ALL SOMEWHAT COMPLETELY 1 2 3 4 5 6 7

Failure and Reflection

Instead, the metascore for *Curse* did not match up with our test results. In fact, our participants were more "engaged" with the educational software, *The Oregon Trail*, than with the highly lauded adventure game! This shocked one researcher's sensibilities, having grown up on the point-and-click adventure gaming genre that *The Curse of Monkey Island* claims as its pedigree. This made clear that common wisdom could be wrong and that basing comparisons on its untested assumptions could lead to failure.

One reason for this mismatch was that the introductory puzzle in *The Curse of Monkey Island* could pose an immediate space for frustration as it included a genre cliché of "pixel hunting," where players had to move their mouse around the game screen, hoping to find a particular object or area of the screen that could be interacted with. Players with little or no familiarity with the genre did not know that they needed to move the mouse around to find hotspots; in fact, it seemed like they did not initially know that the mouse cursor would change when it was over hotspots and certainly didn't know that right-clicking the mouse would bring up an inventory and holding down left-mouse button would bring up a context-sensitive menu. In other words, the tests did in fact measure our participants' frustration with the game interface and thus gave us an accurate measure of lack of engagement, but, again, this was counter to what we had expected based on the metascore for *Curse*. It's possible that familiarity with a genre is needed for players to be fully immersed with later-generation iteration of that genre. If this is true, however, how does it affect ideas on how to accurately measure engagement? Perhaps our measurement instruments would be appropriate for certain purposes but not others. Usability testing or testing for the purposes of minimizing player frustration so player learning would increase, for examples, may still find our testing methods useful.

Measuring engagement while recognizing gaming practice as part of a larger cultural ecology, however, would require different or supplemental testing methods. Games as memorable experiences often require hours upon hours of play time. This is due in part to their fundamental nature as exploration machines where players must perform a series of actions, navigating a path within a rule-based system with its own signs and signifiers and internally consistent meanings. Some consequences or results--and therefore opportunities for meaning-making--can be predicted; others are unexpected. All are emergent out of the complex interaction between game and players, and these experiences are made more meaningful when a player is rooted in the cultural-historical community around the games' genre. Recognizing this, game reviews are typically written by professionals who are immersed in gaming culture after many hours of play. Some reviewers will only review a game after completing it, either "winning" it or otherwise reaching some sort of final conclusion in the game's designed story. By contrast, our lab tests lasted 45 minutes. It's possible--very probable--that for the games we were testing as with most games, 45 minutes is not enough time for players to get a good sense of the underlying rule system of the game. It's not enough time to move out of disequilibrium and into pattern recognition, and Koster (2005) argues that fun in video games comes from the player's ability to recognize patterns to exploit.

Further Research

During the months following our pilot test results, one issue immediately jumped out at us: it seemed clear that something was off about our measurement instruments. While our model may still be useful in helping scholars think about engagement from an immersion and interface perspective, other testing methods needed to be added to adequately account for its social and affective components. Further pushing this idea was the fact that we never confirmed that metascores have any correlation to engagement. Perhaps metascores reflected reviewers' greater sense of gaming culture and history and lasting impressions of game experiences, where engagement (as we modeled and measured it) over emphasized interface and immersion. It would seem, then, that a stronger, broader battery of ways to look at engagement should be devised and tested, especially methods that could account for the situatedness of gaming experiences. This thought helped push one of the researchers into focusing on ethnographic methods for dissertation research (Chen, 2009).

Meanwhile, our model and testing methods did seem useful for others who successfully took them and modified them to better suit their needs. These include some fantastic work on posture, movement, and embodiment and games from Bianchi-Berthouze and team who use a modified version of our concept-map model and questionnaires (cf. Bianchi-Berthouze et al., 2006; Bianchi-Berthouze, Kim, & Patel, 2007; Lindley, Le Couteur, & Bianchi-Berthouze, 2008; Mueller & Bianchi-Berthouze, 2010). Furthermore, there has been some good concurrent research on engagement and player experience in the last few years that an updated version of our model would need to consider. These include a closer look at player experience and immersion (Ermi & Mäyrä, 2005), an exhaustive synthesis of presence literature (Beck et al., 2011), and even a different research group's independent modification of Witmer and Singer's Presence Questionnaire (Brockmyer et al., 2009).

End Notes

(1) All of our test instruments can be downloaded from http://markdangerchen.net/pubs/engagement.tools.zip

References

Alessi, S., & Trollip, S. (2001) *Multimedia for learning: Methods and development 3rd Edition.* Allyn & Bacon.

Bianchi-Berthouze, N., Cairns, P., Cox, A., Jennett, C., & Kim, W.W. (2006). On posture as a modality for expressing and recognizing emotions. Emotion and HCI workshop at BCS HCI London, September, 2006. Retrieved from http://www.uclic.ucl.ac.uk/people/c.jennett/BerthouzeHCI06.pdf

Bianchi-Berthouze, N., Kim, W.W., & Patel, D. (2007). Does body movement engage you more in digital game play? and why? In A. Paiva, R. Prada, & R.W. Picard (Eds.), *Proceedings of the 2nd Annual international conference on Affective Computing and Intelligent Interaction (ACII) 2007 vol. 4738* (pp. 102-113). doi:10.1007/978-3-540-74889-2_10

Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E, Burkhart, K.M., & Pidruzny, J.N. (2009). The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624-634. doi:10.1016/j.jesp.2009.02.016

Bystrom, K., Barfield, W., & Hendrix, C. (1999). A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 8*(2), 241-244. doi:10.1162/105474699566107

Crawford, C. (1982). *The art of computer game design.* MacGraw-Hill Osborne Media.

Chen, M. (2009). Communication, coordination, and camaraderie in *World of Warcraft. Games and Culture, 4*(1), 47-73. doi:10.1177/1555412008325478

Chen, M., Kolko, B., Cuddihy, E., & Medina, E. (June 2005). Modeling and measuring engagement in computer games. Presentation at the annual conference for the Digital Games Research Association (DiGRA), Vancouver, Canada.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. HarperCollins Publisher.

Ermi, L., & Mäyrä, F. (2005). Fundamental components of the gameplay experience: Analysing immersion. *Proceedings of DiGRA 2005 Conference: Changing Views – Worlds in Play.* Retrieved from http://www.digra.org/dl/db/06276.41516.pdf

Gee, J.P. (2003). *What video games have to teach us about learning and literacy.* Palgrave Macmillan.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments, 1*(2), 262-271.

Heeter, C., Chu, K., Maniar, A., Mishra, P., & Egidio, R. (2003). Comparing 14 forms of fun (and learning and gender issues) in commercial versus educational space exploration digital games. *Proceedings of the International Conference on Digital Games Research.*

Jones, M. G. (1998). Creating electronic learning environments: Games, flow, and the user interface. *Proceedings of Selected Research and Development Presentations at the National Convention of the Association for Educational Communications and Technology (AECT)* (pp. 205-214). Retrieved from http://eric.ed.gov/ERICWebPortal/detail?accno=ED423842

Keller, J.M., & Suzuki, K. (1988). Use of the ARCS motivation model in courseware design. In D.H. Jonassen (Ed.), *Instructional designs for microcumputer courseware* (pp. 401-434). Lawrence Erlbaum.

Koster, R. (2005). *A theory of fun for game design.* Paraglyph Press.

Lazzaro, N. (2004). Why we play games: Four keys to more emotion in player experiences. *Proceedings from the Game Developers Conference*. Retrieved from http://www.gamasutra.com/gdc2004/features/20040326/postcard-kane_04.shtml

Lindley, S., Le Couteur, J., & Bianchi-Berthouze, N. (2008). Stirring up experience through movement in game play: Effects on engagement and social behaviour. *Proceedings of the 2008 SIGCHI conference on Human Factors in computing systems (CHI 2008)* (pp. 511-514). doi:10.1145/1357054.1357136

Malone, T.W. (1980). *What makes things fun to learn? A study of intrinsically motivating computer games* (Xerox Palo Alto Research Center Technical Report No. CIS-7). Retrieved from http://cci.mit.edu/malone/tm%20study%20144.html

Malone, T.W. (1982). Heuristics for designing enjoyable user interfaces: Lessons from computer games. *Proceedings of the 1982 conference on Human Factors in Computing Systems.* doi:10.1145/800049.801756

Malone, T.W., & Lepper, M. (1987). Making learning fun: A taxonomy of intrinsic motivation for learning. In R.E. Snow & M.J. Farr (Eds.), *Aptitude, learning, and instruction, vol. 3: Cognitive and affective process analysis* (pp. 223-253). Lawrence Erlbaum.

Mayes, D.K., & Cotton, J.E. (2001). Measuring engagement in video games: A questionnaire. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 692-696).

Mueller, F., & Bianchi-Berthouze, N. (2010). Evaluating exertion games: Experiences from investigating movement-based games. In R. Bernhaupt (Ed.), *Evaluating user experiences in games* (pp. 187-207). Springer. doi:10.1007/978-1-84882-963-3_11

Prensky, M. (2000). *Digital game-based learning*. McGraw-Hill.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments, 8*(5), 560-565. doi:10.1162/105474699566477

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285. doi:10.1016/0364-0213(88)90023-7

Witmer, B. & Singer, M. (1998). Measuring presence in virtual environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225-240. doi:10.1162/105474698565686

Zeltzer, D. (1992). Autonomy, interaction, and presence. *Presence: Teleoperators and Virtual Environments, 1*(1), 127-132.