

This document is downloaded from DR-NTU, Nanyang Technological University Library, Singapore.

Title	Evolutionary nature of virtual experience
Author(s)	Lee, Kwan Min; Jung, Younbo
Citation	Lee, K. M., & Jung, Y. (2005). Evolutionary nature of virtual experience. 2005 Annual Conference - Communication: Questioning the Dialogue.
Date	2005
URL	http://hdl.handle.net/10220/19369
Rights	© 2005 The Authors. This is the author created version of a work that has been peer reviewed and accepted for publication by 2005 Annual Conference - Communication: Questioning the Dialogue, International Communication Association. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [http://citation.allacademic.com/meta/p_mla_apa_research_citation/0/1/2/8/8/p12884_index.html?phpsessid=f2dcef9ef59f22d83deb1b55a36e7b11].

Running Head: VIRTUAL EXPERIENCE

EVOLUTIONARY NATURE OF VIRTUAL EXPERIENCE

KWAN MIN LEE¹, YOUNBO JUNG

University of Southern California

Correspondence should be directed to Kwan Min Lee at Annenberg School for Communication, University of Southern California, 3502 Watt Way, Los Angeles, CA 90089-0281, USA; kwanminl@usc.edu

Abstract. Despite increasing dominance of virtual experience in our everyday life, there have been relatively few discussions on human responses to virtual experience. In the current paper, we define what virtual experience is based on the types of objects being experienced and the ways of experiencing. Then, we review three lines of research traditions examining the evolutionary nature of human responses to virtual experience—a) studies on Media Equation (ME) which investigate human responses to physical and/or social features of virtual objects; b) studies on Doubly Disembodied Language (DDL) which examine human responses to synthesized speech; and c) studies on human preferences for specific content types in media. Based on the general review of the three research traditions, we propose that people's natural and social responses to virtual objects and contents come from the cognitive hijacking of the general thinking mechanism by specialized mind modules.

Keywords: Virtual Experience, Mediated Experience, Domain Specificity, Evolutionary Psychology, Cognitive Hijacking, Media Equation, Doubly Disembodied Language, Content Choice

EVOLUTIONARY NATURE OF VIRTUAL EXPERIENCE

According to LEE (2004a), modern humans have three types of experiences—real experience, hallucination, and virtual experience. This typology is based on the ways of experiencing (sensory vs. non-sensory) and the ontological nature of experienced objects (actual vs. virtual vs. imaginary). Sensory experience is the perceptual experience mediated by human sensors. Actual, mediated, or simulated objects can be perceptually experienced. Non-sensory experience is the non-perceptual experience of imaginary objects and is usually the result of the human imagination (see LEE, 2004a). Sometimes, however, even media can cause non-sensory experience. For example, a good book can bring a rich non-sensory experience to its reader (BRAUN & CUPCHIK, 2001) which is sometimes as vivid as high fidelity audio-visual stimuli.

Actual objects are objects that can be experienced by human sensory systems without using technology. By using the term “actual objects” instead of “real objects,” LEE (2004a) purposefully avoids the Cartesian dualism which assumes the independence of the subjective mental world from the objective physical world (for criticisms against the dualistic assumption in the domain of human experience, see DREYFUS, 1992; HEIDEGGER, 1927/1996; WINOGRAD, 1995; WINOGRAD & FLORES, 1986). Unlike the term “real objects,” which assumes the *existence* of objects independent of human mentality, the term “actual objects” concerns only the *possibility* of objects being experienced without the use of any technology. According to LEE (2004a), real experience is the sensory experience of actual objects.

Virtual objects are objects that are either a) experienced through or b) simulated by the use of technology. Accordingly, there are two types of virtual objects—para-authentic objects vs. artificial objects. Para-authentic objects are the mediated version of actual objects. When an actual object is experienced through mediation technologies such as media and telecommunication devices, the mediated object has some authentic connections to the actual object. For this reason, we use the term “para-authentic” for these types of objects. Artificial objects are objects that are created or simulated by technology. Unlike para-authentic objects, artificial objects have no actual real-life counterparts. For example, a software agent in a massive multiplayer online game does not have any authentic connection to an actual human no matter how well it simulates actual human behaviors, whereas even a rudimentary representation of an actual player of the game (i.e., avatars) has an ontologically authentic connection to the player.

To sum up, para-authenticity is related to mediation technologies and artificiality is related to simulation (e.g., AI) technologies. Virtual experience is the experience of either para-authentic or artificial objects (see LEE, 2004a for more detailed discussions).

Imaginary objects are objects that are experienced in pure hallucination (e.g., daydreaming). They are not actual because they are not experienced by human sensory systems. They are not virtual because neither mediation nor simulation technologies are involved in experiencing or creating them. Imaginary objects are simply the results of human imagination. Hallucination is the non-sensory experience of imaginary objects (see LEE, 2004a for more detailed discussions).

Human experience before the Information Ages was predominantly comprised of “real experience.” With the development of media and simulation technologies, “virtual experience” is becoming a dominant part of human experience. As technologies are being embedded more and more into human sensory systems on a 24-hour basis (e.g., brain chips, wearable computers, etc.), “real experience” might eventually disappear (cf. MCLUHAN, 1964) to the point that the distinction made here will seem less applicable to the study of human experience in future. With regard to the portion of “pure hallucination” in human experience, it can be argued that this portion has been fairly stable and independent of technological developments.

Despite increasing dominance of virtual experience in human experience, there have been relatively few discussions of human responses to virtual experience. Even though we have vast amounts of knowledge on how people process and respond to mediated and/or artificial stimuli, there have been few attempts to provide a general explanation of human responses to virtual experience. In the current paper, we explain the evolutionary nature of virtual experience. More specifically, we review three lines of research traditions examining human responses to virtual experience—a) studies on Media Equation (ME) which investigate human responses to physical and/or social features of virtual objects; b) studies on Doubly Disembodied Language (DDL) which examine human responses to synthesized speech; and c) studies on human preferences for specific content types in media.

MEDIA EQUATION: EVOLUTIONARY NATURE OF HUMAN RESPONSES TO PHYSICAL AND SOCIAL FEATURES OF VIRTUAL OBJECTS

The term “media equation” simply means “media equal real life” (REEVES & NASS, 1996, p. 5). In detail, it implies that an “individual's interactions with computers, television, and new media are fundamentally social and natural, just like any interaction in real life” (p. 5). The main reason is that human brains evolved in a world in which all perceived objects were real physical objects and only humans possessed human-like shapes and human-like characteristics such as language, rapid interaction, emotion, personality, and so on. Therefore, to human minds, anything that seems to be real is “real” and any object that seems to possess human characteristics such as language is a “real human”. When people use media such as television and computers, they usually do not overcome the evolutionary limitation of accepting everything on media at face value. In other words, when people respond to virtual objects, domain-specific mind modules—which evolved around 100,000 years ago (about the time when *Homo sapiens* were first seen in South Africa and the Near East; see MITHEN, 1996, p. 22)—cognitively hijack people’s logical understanding of the ontological limitations of the virtual objects. As a result, people tend to respond to virtual objects as if they were actual.

We believe that human mind is expected to include a number of functionally distinct domain-specific modules, because domain-specific systems are more efficient and effective than a domain-general system (CHOMSKY, 1980; COSMIDES & TOOBY, 1992, 1994; SHERRY & SCHACTER, 1987; see LEE, 2004b for more detailed discussion on domain-specificity and human responses to media). By modules, we mean localized brain areas (or elements) which govern specific human responses to special stimuli or situations (e.g., amygdala for emotional responses [see GOLEMAN, 1995], theory of mind module for social interaction [see BARON-COHEN, 1996], shared-attention mechanism module for triadic representation of self, others, and objects [see BARON-COHEN, 1996]). With few exceptions, the specific locations of numerous domain-specific modules are yet to be discovered. Our cognitive hijacking argument is similar to GOLEMAN’s (1995) theory on amygdala hijacking. GOLEMAN (1995) argues that amygdala temporarily intercepts the normal flow of sensory signals from thalamus to cortex and provides initial conditioned responses to potential threats. In a similar vein, we believe that various mind

modules evolved to provide conditioned responses to certain types of actual objects and information intercept and govern technology users' initial responses to virtual objects and information.

Almost 15 years of cumulative works on Media Equation (ME) by NASS, REEVES and their colleagues (see NASS & MOON, 2000; REEVES & NASS, 1996) at the Social Responses to Communication Technologies (SRCT) Lab of Stanford University have successfully demonstrated how old-time domain-specific mind modules, which govern natural human responses to actual objects, influence our responses to physical and social features of virtual objects. We will summarize their main research results and explain why those results were obtained. Throughout the explanation, our goal is to provide evidence that *old* mind modules are in play when we are using *new* media and technologies (COSMIDES & TOOBY, 1994).

Before turning to NASS, REEVES and their colleagues' research results, it should be mentioned that there is a critical difference between the evidence from studies of child development and the evidence from the media studies conducted at the SRCT lab. The logic of the studies about child development, at least from the evolutionary psychology viewpoint, is that since it is unlikely that very young infants have enough experiences to develop their knowledge of physical and social causation, their immediate understanding of physical and social world should come from innate mind modules. In contrast, the logic suggested by the Media Equation scholars (e.g., NASS, REEVES, and their colleagues) is that since every adult knows that virtual experiences are not real, their natural and social responses to virtual objects come from the *cognitive hijacking* of the general thinking mechanism by domain-specific mind modules.

The ME phenomena that have been reported so far can be categorized into two themes— a) human responses to physical features of virtual objects, and b) human responses to social characteristics of virtual objects. Below, we explain them one by one.²

HUMAN RESPONSES TO PHYSICAL FEATURES OF VIRTUAL OBJECTS

In this section, we review human responses to the following physical features of virtual objects—size, visual fidelity, and motion.

With regard to size, humans are likely to have evolved to pay more attention to big objects (REEVES & NASS, 1996) because big objects more readily become challenges to, or opportunities for, survival. As a result of the huge amount of attention paid, big objects are more easily recalled. There is also an interesting finding that people have a general tendency to like big things and even infants like larger objects over smaller ones (FANTZ, FAGAN, & MIRANDA, 1975). In U.S. presidential elections, taller candidates have continuously won the elections, with the exception of Jimmy Carter. Studies by Reeves and his colleagues show that bigger virtual objects yield more arousal, better memory (descriptive, not image recognition), and more positive social responses (e.g., social attraction, credibility) than smaller ones even when the content is identical (see REEVES & NASS, 1996, Chap. 17 for review). These results indicate that people continue to use the same mind module of judging actual objects based on size, even when they see virtual objects.

In contrast to the size of objects, the visual fidelity of objects does not have a great survival consequence. For survival, humans need to rapidly decide their next actions—fight or flight; stay or run—based on even a slight visual outline of an actual object. Therefore, humans have evolved to see objects and the outside environment mainly through the peripheral vision field, rather than the foveal vision field (HOCHBERG, 1986). Consequently, the visual fidelity of virtual objects is not important *psychologically*—although it might be important aesthetically. Studies by Reeves, Detenber, and Steuer (1993) show that the manipulation of visual fidelity of a scene does not bring any significant differences in arousal, attitudes, and memory. These results indicate that virtual and actual objects are visually processed by the same mechanism.

In a similar way that even a slight outline of an object demands a rapid decision on the next movement—stay or run—to humans, even a slight motion of an object alerts humans to pay close attention to the moving object. Motion has huge survival consequences: moving objects are more likely to be significant threats to survival (see REEVES, THORSON, ROTHSCHILD, McDONALD, GOLDSTEIN, & HIRSCH, 1985 for a review of the visual orienting response). Studies by REEVES et al. (1985) show that people pay more attention to moving objects on the screen, even when they clearly know that the moving virtual objects provide no real actual threats (see REEVES & NASS, 1996, chap. 20 also). Their results indicate that human beings' initial responses to moving virtual objects are governed by an evolutionary hardwired mind module, rather than the general thinking mechanism.

Human responses to social characteristics of virtual objects

In this section, we will review human responses to the following social characteristics of virtual objects—reciprocity, in-groupness, personality and gender.

Reciprocity is one of the major characteristics of being human (CIALDINI, 1993). It is almost culturally universal; across different cultures, reciprocity has been the major theme of education and folk tales. We are trained to help others who have previously provided help to us (CIALDINI, 1993; FOGG, 1997). The most important thing for preserving reciprocity is to detect cheaters. Without cheater detection, reciprocity can no longer be preserved, because cheaters do not reciprocate even though help has previously been given. As a result, humans have evolved to have a specialized cheater detection mind module that is very sensitive to the reciprocity rule in social life (COSMIDES & TOOBY, 1992). A study by FOGG and NASS (1997) shows that humans reciprocate, even to computers. In their experiment, participants calibrated the resolution of a computer screen more when it previously had helped them than when it had not provided any help. A study by MOON (1998) shows that participants even disclosed more about themselves when a computer had disclosed about itself first. These results indicate that, when interacting with virtual social objects such as computers, humans reciprocate before thinking about the ontology of the reciprocated virtual objects.

Humans are very sensitive about genetic relevance of other people, because the survival of a genetically relevant other increases the possibility of one's own gene pool (see DAWKINS, 1989; HAMILTON, 1964). As a result, humans have evolved to have special mind modules for diagnosing genetic in-groupness (see SPEARS, 1989) and this sensitivity generalizes to other types of in-group characteristics (e.g., team membership, community membership, dialect, nationality, ethnicity and so on). NASS, FOGG, and MOON (1996) show that participants evaluated a computer more positively when it was labeled as a teammate than when it was not (see REEVES & NASS, 1996, chap. 13). NASS, ISBISTER, and LEE (2000) also demonstrated that even the ethnicity of software agents matters in the same way as the ethnicity of actual humans (see WALLACE, 1997 for the judgment of other people based on their ethnicity). Specifically, participants evaluated agents with seemingly similar ethnicity to be more attractive, trustworthy, intelligent, and persuasive. The above results indicate that in-groupness of virtual objects is

judged first before cognitive thought of the ontological impossibility of virtual objects' having an ethnicity.

Personality is a trait applicable only to humans (FUNDER, 1997). Humans have evolved to categorize other people according to relatively simple personality types, such as extroversion, dominance, and so on (DIGMAN, 1990; JOHN, 1990; PERSSON, LAAKSOLATHI, & LÖNNQVIST, 2000). This simple categorization scheme significantly reduces cognitive load when humans engage in social interaction, and enables humans to predict other people's behaviors fairly accurately. Rapid understanding and prediction of other people's behaviors is very important for human survival, because, for humans, other humans are either the most dangerous enemies or desirable mating partners. Studies by NASS and his colleagues (ISBISTER & NASS, 2000; MOON & NASS, 1996) demonstrate that people keep using their personality-detection mind modules, even when they interact with virtual objects. In both studies (ISBISTER & NASS, 2000; MOON & NASS, 1996), participants attributed personality (extroversion, introversion, submissiveness, and dominance) to computers and software agents. More importantly, they also unknowingly applied personality-based social rules, such as attraction to similarity, to virtual objects. That is, they evaluated a computer or a software agent more positively when it manifested a personality similar to their own than when it displayed the opposite. These results imply that traits of virtual objects are automatically judged, despite the ontological impossibility of the virtual objects' having a personality.

One of the most important social categories in evolution is gender (BEM, 1981). Correct identification of gender is critical for the successful reproduction of one's own gene. Consequently, humans have evolved to identify the gender of other people almost immediately with or without notice. Gender is used to judge other people and their behaviors—gender stereotyping. A study by NASS, MOON, and GREEN (1997) showed that people apply gender stereotyping even when they interact with computers. One of their results indicated that participants evaluated a male-voiced computer as being more informative about technology than a female-voiced one, whereas the female-voiced computer was evaluated as being more informative about love and relationships compared to the male-voiced one. This is surprising because most participants acknowledged that both male and female voiced computers were programmed by the same person. Again, this result implies that people apply social stereotypes

to virtual social objects, even when they clearly know that it is ridiculous to apply those stereotypes to non-human objects.

DOUBLY DISEMBODIED LANGUAGE (DDL): EVOLUTIONARY NATURE OF HUMAN RESPONSES TO SYNTHESIZED SPEECH

The ability to perceive and produce speech is one of the most important characteristics of humans. Humans have an innate ability for speech perception. For example, even a fetus in the womb can distinguish its mother's voice from other voices (SLOBIN, 1979). The mother's voice is preferred over those of strangers within minutes after birth (DECASPER & FIFER, 1980). Newborns have an innate ability to perceive a number of phonemic distinctions (CARROLL, 1999, p. 257). One-day-old infants show more activity on the left-side of their brains in response to human speech (SLOBIN, 1979). By four days of life, infants can distinguish between utterances in their native language and those in foreign languages (MEHLER, JUSCZYK, LAMBERTZ, HALSTED, BERTONCINI, & AMIEL-TISON, 1988).

The ability to produce speech is also innate or at least is learned very rapidly after birth. By the end of the second month, babies begin to utter coos which are acoustically more varied than simple cries. Reduplicated babbling, the repetition of a consonant-vowel sequence (e.g., *babababa*), begins by about six to seven months. Variegated babbling, murmuring of syllable strings consisting of varying consonants and vowels (e.g., *bigodabu*), appears by 11 to 12 months. By the end of the first year, babies become capable of using speech sounds to communicate meaning (see CARROLL, 1999, chap. 10 for a detailed explanation of early language acquisition).

Speech is not merely a tool for communicating language (linguistic cues); it also conveys a very rich set of socially relevant information through voices (paralinguistic cues). Embedded in voice is fundamental social information about a speaker, such as individual identity, gender, personality, ethnicity, locality, and emotion (see PITTAM, 1994, chap. 4 for a detailed explanation of the vocal communication of social identity, and see chap. 5 for the vocal communication of emotion). Listeners constantly and automatically extract relevant social information about speakers whenever involved in embodied and/or disembodied language situations (see NASS & GONG, 2000; NASS & LEE, 2001; NASS & STEUER, 1993). To borrow NASS and BRAVE's (in press) phrase, human brains have evolved to become "*voice-activated*."

When a computer synthesizes a voice and produces disembodied speech from a given text, the speech becomes doubly disembodied language (see LEE & NASS, 2004 for a detailed discussion on the concept of doubly disembodied language). Computer-synthesized speech is *doubly* disembodied because of the absence of an actual speaker at the moment of its interpretation (*first-degree disembodiment*) and the broken association between its paralinguistic cues and its source, the speaker (*second-degree disembodiment*). Ontologically problematic computer-synthesized speech—“a voice from a non-human source”—provides a serious question regarding human responses toward it. Will evolutionary-hardwired *voice-activation* occur even when people respond to *doubly disembodied language*? A series of studies by LEE and his colleagues (LEE, 2002; LEE & NASS, 2002, 2004, 2005; MORISHIMA, NASS, BENNETT, & LEE, 2001; NASS & LEE, 2001) show that people automatically extract social information even from doubly disembodied language. In other words, even an absolutely machine-like voice³ automatically activates a specialized mind module made for processing actual human voices. The automatic activation of the specialized mind module, in turn, results in social responses to doubly disembodied language.

Below, we explain previous studies on human responses to doubly disembodied language based on three categories—voice identification; gender; and personality inference.

Voice identification

One of the most fundamental social responses to human voice is to identify and assign one human actor to one human voice. As we can see from the cocktail party effect (CHERRY, 1953), humans are very sensitive about the identity of a voice and can pay attention to a particular voice even in a very noisy environment. Throughout the evolutionary history, it is only humans that can make human voice sounds. Therefore, it is always safe to assume the existence of another human when people hear a voice. One voice is always one person and multiple voices indicate the existence of multiple people. Humans are evolutionarily hardwired to believe the existence of another human when they hear a voice (see LEE, 2002).

LEE and NASS (2004) examined whether this fundamental social response will happen even when people hear synthetic voices. They tested this in an e-commerce situation where participants heard product endorsements by multiple people. In one condition, the participants heard endorsements narrated by multiple synthetic voices, whereas in the other condition the

participants heard the same multiple endorsements narrated by a single synthetic voice. They hypothesized that if people were hardwired to identify and assign one human actor to one human voice, they would show the same behavior when they hear a synthetic voice. That is, people will identify and assign one social actor to one synthetic voice. In addition, they also hypothesized that product endorsements narrated by multiple synthetic voices would have higher persuasive impacts than the same endorsements that were narrated by a single synthetic voice because multiple product endorsers have more persuasive power than a single one. The results confirmed both hypotheses. People felt the presence of more social actors when they heard multiple synthetic voices. Moreover, they evaluated a product more positively if the product endorsements were narrated by multiple synthetic voices rather than a single one. To sum up, this experiment clearly shows that humans are compelled to identify and assign a social identity even when they hear a clearly non-human voice.

Gender

Along with the judgment of the existence of someone based on a voice, social judgment of the gender of a voice is almost automatic. In human-human interaction, gender influences a variety of dimensions of persuasion. For example, male persuaders elicit more conformity than female persuaders, and male communicators are considered to be more competent and to occupy a higher social status than female communicators (LEE, NASS, & BRAVE, 2000). The gender of a spokesperson influences a product image. Communication effectiveness is greater when the gender of the spokesperson matches the product than when they mismatch (DEBEVEC & IYER, 1986).

MORISHIMA et al. (2001) tested whether the above mentioned gender stereotyping applies when people hear synthetic voices in an Amazon.com-like e-commerce web site. The results showed that participants trusted a product description more if the gender of the synthetic voice matches the gender of the product. In addition, the gender of voice affected the participants' perceptions of the masculinity/femininity of the product. A product that was advertised using female voices was perceived as more feminine, whereas a product with masculine voice was perceived as more masculine. Conversely, female voices were perceived as less feminine when describing a male product and male voices were perceived as less masculine when describing a female product. These results suggest that gender stereotyping is an almost automatic social

behavior that happens whenever people hear human-voice like sounds. Even a clearly machine-like voice successfully generates gender stereotyping from its listeners automatically and probably unconsciously.

Personality Inference: Proximate source orientation

People use paralinguistic cues of speech to assess the personality of the speaker (SCHERER, 1978, 1979). Although the correlation between people's voices and their personalities are somewhat weak (see NASS & LEE, 2001), this way of voice-based personality inference provides a quick and useful method for rapidly judging the personalities of other people.

A study by NASS and LEE (2001, see Experiment 1) shows that paralinguistic cues in synthesized speech influence the judgment about the personality of a virtual speaker. When listening to book reviews on a book-buying web site, participants who heard an extroverted machine voice judged the reviewer to be an extrovert, whereas participants who heard an introverted machine voice judged the personality of the reviewer to be introverted. This is a surprising result and should not be confused with the personality inference in an embodied language situation. In an embodied language situation, listeners' interpretation of paralinguistic variations of a narrator's voice is logically sound and natural in that in order to appreciate a story, a joke, or a discourse more, listeners should actively imagine a virtual speaker in a way that is set by the narrator (see CLARK, 1996, chap. 12). In a synthetic voice situation, however, the influence of paralinguistic cues on the mental formation of a virtual speaker is logically problematic since machines ontologically lack consciousness. The results, thus, are very surprising and provide evidence that our responses to synthesized speech do not reflect our understanding of the ontological limitations of synthesized speech. Our response to speech is so evolutionarily hardwired that we react to synthesized speech, either consciously or subconsciously, as if it were natural human speech.

CONTENT PREFERENCE: EVOLUTIONARY NATURE OF HUMAN PREFERENCE FOR SPECIFIC CONTENT IN MEDIA

In general, most studies in the field of mass communication are based on social or environmental learning perspectives on human behaviors (SHERRY, 2004) and focus on the issue

of how specific media content affects human minds and behaviors—i.e., media-effect studies. For example, influential theories in media studies such as the cultivation theory (e.g. GERBNER, GROSS, MORGAN, SIGNORIELLI, & SHANAHAN, 2002), the general aggression model (e.g. ANDERSON & BUSHMAN, 2002), and the media priming theory (e.g. BERKOWITZ, 1990) focus on the issue of how certain types of media contents—usually violent media contents—generate certain thoughts, affects, or behaviors related to the portrayed media contents. Although the question of *how* media influence human minds and behaviors is important, equally important questions are *why* human beings are hardwired to be fascinated with particular contents (or genres) of media even without sufficient social or environmental learning, and *why* human beings continuously pursue these media products despite their potential harmful effects.

The “why” questions are especially important if we consider that human engagement in fictional content in general is a cross-culturally universal, species-typical phenomenon (TOOBY & COSMIDES, 2001). Human engagement in fictional content is a product of adaptation designed by natural selection in order to solve the everyday survival and reproduction problems that our ancestors faced throughout generations (TOOBY & COSMIDES, 2001; cf. PINKER [1997] sees it as a byproduct of adaptation, not a direct product of adaptation). It is an intrinsically rewarding activity, as we can see from the continuing pursuit of engagement in fictional worlds even with the existence of potential harmful side effects (e.g., continuing exposure to violent or sexually explicit media even when these genres of media tend to provide negative social and psychological effects). It is made possible because humans have specialized mind modules (e.g., pretense, see LESLIE, 1987) allowing them to enter and participate in imagined worlds.

Unfortunately, only a few attempts have been made in media studies to provide at least partial explanations for the “why” questions. The existing theories to explain human preferences for particular media content (e.g. selective exposure by BRYANT & ZILLMANN, 1984; uses and gratifications by BLUMLER & KATZ, 1974) are limited in that their explanations are not based on long-term evolutionary perspectives. Their explanations are individual-specific, not species-specific. Therefore, the existing theories cannot explain well this culturally-universal, species-typical phenomenon. In this section, we will provide explanations for the evolutionary nature of human preferences for particular media content. More specifically, we will discuss why humans are especially fascinated with five culturally universal media contents—news, sexually explicit media, violence, horror, and humor.

News

SHOEMAKER (1996) explains that people's attention to news or surveillance is a natural characteristic of all humans. According to her, humans are "hardwired to news." Specifically, she states that humans are hardwired to pay more attention to the threatening types of negative news stories because it has been advantageous for our survival and reproduction to pay attention to potential dangers surrounding us (a surveillance function). Historically, the watchman in a tribe broadcasted information about what was happening in the environment. Nowadays, the media, including the Internet, perform the function of a gatekeeper to filter overwhelming information, thereby taking the place of the watchman (SHOEMAKER, 1996).

It is important to note that SHOEMAKER does not take an extreme position that privileges biological evolution. SHOEMAKER proposes two possible explanations for human-surveillance behavior by embracing a compromise perspective on the interaction between nature and nurture. In addition to evolutionary mechanisms for reproductive advantage, SHOEMAKER indicates that the cultural evolution we learn from one another and inherit from our ancestors through non-genetic means also stimulates us to pay attention to the environment (e.g. social learning theory by BANDURA, 1973).

To sum up, the motivation of human obsession with news can be explained as a hardwired social behavior adapted via natural selection because attending to negative deviance in the environment has been advantageous to human reproduction and survival. The reason for the predominance of negative news in news content can be understood in this vein.

Sexually explicit media

MALAMUTH (1996) explains that the gender difference in the use of sexually explicit media is a result of different evolutionary reproductive demands or a result of different mating strategies. According to him, psychological mechanisms (or modules) governing sexuality for each of the two genders have evolved differently via natural selection because reproductive consequences of sexual behavior for male and female are different. For example, females usually experience restrictions during pregnancy and also invest more time and effort in each offspring. As a result, females' psychological mechanisms have evolved to carefully select a mate who can responsibly contribute himself to raising offspring and taking care of the family. Therefore, it

might be natural for females to prefer sexually explicit media that appeal to a long-term orientation of female mating strategy. Males, however, historically took advantage of short-term mating opportunities because intercourse with a larger number of fertile females leads to a more reproductive success. As a result, males prefer sexually explicit media that contain a short-term orientation of male mating strategy which predominate in the current sexually explicit media products (MALAMUTH, 1996).

Evolutionary perspectives that look at different mating strategies between male and female can explain not only human responses to sexually explicit media but also males' inevitable vulnerability to sexually explicit media that appeal to the short-term mating strategy. Paying attention to sexuality is one of the basic human instincts because sexuality is highly related to successful reproduction. These sexuality modules in the human brain may automatically seduce humans into paying attention to or even enjoying sexually explicit media, because it is functionally advantageous to do so. However, different mating strategies between males and females may result in their different preferences in selecting sexually explicit media.

Violence

PINKER (2002) indicates three examples of the innate-hardwired violence in the human mind and body: a) "the larger size, strength, and upper-body mass of men is a zoological result of violent male-male competition in evolutionary history"; b) "toddlers just past the age of two show the signs of violence by engaging in hitting, biting, and kicking"; and c) "people enjoy watching sports, which are the stylized combats consisting of aiming, chasing, and fighting" (p. 316-317). Thus, PINKER (2002) argues that aggression or violence is an "organized, goal-directed activity" that is derived from the nature of the evolved human mind, rather than a random malfunction (p.316).

In addition, PINKER (2002) explains that violence has been historically justified within the stereotypical dichotomous view of a good guy versus a bad guy. Skillful violence against animals to get food and against nemeses to protect one's own tribe was regarded as heroic masculinity or a sign of being a "real man" (p. 309), and thus was adored by people. Fellow humans who are outside the circle of "we" are often treated like animals that do not deserve our sympathy (PINKER, 2002). For example, movie viewers often get excited to see that protagonists kill antagonists in even the cruelest ways. Violence by a protagonist is accepted well without any

disturbance because people tend to identify with the protagonist—the “our guy, thus a good guy” thinking (OATLEY, 1994).

Violence can be also viewed as a result of the mating competition throughout human history. Due to the difference in the parental investments of males and females, males had to compete over a scarce-reproductive capacity of females (PINKER, 2002). Females preferred males who showed a strong masculinity, which meant better protection, more food, and higher survival rates for the family in the Hunting Age. Thus, males have become the violent gender and, as a consequence, enjoy violent media more than do females in general.

To summarize, it is not so surprising to see human beings’ enjoyment of violent media because the sensitivity to violence has evolved from the Stone Age. The innate modules of violence reside deeply or subliminally inside all human minds. Virtual (vicarious) experience of violence may unlock the button or wake up the innate instinct of violence (PINKER, 2002). Because the human mind still regards violence as an important function or rewarding activity for reproduction and survival, human responses to this wake-up call for violence (exposure to violent media) could be perceived as something enjoyable, exciting, and satisfactory, which are similar to meeting old friends whom people have not seen for a long time.

Horror

MARKS and NESS (1994) argue that phobias are innate fears that never went away. To support this argument, PINKER (1997) states that children between the ages of three and five become fearful of phobic objects, and then master these phobic objects one by one as they grow up. Thus, if a person has not had a chance to master his/her innate phobias, the person is likely to fear the phobic objects in his/her adulthood. A good example would be why city-dwellers fear snakes the most (PINKER, 1997). Having said that, horror movies may have their unique value in training humans to master their fears in a safer way. Similarly, TOOBY and COSMIDES (2001) state that virtual experiences (e.g., reading a fiction) helps us experience certain horrible tasks virtually without facing any real danger.

Another possible explanation for enjoying horror movies is that constant vigilance had been a must-have attitude for a long time and is even required in the 21st century. For example, our ancestors in the Hunting Age needed to be alert constantly in order to increase their survival rates. They were exposed to many dangers in the environment without much protection.

Although we do not need to worry about animal attacks nowadays, these innate modules of vigilance may have been adapted in the human mind and may keep sending us a signal that we need to be alert and prepared. Thus, people may enjoy horror movies because the virtual experience of watching horror movies allows the exposure to some levels of danger and alert without any actual harmful consequences. In other words, the human preference for horror movies can be explained as a result of the innate need for being alert all the time.

To sum up, the first argument is based on the assumption that human responses to a specific media content are functional products of human adaptations. Watching horror movies is a result of a functional desire to master innate phobias via virtual experience without facing real dangers, thus is pleasurable and enjoyable to humans. The second argument is based on the assumption that human responses to media are functionless byproducts of human adaptations. It has been advantageous for humans to be constantly vigilant, which is highly related to our survival. Because of these adapted mind modules, people may have an urge to be alert without knowing precisely why. In other words, if there is no warning or alert, people may not realize that they are alive in the real world. In order to have a feeling of being alive, people may seek scary media contents that provide surprise, shocks, and/or alerts without taking any real harmful consequences.

Humor

Humor has been related to smiles, laughs, and positive moods. Watching humorous media products can enhance people's moods. In fact, ZILLMANN (1988) proposed a theory of mood management in order to explain why individuals select certain messages or media products based on the hedonistic premise. The hedonistic premise posits that individuals naturally strive to terminate bad moods and to maintain or improve good moods, which is one of the innate human instincts. This desire for maintaining good moods (and eventually good mental health) is an important survival weapon. Many medical studies show the evidence for the positive correlation between good mental health and good physical health. Bad mood and stresses cause diseases and cut life expectancies. We can safely conclude that this natural desire to maintain positive moods (or the hedonistic premise module) has caused human beings to consume humorous media contents, especially when they are in a bad mood (see OLIVER, 2003 for a detail discussion about the paradoxes of negative-affect producing entertainment).

CONCLUSION

In the current paper, we reviewed three lines of research traditions examining the evolutionary nature of human responses to virtual experience—a) studies on Media Equation (ME); b) studies on Doubly Disembodied Language (DDL); and c) studies on human preferences for specific media contents. Throughout the review, we proposed that people's natural and social responses to virtual objects and contents come from the *cognitive hijacking* of the general thinking mechanism *by* domain-specific mind modules.

We understand that our cognitive hijacking model is based on *indirect* empirical evidence. Although we have reviewed many empirical studies in the current paper, they by themselves cannot provide any direct empirical support for our claim. In fact, current research methods in social science can never empirically test the cognitive hijacking model of human responses to virtual experience, because domain-specific modules in the brain are not directly observable. There are two ways to overcome these limitations in future research. First, future studies should try to employ new methodologies (e.g., fMRI) which enable direct observation of brain activities. Second, scholars can come up with a clever research paradigm (e.g., COSMIDES and TOOBY's series of studies based on Wason selection tasks [see COSMIDES, 1989]) which makes the modularity argument as the most compelling explanation for given experimental results.

As a final remark, we want to emphasize that we are not championing a pure biological determinism (see CARTWRIGHT, 2000 for more detail discussion about the abuse of evolutionary psychology). We embrace a compromise perspective of “the interaction between nature and nurture” in explaining human responses to virtual objects and contents (SHERRY, 2004). We are focusing on the “nature” part in the current paper, especially because this part has been underestimated and understudied in the field of media studies and communication. Mixed with the predominant nurture-based (or culture-based) explanations on human responses to virtual objects and contents, the current paper will provide a balanced view of the complicated nature of human virtual experience.

References

- ANDERSON, C. A., and BUSHMAN, B. J. (2002): Human aggression. *Annual Review of Psychology*, 53, 27-51.
- BANDURA, A. (1973): *Aggression: A social learning analysis*. Upper Saddle River, NJ: Prentice Hall.
- BARON-COHEN, S. (1996): *Mindblindness: An essay on autism and theory of mind*. Boston, MA: MIT Press
- BEM, S. L. (1981): Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88, 354-364.
- BERKOWITZ, L. (1990): On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *Psychological Bulletin*, 95, 410-427.
- BLUMLER J. G., and KATZ, E. (1974): *The uses of mass communications: Current perspectives on gratifications research*. Beverly Hills, CA: Sage.
- BRAUN, I. and CUPCHIK, G.C. (2001): Phenomenological and quantitative analyses of absorption in literary passages. *Empirical Studies of the Arts*, 19, (1), 85-109.
- BRYANT, J., and ZILLMANN, D. (1984): Using television to alleviate boredom and stress: Selective exposure as a function of induced excitational states. *Journal of Broadcasting*, 28, 1-20.
- CARROLL, D. W. (1999): *Psychology of language* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing Co.
- CARTWRIGHT, J. (2000): *Evolution and human behavior*. Cambridge, MA: MIT.
- CHERRY, E. C. (1953): Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustic Society of America*, 25, 975-979.
- CHOMSKY, N. (1980): *Rules and representations*. New York: Columbia University Press.
- CIALDINI, R. B. (1993): *Influence: Science and practice* (3rd ed.). New York: Harper Collins.
- CLARK, H. H. (1996): *Using language*. New York, NY: Cambridge University Press.
- COSMIDES, L. (1989): The logic of social exchange: Has natural selection shaped how humans reason? Studies with the wason selection task. *Cognition*, 31, 187-276.
- COSMIDES, L., and TOOBY, J. (1992): Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides and J. Tooby (eds): *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford, UK: Oxford University Press, 163-228.

- COSMIDES, L., and TOOBY, J. (1994): Origins of domain specificity: The evolution of functional organization. In L. Hirschfeld and S. Gelman (eds): *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press, 85-116.
- DAWKINS, R. (1989): *The Selfish gene* (2nd ed.). Oxford: Oxford University Press.
- DEBEVEC, K., and IYER, E. (1986): The influence of spokespersons in altering a product's gender image: Implications for advertising effectiveness. *Journal of Advertising*, 15, 12-20.
- DECASPER, A. J., and FIFER, W. P. (1980): Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1174-1176.
- DIGMAN, J. M. (1990): Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
- DREYFUS, H. (1992): *Being-in-the-world*. Cambridge, MA: MIT Press.
- FANTZ, R. L., FAGAN, J. F., and MIRANDA, S. B. (1975): Early visual selectivity as a function of pattern variables, previous exposure, age from birth and conception, and expected cognitive deficit. In L. P. Cohen and P. Salapatek (eds): *Infant perception: From sensation to cognition: Basic visual processes* Vol.1. New York: Academic Press, 249-345.
- FOGG, B. J. (1997): *Charismatic computers: Creating more likable and persuasive interactive technologies by leveraging principles from social psychology*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- FOGG, B. J., and NASS, C. (1997): Do users reciprocate to computers? *Proceedings of the CHI Conference*, Atlanta, GA.
- FUNDER, D. C. (1997): *The personality puzzle*. New York: W. W. Norton.
- GERBNER, G., GROSS, L., MORGAN, M., SIGNORIELLI, N. and SHANAHAN, J. (2002): Growing up with television: Cultivation processes. In J. Bryant and Zillmann, D., *Media Effects: Advances in Theory and Research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, 43-68.
- GOLEMAN, D. (1995): *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam.
- HAMILTON, W. D. (1964): The genetical theory of social behavior: I and II. *Journal of Theoretical Biology*, 7, 1-52.
- HEIDEGGER, M. (1927/1996): *Being and time: A translation of Sein und zeit* (J. Stambaugh, Trans.). Albany, NY: SUNY Press.

- HOCHBERG, J. (1986): Representation of motion and space in video and cinematic displays. In K. R. Boff, L. Kaufman, and J. P. Thomas (eds). *Handbook of perception and human performance* (Vol. 1: Sensory processes and perception). New York: John Wiley and Sons, 22.21-22.64.
- ISBISTER, K., and NASS, C. (2000): Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Interaction*, 53, (1), 251-267.
- JOHN, O. P. (1990): The "big five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66-100). New York: Guilford.
- LEE, E. J., NASS, C., and BRAVE, S. (2000): Can computer-generated speech have gender? An experimental test of gender stereotypes. *CHI 2000 Conference Companion*, Amsterdam, Netherlands.
- LEE, K. M. (2002): *Social responses to synthesized speech: Theory and application*. Unpublished doctoral dissertation. Stanford, CA: Stanford University.
- LEE, K. M. (2004a): Presence, Explicated. *Communication Theory*, 14, (1), 27-50.
- LEE, K. M. (2004b): Why presence occurs: Evolutionary Psychology, Media Equation, and Presence. *Presence: Teleoperators and virtual environments*, 13, 494-505.
- LEE, K. M., and NASS, C. (2002): *User choice and computer-synthesized speech*. Unpublished Document.
- LEE, K. M., and NASS, C. (2004): The multiple source effect and synthesized speech: Doubly disembodied language as a conceptual framework. *Human Communication Research*, 30, 182-207.
- LEE, K. M., and NASS, C. (2005): Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology*, 7, 31-45.
- LESLIE, A. M. (1987): Pretence and representation. The origins of "theory of mind". *Psychological Review*, 94, 412-426.
- MALAMUTH, N. (1996): Sexually explicit media, gender differences, and evolutionary theory. *Journal of Communication*, 46, 8-31.
- MARKS, I. M. and NESSE, R. M. (1994): Fear and fitness: An evolutionary analysis of anxiety disorders. *Ethology and Sociobiology*, 15, 247-261.
- MCLUHAN, M. (1964): *Understanding media: The extensions of man*. New York: McGraw Hill.

- MEHLER, J., JUSCZYK, P. W., LAMBERTZ, G., HALSTED, N., BERTONCINI, J., and AMIEL-TISON, C. (1988): A precursor of language acquisition in young infants. *Cognition*, 29, 144-178.
- MITHEN, S. J. (1996): *The prehistory of the mind: The cognitive origins of art, religion and science*. New York: Cambridge University Press.
- MOON, Y. (1998): *When the computer is the "salesperson": Computer responses to computer "personalities" in interactive marketing situations*. Working Paper No. 99-041, Harvard Business School. Boston, MA.
- MOON, Y., and NASS, C. (1996): How "Real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, 23, (6), 651-674
- MORISHIMA, Y., NASS, C., BENNETT, C., and LEE, K. M. (2001): Effects of "Gender" of Computer-Generated Speech on Credibility. *Technical Report of IEICE TL 2001-16*, 31, (8), 557-562.
- NASS, C., and BRAVE, S. (in press): *Voice activated: How people are wired for speech and how computers will speak with us*. Boston, MA: MIT Press.
- NASS, C., FOGG, B. J., and MOON, Y. (1996): Can computers be teammates? *International Journal of Human-Computer Studies*, 45, (6), 669-678.
- NASS, C., and GONG, L. (2000): Speech interfaces from an evolutionary perspective. *Communications of the ACM*, 43, (9), 36 – 43.
- NASS, C., ISBISTER, K., and LEE, E. J. (2000): Truth is beauty: Researching embodied conversational agents. In J. Cassells (ed.): *Embodied conversational agents*. Cambridge, MA: MIT Press, 374-402).
- NASS, C., and LEE, K. M. (2001): Does computer-generated speech manifest personality?: Experimental test of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology, Applied*, 7, (3), 171-181.
- NASS, C., and MOON, Y. (2000): Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56, (1), 81-103.
- NASS, C., MOON, Y., and GREEN, N. (1997): Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology*, 27, (10), 864-876.
- NASS, C. and STEUER, J. S. (1993): Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research*, 19, (4), 504-527.

- OATLEY, K. (1994): A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23, 53-74.
- OLIVER, M. B. (2003): Mood Management and Selective Exposure. In J. Bryant, D. R. Roskos-Ewoldsen, and J. Cantor (eds): *Communication and emotion: Essays in honor of Dolf Zillmann*. Mahwah, NJ: Lawrence Erlbaum Associates, 131-154.
- PERSSON, P., LAAKSOLATHI, J. and LÖNNQVIST, P. (2000, November): *Stereotyping characters: A way of triggering anthropomorphism?* Paper presented at AAAI Fall symposium: Socially Intelligent Agents—The Human in the Loop, North Falmouth, MA. Retrieved January 11, 2002, from <http://www.sics.se/~jarmo/SocIntAgent.htm>
- PINKER, S. (1997): *How the mind works*. New York, NY: W. W. Norton and Company, Inc.
- PINKER, S. (2002): *The blank slate: The modern denial of human nature*. New York, NY: Penguin Group.
- PITTAM, J. (1994): *Voice in social interaction: An interdisciplinary approach*. Thousand Oaks, CA: Sage.
- REEVES, B., DETENBER, B. and STEUER, J. (1993, May): *New televisions: The effects of big pictures and big sound on viewer responses to the screen*. Paper presented at the Information Systems Division of the International Communication Association, Washington, D.C.
- REEVES, B. and NASS, C. (1996): *The media equation: How people treat computers, television, and new media like real people and places*. New York, NY: Cambridge University Press.
- REEVES, B., THORSON, E., ROTHSCHILD, M. L., MCDONALD, D., GOLDSTEIN, R., and HIRSCH, J. (1985): Attention to television: Intrastimulus effects of movement and scene changes on alpha variation over time. *International Journal of Neuroscience*, 27, 241-255.
- SCHERER, K. R. (1978): Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, 8, 467-487.
- SCHERER, K. R. (1979): Personality markers in speech. In K. R. Scherer and H. Giles (eds): *Social markers in speech*. New York, NY: Cambridge University Press, 147-209.
- SHERRY, D. F. and SCHACTER, D. L. (1987): The evolution of multiple memory systems. *Psychological Review*, 94, 439-454.
- SHERRY, J. (2004): Media effects theory and the nature/nurture debate: A historical overview and directions for future research. *Media Psychology*, 6, (1), 83-109.

- SHOEMAKER, P. (1996): Hardwired for news? Using biological and cultural evolution to explain the surveillance function. *Journal of Communication*, 46, 32-47.
- SLOBIN, D. I. (1979): *Psycholinguistics* (2nd ed.). Glenview, IL: Scott, Foresman and Co.
- SPEARS, S. C. (1989): *Controlling for exposure to intragroup communication: Effects of heterogeneous social perspectives and dependence in intragroup behavior and attitudes*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- TOOBY, J. and COSMIDES, L. (2001): Does beauty build adapted minds? Toward an evolutionary theory of aesthetics, fiction and the arts. *SubStance, Issue 94/95*, 30, (1), 6-27.
- WALLACE, W. L. (1997): *The future of ethnicity, race, and nationality*. Westport, CT: Praeger.
- WINOGRAD, T. (1995): Heidegger and the design of computer systems. In A. Feenberg and A. Hannay (eds): *Technology and the politics of knowledge*. Bloomington, IN: Indiana University Press, 108-127.
- WINOGRAD, T., and FLORES, F. (1986): *Understanding computers and cognition: A new foundation for design*. Norwood, NJ: Ablex Publishing Corporation.
- ZILLMANN, D. (1988): Mood management through communication choices. *American Behavioral Scientist*, 31, (3), 327-340.

Endnote

¹ This work was supported in part by the Nuri grant to the first author, which is jointly funded by the Ministry of Education in S. Korea and Korea Game Industry Development Agency (KGIDA).

² Our explanation of the evolutionary nature of the media equation phenomena is minimal, because it was fully described in our pervious work (Lee, 2004b).

³ No participants in Lee and his colleagues' previous studies (Lee, Nass, & Brave, 2000; Lee, 2002; Lee & Nass, 2002, 2004, in press; Nass & Lee, 2001) misrecognize synthetic voices with actual human voices. Thanks to the technological development, synthetic voices developed by current TTS (Text-to-Speech) engines sound more human-like than the synthetic voices that Lee and his colleagues used.