A CAPACITY LIMITED, COGNITIVE CONSTRUCTIONIST MODEL OF VIRTUAL PRESENCE

David Nunez, MPhil.

Dissertation Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Computer Science

UNIVERSITY OF CAPE TOWN

December, 2007

ACSENT

Department of Psychology
University of Cape Town

Copyright 2007
by
David Nunez

Abstract

The Capacity Limited, Cognitive Constructionist (CLCC) model of presence is proposed as an information processing model of presence, which is demonstrated to have more theoretical power than extant models. The CLCC model defines information processing paths between attention, working memory, declarative memory and procedural memory, which operate to create and maintain a semantic context or bias. Bottom-up information entering the sensory cortices is filtered by attention into working memory where it forms temporary structures encoding the subject's experience of the VE. These structures also receive top-down information, which arises in declarative memory. This interaction of top-down and bottom-up data gives the entire model a semantic bias which attempts to keep the subject's construction of the environment semantically coherent. This allows for inferences and decision making, which translates into high presence. A semantically incoherent construction, or one which does not have enough working memory capacity allocated to it will result in poorer inferences about the environment, and reduced presence. If, as the CLCC model contends, presence involves information processing rather than simple perception, then one would expect to see working memory interference effects and semantic content effects in presence. Six studies were conducted to test these conjectures and validate the CLCC model. Studies 1 – 3 examined the role of working memory and attention on presence (the bottom half of the model), while Studies 4 – 6 examined semantic content and processing effects on presence (the top half of the model).

Study 1 manipulated working memory (WM) load during VE exploration. The CLCC prediction was that WM load would interfere with presence. Data from 177 subjects showed smaller effects than predicted: No WM effects on spatial presence, lower naturalness under spatial WM load, and lower engagement under verbal WM load. This suggests that spatial presence makes no use of WM, and that engagement and naturalness make limited use of it. While engagement seems to make use of semantic processing as predicted, naturalness seems to use spatial processing. Study 2 examined WM use by media decoders by repeating Study 1 with a text-based VE. Data from 114 subjects shows no WM effects exist on any of the four ITC-SOPI factors. This supports Study 1's finding that spatial presence does not use WM, but

contradicts results engagement and naturalness. Study 3 examined the relative contribution of attention and WM. 46 subjects viewed VE walkthroughs in three conditions: Viewing one walkthrough only (baseline), viewing two walkthroughs simultaneously (WM load condition), or viewing one walkthrough and a jumbled video simultaneously (attention load condition). The CLCC model predicted the WM load condition would interfere with presence the most, followed by the attention load condition, followed by the baseline. No difference was found across conditions, although naturalness and engagement predicted task performance, indicating that spatial presence is distinct from these factors, in agreement with the findings of Study 1 and 2.

Study 4 was a survey of semantic and processing effects on presence. Data from 101 computer gamers indicate that it is how often gamers play presence games (and not how many years they have been playing) that predicts how important they consider presence to their gaming experience. This suggests a moderate term activation effect rather than a long term learning effect. Furthermore, gamers with a high thematic inertia rate presence as important to gaming, indicating a processing effect. Finally, gamers who are capable of integrating non-diegetic music into their experiences rate presence as more important, which supports the CLCC notion that information processing of both semantic and perceptual information is important to presence. Study 5 followed up Study 4 by focusing on one specific content area. 461 flight simulation gamers completed the survey. Findings largely agree with those of Study 4, and strongly support the CLCC model prediction that highly specific expectations of content will reduce presence, while generalized expectations will increase it. Thematic inertia and priming were are also positively associated with presence, as predicted by the CLCC model. Study 6 manipulated non-diegetic information (background music) and semantic priming to test semantic processing in presence. The CLCC model predicted that all VE related information (semantic or perceptual) contributes to presence, particularly engagement and naturalness. 181 subjects were primed with materials semantically relevant or irrelevant to VE content, and then experienced the VE with no background music (baseline), music which semantically fit the VE, or VE music which was not a semantic fit. Priming did not influence presence as predicted, but non-diegetic music which fit the VE increased naturalness as predicted. The no-fit music produced the same presence scores as the baseline condition, indicating that it was filtered out by attention, as predicted by the CLCC model.

Overall, the CLCC model and data show that content effects occur in presence, and how these are mediated by declarative memory. It also shows that presence is a complex multi-level processing phenomenon. Spatial presence is at a cognitively low level, relying on perceptual (bottom-up) information, while engagement and naturalness are heavily dependent on conceptual (top-down) information, operating as a set of expectation-content comparisons which, when met by the content, lead to enhanced presence. These high and low cognitive forms of presence are largely independent, but do share some semantic effects, likely due to a reliance on common underlying cognitive processes such as priming and thematic inertia. The top half of the CLCC model (which encodes semantic meaning and explains content effects) is better supported that the bottom half (which predicted interference and attention effects). This finding is highly unexpected, as the literature on almost all extant models predicts an important role for attention in presence. From the data however, one must conclude that spatial presence makes no use of working memory, while cognitive higher forms of presence make use of limited amounts of working memory.

Acknowledgements

This was a large project which was not cheap to run. It would not have been possible without the financial support of these entities:

- The Collaborative African Virtual Environment System (CAVES), which was funded by the South African government's Department of Arts, Culture, Science and Technology.
- The South African National Research Foundation
- The Faculty of Humanities at the University of Cape Town

Many people supported an encouraged me during the project. Most prominent among them is Edwin Blake, with whom I have been researching presence for almost ten years now; he has as always had great faith in my insights, and has always been an enthusiastic supporter of my work. Sally Swartz, Colin Tredoux and Paula Ensor, who took an interest in my budding research career and offered me a research post and a lab to work in. Cara Winterbottom, Katherine Rix, Steven Bowles and Chomsky (my little web server) who assisted me in collecting data and running the seemingly endless line of subjects through the lab. The VEG research group, who patiently listened to and vetted many of the ideas in this dissertation. My labmates at the CVC lab with whom I shared sleepless nights during the dark days of CAVEAT development and testing, as well as many happy movie nights. Finally, Ilda Ladeira, who has to listen to my gripes (and the odd success) day in and day out for little more than scraps of cheese.

This project taught me a lot about research which you don't find in methodology texts. I subsequently found most of what I needed to know in this proverb by Antonio Machado:

Cuatro cosas tiene el hombre
que no sirven en la mar:
ancla, gobernalle y remos,
y miedo de naufragar.

Four things man has
which are of no use upon the sea:
anchor, tiller and oars,
and fear of wrecking his ship.

Table of Contents

Table of Figures

Chapter 1: Introduction

Presence, the sense that one is experiencing a mediated or virtual environment as if it were real and non-mediated, is strongly associated with mediation and technology, but in the end there is always a person, with intelligence, memories, biases and expectations, who takes in the stimuli produced by the environment, and has (or fails to have) a presence experience. This dissertation is about the role of the subject in presence. People are not passive agents in the world – they process what they experience, and they act based on what they know. This dissertation aims to uncover how much of the presence experience is attributable to this processing and knowledge. It is probably best to begin by giving a little history to put this research in context. Given the topic of this dissertation (virtual presence), one might expect this introduction to say something about Ivan Sutherland's "perfect display" from the late 1960s (Sutherland, 1965), or Marvin Minsky's early work on on telepresence in 1970s (Minsky, 1980); the kind of thing one might call a "standard history" of virtual reality and presence. Such histories are interesting, but they do not reflect what this dissertation is about at all, because being histories of science and engineering, they place too much emphasis on presence as an aspect of a technology. Rather, I would like to set a broader context, free of reference to any particular technology, by beginning this historical excursion a little earlier than the 1960s. I would like to begin in the mid 1660s.

During this time, Japanese poets were discovering a new medium, the *haiku* (Ueda, 1996). Much like virtual environments (VEs) which evolved from television and film, haiku evolved from an earlier, well respected and well established poetic form called *tanka* (Ueda, 1996). A *tanka* aims to evoke a moment or emotion chosen by the poet, within the constraints of only thirty-one syllables (Ueda, 1996). *Haiku*, however, was far more ambitious. Being largely defined by minimalist-minded Zen Buddhist monks such as Matsuo Basho (1644 – 1694), all *haiku* aimed to evoke one particular emotion – *sabi* – using only seventeen syllables. Basho himself became quite proficient at evoking *sabi* within the difficult technical constraints of *haiku* – consider these two examples of his work:

Moonlight slanting
through the bamboo grove;
a cuckoo crying.

Another year is gone;
and I still wear
straw hat and straw sandal.

(Ueda, 1982)

What exactly is *sabi*? *Sabi* was difficult to define, even by the *haiku* masters of that era. In modern Japanese, the word has transformed into *sabishii*, and refers to a bittersweet, peaceful loniless (Ikeno & Davies, 2002). For the *haiku* masters and their disciples, *sabi* was a complex emotion, and much discussion was devoted to defining it and understanding its nature. One of the most famous definitions of *sabi* was given by one of Basho's disciples, Kyoraisho. During a discussion group, Kyoraisho is quoted as having said

> *sabi is the colour of haiku;*
> *it is different from tranquility.*
> *For example, if an old man dresses up in armour and helmet*
> *and goes to the battlefield,*
> *or in colorful brocade kimono, attending his lord*
> *at a banquet, sabi is like this old figure.* (Ueda, 1982)

Perhaps *sabi* is a concept not too far from presence. Much like *sabi*, we who study presence struggle to define and capture its nature, although anyone can identify it when they experience it. Like *sabi* in *haiku*, presence is the colour of virtual reality (if one can paraphrase Steuer's 1992 definition of presence that way). Like *sabi* which can be evoked using only seventeen syllables, very little external stimulation is required to bring about presence (as little as a few polygons on a screen; Slater *et al.*, 1995c), but at the same time it is difficult to arrange stimuli so as to guarantee a good presence experience. The most fundamental and important similarity between *sabi* and presence, however, is that they are both emergent properties of an information processing system which includes a human subject and a mediated space. The stimuli in *haiku* (the painstakingly selected seventeen syllables), much like those of a VE, need to be perceived by a subject, and decoded; then, as these stimuli are interactively processed in the context of the subject's previous experiences and knowledge of the world, the desired experience (be it *sabi* or presence) emerges. This dissertation is my attempt to convince the reader that this simple explanation of presence is true.

1.1 Aim of this dissertation

The past five years have seen significant advances in presence theory. Several important predictive models have appeared (see chapter 3), and there has been significant improvements in thinking about the presence phenomenon itself. Theorists such as Biocca (2003), Timmins and Lombard (2005), Lee (2004) and Slater (2002) have identified phenomena and constraints to presence theory which allow for more exact development and evaluation of models (see chapters 2 and 3). It is from these developments that this dissertation begins. It aims to develop a model of presence which is able to satisfy all of the following criteria:

1. Can explain presence in immersive as well as non-immersive media (Biocca, 2003; Lombard & Ditton, 1997)
2. Is equally suitable for explaining behaviour in real and virtual environments (Biocca, 2003; Timmins & Lombard, 2005)
3. Is consistent with current models of cognition, and can be explained as an evolved natural process (Biocca, 2003; K. M. Lee, 2004)
4. Can explain perceptual as well as content effects (Lessiter *et al.*, 2001; Robillard, 2003)
5. Can explain the duality of presence as a binary (Slater, 2002) and continuous phenomenon (Wirth *et al.*, 2007)
6. Is well supported empirically

Although a number of useful models of presence already exist (see chapter 3), none are able to meet all the constraints listed above. By careful examination of each extant model, this dissertation will define the Capacity Limited, Cognitive Constructionist model of presence (CLCC), conceptually show that it has more predictive power than extant models, and empirically validate its central aspects using six studies.

1.2 Assumptions and methods

This dissertation begins with the assumption that presence is a psychological phenomenon which occurs due to the interaction of a subject with external stimuli, which may or may not be related to a mediated environment (after Biocca, 2003; K. M. Lee, 2004). Following this assumption, the methods used are those of computational cognitive psychology; the subject's mind is assumed to be an information processing system which can be modeled to produce specific predictions which can be verified empirically (after Neisser, 1967). By and large, these assumptions and methods are non-controversial in the presence field, although this dissertation places far more emphasis on the role of information content than previous work (this is justified in chapter 4). The measurement of presence in all studies is done by the use of self-reports, which have been at the center of some debate (see 2.4.6 in chapter 2). This choice was made after a review of the psychometric properties of various self-report and other measurement methods, and is justified in section 2.6 (chapter 2).

1.3 Outline of the dissertation

This dissertation is divided into two parts. Part I (chapters 2 – 4) presents the capacity limited, constructionist cognitive (CLCC) model of presence, with its theoretical background. Part II (chapters 5 – 10) presents six studies which were used to empirically evaluate the CLCC model:

1.3.1 Outline of Part I

Chapter 2 presents a review of major conceptualizations of presence, and some efforts to unify theoretical strands. It then moves on to a review of measurement techniques, and concludes by selecting a definition and measure of presence to be used for the remainder of the dissertation. Chapter 3 reviews major presence models, and evaluates them in terms of theoretical soundness, empirical evidence, and their ability to deal with major theoretical problems in presence. Chapter 4 presents the CLCC model, and evaluates it theoretically, using the same criteria applied in chapter 3.

1.3.2 Outline of Part II

Chapters 6 and 7 present two working memory loading studies which examine presence under a dual-task paradigm. Chapter 7 uses a similar paradigm, but examines presence under divided attention. Chapters 8 and 9 present two survey conducted on computer gamers to examine the role of experience, content expertise and other processing factors in presence. Chapter 10 follows this with an experiment which examines the role of semantic coherence of the VE on presence. Chapter 11 presents conclusions for the dissertation as a whole.

Part I:

The CLCC model and its theoretical foundations

This part of the dissertation presents the capacity limited, cognitive constructionist model of presence as a theoretical entity. Chapter 2 begins by examining the concept of presence as it currently exists in the literature, and derives a flexible and powerful concept of presence to model. Measures of presence are then considered in order to find a valid and reliable basis for the empirical evaluation of the CLCC model in Part II. Chapter 2 critically reviews the most influential current models of presence in terms of their explanatory power and degree of supporting empirical evidence. Finally, Chapter 3 presents and justifies the CLCC model, and compares it, on a theoretical basis, with the extant models of presence reviewed in Chapter 2.

Chapter 2: Concepts and measures of presence

Currently, two significant theoretical hurdles stand in the way of presence research. First, although the presence experience is easily identifiable for the subject, a common scientific definition remains elusive; and second, little agreement exists in the field about how presence should be measured. Any empirical work must find some way to overcome these problems in order to claim some form of theoretical validity. This chapter has two goals: First, to critically review the current concepts of presence while focusing on efforts to unify them into a global concept, and argue for the importance of a cognitive perspective in unification; and second, to evaluate current presence measures so as to select a measure for the empirical component of this project which is consonant with the chosen presence concept.

2.1 The phenomenon of presence

Presence is a phenomenon which is associated with being a user of a system which mediates an environment. One definition which is a useful starting point for discussion is that provided by the International Society for Presence Research (ISPR), which is an international association of mostly academic presence researchers. That definition, which is heavily influenced by the synthesis of Lombard and Ditton (1998) is:

> *Presence (a shortened version of the term "telepresence") is a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience. Except in the most extreme cases, the individual can indicate correctly that s/he is using the technology, but at some level and to some degree, her/his perceptions overlook that knowledge and objects, events, entities, and environments are perceived as if the technology was not involved in the experience. Experience is defined as a person's observation of and/or interaction with objects, entities, and/or events in her/his environment; perception, the result of perceiving, is defined*

as a meaningful interpretation of experience. (International Society for Presence Research, 2000, paragraph 1)

This standardized sounding definition is actually a little misleading. In practice, pinning down exactly what presence is has been made far more difficult due to a proliferation of subtypes of presence. Here are some examples:

- Social presence: When subjects interact with simulated characters and mediated users as if they were not mediated (K. M. Lee, 2004)

- Co-presence: When the subject feels as if other mediated users are really in the same space as themselves (Slater *et al.*, 1999)

- Story presence: When subjects feel as if they are inside the events of a story (Brown *et al.*, 2003)

- Cognitive presence: When subjects take the VE as the basis for cognition rather than the non-mediated world (Nunez & Blake, 2001)

- Relational presence: When two subjects who experience a mediated relationship feel an emotional connection (Maguire & Connaughton, 2006).

- Spatial presence: Where subjects feel as if they are occupying the virtual space, and they are sharing the space with the objects in it (S. Lee *et al.*, 2004b) – more confusingly, this is often simply referred to as 'presence' (Slater, 2003a), which gives the impression that it is somehow the top of a hierarchy of presence concepts.

This is by no means an exhaustive list - Lombard and Jones (2006) are in the process of compiling a definitive taxonomy. A more constructive way to understand presence might be to consider all definitions as instances of a broader principle. In essence, presence arises when one person interacts with an environment (which may or may not be populated). This interaction could be mediated, or unmediated (Biocca, 2003;

IJsselsteijn, 2002; K. M. Lee, 2004; Lombard & Ditton, 1997). This very basic notion of presence originates from reports of operators of remote systems who, while working with their mediated systems, sometimes felt as if they were located at the remote site rather than at the workstation (Minsky, 1980). This phenomenon was described in detail by Sheridan (1992a), who suggested that presence could be applied to a user of a virtual environment, as this was simply a special case of teleoperation where the remote site was virtual rather than real. This extremely broad definition ("a feeling of being at the remote or virtual site rather than at the workstation") has continued to be extremely influential. This basic concept has never truly been replaced, but rather refined. (a few notable exceptions exist – see Floridi, 2005 for an example of a different but not widely adopted notion of presence based not on the sense of being at a remote site, but on the notion of a mismatch between perceptual sensors at the local and remote sites).

2.2 Unifying presence

Sheridan's simple concept may have been influential, but ultimately it was too nebulous to base a program of empirical research on. This led to a need for something more specific, and is probably what sparked the creation of new 'types' of presence. The creation of new definitions of presence solved the specificity problem, but created another: Along with new definitions came new measures, and comparing findings between studies became progressively more difficult (Kalawsky, 2000). Attempts were made to unify these concepts under a smaller number of umbrella concepts. For instance, Schloerb (1995) created the categories *subjective presence* and *objective presence* to simplify the picture (Schuemie *et al.*, 2001), but these were effectively categorizations of presence *measures* rather than presence *concepts*, and did not alleviate the problem. A similar but more substantial categorization was proposed by Heeter (1992), which operates on a level of analysis principle: Presence can be *personal* (the subject feels that they are in the virtual environment), *social* (the subject feels that the other characters and avatars in the virtual environment are real social actors) or *environmental* (the environment responds and reacts to the subject). As with Schloerb's effort, this is more of a methodological advance than a theoretical one. The categories help with deciding which measures and events are of interest, but in an overly reductionistic way. Consider a simple on-line gaming session as a common example of a virtual environment experience. Here players may experience personal presence (they feel present in the space), and simultaneously, social presence (as they work in teams to achieve the game's goals), and environmental presence (as they see doors open and objects tumble as they bump into them). It is methodologically useful to say that this subject is experiencing environmental or social presence, as it clears up what the variables of interest are; but from a theoretical perspective, it does not provide much insight, particularly because it forces the assumption that these forms of presence might be phenomenologically different beyond simply being different levels at which the phenomenon can be analyzed.

An important move towards a theoretically useful and empirically workable definition was published by Lombard and Ditton (1998). They began by reviewing both the theoretical and empirical literature. This included sources not only from the field of VE research, but from communications and media research as well. From this review, they created six categorizations for both explicitly stated and implicitly held formulations of presence:

1. *Presence as social richness* – presence arises when subjects perceive the environment as warm, sensitive, personal, intimate and immediate (these definitions typically arise from communications researchers).

2. *Presence as realism* –presence occurs when the medium is able of reproducing realistic objects, characters and events (these definitions arise mostly from human factors and systems engineers)

3. *Presence as transportation* – subjects experience either a sense of being in another place, that objects have been transported to share the same space as the subject, or that a group of subjects have been transported to the same place (these definitions arise from various sources, including communications research and literary theorists).

4. *Presence as immersion* – the medium becomes the primary source of sensory data for subjects (perceptual immersion), which may leads to a sense of involvement in the medium (psychological immersion). These definitions are often used by virtual reality and immersive media researchers.

5. *Presence as social actor within the medium* –subjects interact with characters in the medium as if they were social actors (these definitions are found mostly in the communications and shared spaces fields).

6. *Presence as the medium becoming a social actor* – subjects interact with the medium itself as if it were a social actor (these definitions are used mostly in the fields of robotics and human-computer interaction).

This ambitious project aimed to unify not only the presence concepts which cover what Heeter would call personal presence, but also those which deal social presence – a feat which has not been replicated since, and no doubt has led to this being one of the most widely cited papers in the field. Lombard and Ditton went further than simply organizing the definitions into six categories. They examined the categories searching for a common, unifying factor. They concluded that the factor is *the perceptual illusion of non-mediation*. They argue that in all six categories, presence occurs because the subject is responding to environment or characters as if the system (or more specifically, the technological aspects of the system) were not there. This occurs either because the system becomes a sort of window into the environment (in the classic 'ultimate display' sense proposed by Sutherland, 1965), or because the system is perceived as having transformed itself into a social actor (Lombard & Ditton, 1997).

An interesting consequence of Lombard and Ditton's definition is that it shifts the focus of presence onto perception. Much of the early examples provided by, for example, Sheridan (1992a; , 1992b), Ellis (1996) and Slater (1993a; , 1995c), had an underlying assumption that given the correct conditions, presence would occur automatically (part of the *presence as immersion* family of definitions). However, by highlighting perception rather than sensation, Lombard and Ditton suggest that, like perception, presence is a constructive, active process where the subject combines aspects of their own experience with the stimuli arising from the medium to create the experience for themselves:

> *Because it is a perceptual illusion, presence is a property of a person. However it results from an interaction among formal and content characteristics of a medium and characteristics of the media user, and therefore it can and does vary across individuals and across time for the same individual.* (Lombard & Ditton, 1997)

The idea that cognition must play an important role in presence has slowly gained popularity. Heeter has argued that presence can only arise from an interaction of top-down and bottom-up data, and furthermore, can only occur in a meaningful environmental context (Heeter, 2003). Nunez and Blake (2001) re-examined Lombard and Ditton's unifying principle using a cognitive perspective, with the aim of combining explanations of presence as a sensation (such as used by Sheridan, 1992a) with behavioural presence (for example, the notion of behavioural realism proposed by Freeman *et al.*, 2000). This approach (which was termed 'cognitive presence', misleadingly suggesting a new type of presence rather than an omnibus concept), cognition is argued to be the appropriate level of analysis, as all environments must be decoded and processed before any presence or action can occur. Cognition is therefore seen as the bottom-most level at which presence can be analyzed (Nunez & Blake, 2001). It should be noted that like Lombard and Ditton, Nunez and Blake do not present any discussion of the relative difficulty of using this simple level of analysis to explain the more complex forms of presence (such as social presence), and its usefulness in these domains is therefore unknown.

Cognitive presence as an adaptation of Lombard and Ditton's principle has only seen a limited amount of acceptance in the literature (used by, for instance, Hwang *et al.*, 2004; S. Lee *et al.*, 2004a). Rather, it is the examination of presence as a neuropsychological phenomenon which is showing signs of being a major unifying principle for presence. IJsselsteijn (2002) argues that presence should be explained as a multi-level phenomenon, with the bottom-most level being neural processing; then cognitive, phenomenological, social and other explanations should be considered as progressively complex levels which remain consonant with lower levels (IJsselsteijn, 2002). This view is perhaps an improvement over the monolithic approaches to presence unification proposed by Lombard and Ditton and Nunez, and Blake, as they allow switching to a particular level of explanation as required by the research problem without violating the principles of the more fundamental levels of explanation. Unfortunately, the undoubtedly frightening complexities of how a neural theory of presence might mesh with a social theory are not discussed. IJsselsteijn completes his argument by speculating on future research directions for a neural understanding of presence, including possible physiological correlates of presence (some of which are later echoed by Sanchez-Vives & Slater, 2005). Some of these ideas have been examined empirically (see 3.3.3.3 in chapter 3).

The most recent move towards unifying presence (which again comes from a cognitive perspective) is by Lee (2004). It is the first paper to substantially address possible ultimate as opposed to proximate causes of presence (Biocca, 2003 had previously noted the importance of considering evolutionary processes in presence). Lee argues that the interesting feature of presence is not that it occurs, but it sometimes fails to occur, given that it is an automatic reaction to particular stimuli (K. M. Lee, 2004). He therefore opposes presence concepts which emphasize the role of subjects' own conscious efforts to bring about the experience (such as the "suspension of disbelief" argument used by Slater & Usoh, 1993b). The mental architecture which Lee uses to support his argument is a cognitive modular architecture (Fodor, 1983; Tooby & Cosmides, 1990) which has recently been popularized by Pinker (1997; , 2002). In this architecture, the mind consists of various modules which are specialized to take in very particular types of data, transform them, and output to other similar modules. Each module is fairly simple, and optimized by natural selection for its particular purpose (Tooby & Cosmides, 1990). Because these modules are simple, they do not have complex filters for incoming data – they will respond even to stimuli which only resemble the intended objects; they thus respond to mediated scenes in the same way as they do to real scenes (K. M. Lee, 2004). For instance, the face processing module will respond to facial expressions even if the faces are depicted as simple cartoons rather than real faces (Heraz & Frasson, 2006). Of course, the mediated stimuli must fall within certain limits for the modules to take over their processing, hence the general finding that scenes with more fidelity lead to more presence (K. M. Lee, 2004). Although the modules process automatically, they can still be controlled to some extent by the central executive (Pinker, 1997). This is not however done by the willful *activation* of modules (as the "willing suspension of disbelief argument would contend), but rather by the willing *inhibition* of modules so that other reactions and behaviours can be expressed instead (Fox *et al.*, 2005).

Lee (2004) proposes that two types of presence can arise from this mechanism:

1. *Spatial presence through the folk physics module:* The folk physics module, which processes motion, three-dimensional spatial data and basic physics (Pinker, 1997) reacts automatically to sensory stimuli to form cognitive maps, mental models and other structures which are then handed off to other processes for the selection of appropriate action (Tooby & Cosmides, 1990). As these modules respond to simple features of sensory inputs, they can be easily fooled by mediation technologies (K. M. Lee, 2004). For instance, the spatial processing component of vision makes use of a few simple features such as relative size and occlusion to extract a three-dimensional space from a two-dimensional retinal image (Tong, 2002). Lee (2004) argues that most of the published findings which report on immersion effects on presence occur because the technology is capitalizing on the automatic interpretation of these features by the folk physics module. When this occurs, the subject will respond to the displayed scene in the way they react to unmediated objects, and probably experience a sense of spatial presence in doing so.

2. *Social and co-presence through the folk psychology module:* The folk psychology module is used to infer mental states in others (Baron-Cohen, 1995). It takes as input semantic information from the speech processing module, visual information and facial expression data from the face recognition and processing modules (Pinker, 1997). As with the folk physics module, it is activated by a few key features, and can be fooled by a display (K. M. Lee, 2004). The module is used to initiate social behaviours and make psychological predictions about other people. When activated, it leads to a sense that the object being interacted with has intelligence or personality (Baron-Cohen, 1995). As with the folk physics module, an active folk psychology module leads to subjects responding as they do to non-mediated persons, and a sense of social or co-presence could result.

The impressive advantage that Lee's unifying structure has over that of Lombard and Ditton (1998) is that because modules are specialized for particular tasks, he can retain the "typed" flavour of the different presence definitions (allowing spatial and social presence, for instance, to retain their own phenomenological distinctness), while at the same time explaining them using a single principle (i.e. the automatic processing by evolved modules). Nonetheless, for all the differences in the unifying attempts described above, it is interesting that all make use of cognitive explanations (or more generally, human information processing explanations) to produce theoretically reasonable umbrella terms.

2.3 Is presence a continuous or binary experience?

Until the year 2000, presence was generally considered to be a psychological variable, similar to mental workload (Sheridan, 1992a) and situational awareness (Prothero *et al.*, 1995). This led to the assumption that presence existed on a continuous scale, from highly present in the environment of interest, to totally absent and unaware. Perhaps the most compelling recent treatment of presence as a continuous phenomenon comes from Heeter (2003), who argues not only that presence fluctuates from moment to moment within a given environment, but also that individuals differ in terms of their presence experiences of the real world. The notion of presence as continuous is so convincing, that it has been carried through into most models and theories of presence (Biocca, 2003; Riva & Waterworth, 2003; Slater *et al.*, 1994; Waterworth & Waterworth, 2001; Wirth *et al.*, 2007). As all current models of presence are written at a conceptual rather than strictly predictive level, the question of whether presence is continuous or binary is not particularly critical – most models could theoretically cope with both. It is in the area of measurement that the question becomes critical.

Early measures of presence included methods which gave both quantified estimates on a scale and dichotomous ratings. For example, Barfield and Weghorst (1993), Hendrix *et al* (1996a; , 1996b) as well as Witmer and Singer (1998) used questionnaires which asked subjects to rate their experience of 'being there' (to varying degrees of sophistication) on Likert or semantic differential scales. On the other hand, Schloerb (1995) suggested that subjects could be shown a view through a window or peep-hole, which could either be simulated or real; and the more often the subject misidentified the simulated scene as real, the more present the subject could be said to be. Similarly, Hoffman *et al.*(1995) suggested that subjects could be asked to recall if the event occurred in a virtual space or real space; errors in favour of selecting the real space were taken as an indication of presence. It should be noted that these techniques of binary measurement (present/not present) did not make assumptions about the nature of the construct they were measuring. They simply used a binary quantification of what could well be a continuous experience.

That the presence experience is a dichotomous phenomenon was indirectly argued by Lombard and Ditton (1998). For them, presence is a perceptual illusion, and so it must either occur or not (interestingly, since then Lombard and Ditton have published a validated continuous measure of presence, which implies they currently see it as a continuous phenomenon – see Lombard & Ditton, 2004). Currently, the most obvious manifestation of presence as a binary concept is that proposed by Slater (2002). In this view presence is the outcome of a choice made at a perceptual level by the subject (see 1.3.1.1 in chapter 3). Subjects are therefore in one environment or the other, and the change between them is not gradual, but a sudden 'break' experience, which subjects are capable of reporting (Slater & Steed, 2000). Slater argues that presence is a Gestalt, and therefore cannot be graded. By analogy, he compares presence to reversible ambiguous figure illusions (see for instance Girgus *et al.*, 1977), which are also considered to be Gestalts. In these illusions, a stimulus figure seems to switch in meaning (one moment appearing as a rabbit, the next as a duck, for instance), but is never perceived as anything in between. He further argues that the 'break in presence' phenomenon would not exist if the experience were graded. While convincing, Slater's argument is missing two parts: First, while there is no doubt that presence can end by a break (this is documented in Slater & Steed, 2000), it is not clear that breaks are the only mechanism by which presence ends. In fact, Slater and Steed argue that subjects become presence in a VE by a different mechanism which does not lead to a break sensation (Slater & Steed, 2000). But, in order to become present in the VE, the subject must first stop being present in the real environment, which implies that there are at least two ways one can stop being present (see 4.4.2 in chapter 4). It is possible that the presence experience is in fact continuous, but in a 'break in presence' situation, the speed at which the change occurs makes it seem as if it is an instantaneous break. Secondly, Slater does not provide a proper role for attention in his dichotomous model of presence (Slater, 2002), especially given that many breaks in presence occur because of the sudden appearance of distracters (such as rendering glitches, interference from cables, etc.). Models of presence which do take into account modern findings on attention (see for instance Lombard & Ditton, 1997; Waterworth & Waterworth, 2001; Wirth *et al.*, 2007), conclude that because attention can be divided, and does allow for a limited degree of simultaneous processing (Baddeley, 1986; Treisman, 1969), presence should be possible to varying degrees based on the amount of attention that is available for processing the scene. Of course, because large changes in the stimuli field can lead to the sudden selection of a particular subset of stimuli for processing (Treisman *et al.*, 1992), the break in presence phenomenon can also be explained from a divided attention perspective. It is therefore highly likely that presence, if it is indeed as strongly related to attention as the literature claims, is a continuous phenomenon. The fact that it could be continuous, of course, neither precludes nor invalidates *measures* of presence which quantify it into two levels.

A final possibility to explain the apparent contradiction between models and measures in this regard is that presence is in fact a binary phenomenon, and that during any one VE experience, a subject will experience a large number of drifts into and interruptions out of presence. If one is using a binary measure of presence (such as BIPS discussed in 2.4.5 below), then it will seem as if the experience is binary, as each interruption will be reported. However, if a continuous measure is used (such as the TPI discussed in 2.4.1.6 below), it is possible that the subject gives an 'averaged' impression of the experience, interpreting more interruptions to mean less presence when asked to place it on a Likert scale. This would then give the impression that presence is continuous. Given the current empirical data, it is not possible to falsify this notion, and it must remain as a possible (and likely) explanation for the apparent contradiction in the literature.

2.4 Measures of presence

The second great debate in presence is that around measurement. Generally speaking there are three forms of presence measurement – the first two, self-report scales (Lessiter *et al.*, 2001; Slater *et al.*, 1994) and behavioural observation (IJsselsteijn, 2004; IJsselsteijn *et al.*, 2000), have been discussed since the early 1990s, but the third, breaks in presence (Slater & Steed, 2000), is a new and interesting addition to the literature. Any measure can have two uses: one is simply to measure presence in an empirical setting (Nowak *et al.*, 2006; for instance as is done by Robillard, 2003; Vinayagamoorthy *et al.*, 2004). Another, less familiar use is to use the measure to generate a theory of the dimensional structure and organization of presence by the use of factor analysis (as has been done by Lessiter *et al.*, 2001; Witmer *et al.*, 2005). This particular use has been fairly controversial, leading Waller and Bachman (2006) to criticize the technique on the following grounds:

1. Factor analysis (and principal components analysis) are essentially descriptive. They simplify data by projecting it onto a space (Neter *et al.*, 1988). Effectively, a factor analysis re-organizes the data, with some associated loss of variance; it is not an inferential technique and is therefore not capable of testing hypotheses (Neter *et al.*, 1988). It is not appropriate to use these techniques to test theories of the structure of presence (with one exception – see point 3 below).

2. Many decisions required by factor analysis are made *a priori*. These include the items to include in the analysis, the type of rotation applied to the data before projection, and (most importantly) the number of factors to project to (Waller & Bachmann, 2006). Although one can measure how well the data fits the current factor structure (by considering the factor eigenvalues and loadings - Neter *et al.*, 1988), this would only reflect that the data fits the number of factors chosen. If the choice of optimal number of factors scores could be objectively made, then it might be an appropriate method for testing theories, but at best all it can do is provide an estimate of how well a particular number of factors fit a particular sampled data set.

3. Although factor analysis is limited, it has some use. Rather than using the exploratory factor analyses currently in the literature (Waller & Bachmann, 2006), one can employ *confirmatory factor analysis*, where one begins with a specified factor structure (generated from a theory), and the data set is fitted to this structure. One can thus use this technique to compare theories (Waller & Bachmann, 2006). This is done by fitting all competing theories to a given

data set, and seeing which fits with the smallest error (of course, this is assuming that the data has a large degree of external validity, and that the sample is large enough to rule out idiosyncratic artifacts within the data).

Factor analysis is a popular technique in presence research, and it has largely gone unopposed as a method up till the publication of Waller and Bachmann's paper. Although their arguments are correct, it seems that they may have taken the argument too far. First, the way in which factor analysis is being used by at least some presence researchers (such as Lessiter *et al.*, 2001; Wirth *et al.*, 2007) is indeed descriptive. The complexity of the data is being reduced into a manageable number of concepts. In most cases, the factor analyses show that the factors are highly inter-correlated, which is interpreted by some as evidence that there is either a highly interdependent relationship between the factors (Lessiter *et al.*, 2001; Wirth *et al.*, 2007), or as evidence that the processes are causally linked (Wirth et al., 2007). Here, factor analysis is not being used to create theory, but rather to illustrate something about the interactivity of factors. Second, even exploratory factor analysis can provide important evidence for one of the most important questions in presence research: Is presence unidimensional or multidimensional (Slater, 2003a)? An exploratory factor analysis can resolve this issue because it allows one to partial out items from a questionnaire which can then be used to form a subscale which can be subsequently validated. So, for example, if a factor analysis reveals some new factor, one can set up an experiment where one manipulates immersion to see its effect on this factor – if it is related to presence, one can expect differences between immersion conditions. One could still argue (as suggested by Slater, 2003a) that this happens not because that factor is a part of presence, but because it is caused by presence; but this then is an issue not of presence theory, but of presence definition; and no amount of data can resolve definition debates (those must be resolved by consensus). In some sense then, it is true that factor analysis cannot generate theory. But for the problems currently facing presence, it is a very useful tool.

2.4.1 Self-report measures

Although it is common to refer to these measures as *subjective measures of presence* (for example in Schuemie *et al.*, 2001) perhaps to contrast them to physiological measures (which are referred to as objective measures, for example in Meehan, 2001), this review will not follow this tradition, as the word *subjective* has an erroneously negative connotation in the sciences. To avoid passing judgment on these measures on any grounds other than their validity and reliability (which can be evaluated objectively, in the scientific meaning of the word), this review will use the more accurate and descriptive term *self-report measure*.

These instruments have subjects give scaled reports of their experiences, typically using Likert or semantic differential items. These scales can be evaluated using psychometric techniques which can give good estimates of validity and reliability (Anastasi & Urbina, 1996). A large number of self-report scales have been developed, reflecting a lack of standardization in measurement in the field. This lack of standardization should itself not be a concern, as excursions into alternative measures are desirable in any developing field; The concern lies in the wide range of quality found in these measures. For example, Tromp *et al* (1998), for a single study, devised two items without any psychometric evaluation, while Lessiter *et al* (2001) used a four factor, 44 item scale which has been evaluated and refined with data from several thousand subjects in dozens of media conditions. This review will only focus on scales either with published psychometric properties, or those which have reached *de facto* status through wide use.

2.4.1.1 The Slater, Usoh and Steed (SUS) questionnaire (1994)

This early and highly influential questionnaire is perhaps the most widely used of all. It has not been psychometrically evaluated by its creators, but data from a moderate sample (n=101) showed a reasonable level of reliability (Cronbach's alpha of 0.77; Nunez, 2002). The questionnaire seems to be highly valid. Many studies using immersion manipulations show significant differences in the predicted direction (see Schuemie *et al.*, 2001 for a review of some of these studies). Although its validity has been criticized by its authors due to its low ability to distinguish between real and virtual environments in a single study (Usoh *et al.*, 2000), many studies show that it can distinguish between high and low immersion media conditions (Bracken & Skalski, 2006; Nunez & Blake, 2003b; Towell & Towell, 1997).

The SUS measures a single factor, spatial presence (the sense of being in the virtual space), with six items. Sample items from this factor include "I had a sense of 'being there' in the office space" (Not at all/Very much) and "I think of the office space as a place in a way similar to other places that I've been today..." (Not at all/Very much so). This factor is similar to that measured by the *Spatial presence* factor of the IPQ (see 2.4.1.3 below), the *Sense of physical space* factor of the ITC-SOPI (see 2.4.1.4 below), the *Spatial presence (self-location)* item of the MEC-SPQ (see 2.4.1.5 below), and the *Spatial presence* factor of the TPI (see 2.4.1.6 below). No doubt the reliability of the scale could be improved by increasing the number of items, but since the scale's publication, no moves have been made to update it.

There are two scoring methods for the SUS: One is the typical method of averaging the scores across all items, while the other method, which is generally only used by Slater and colleagues (for example Brogni *et al.*, 2003; Slater & Steed, 2000; Slater *et al.*, 1994; Usoh *et al.*, 2000), is to count the number of responses scoring 6 or 7, and then treat that number as a binomial distribution. This technique is intended to disambiguate between subjects who truly had a presence experience from the rest (Slater *et al.*, 1994), although it does not control for acquiescent subjects who have a bias towards giving high scores regardless of their experience.

2.4.1.2 The Presence Questionnaire (1998; 2005)

The first version of the PQ was not widely circulated, but the second version, published by Witmer and Singer (1998), reached fairly widespread use (for example, Darken *et al.*, 1999; Johns *et al.*, 2000; Mania & Chalmers, 2001; Schubert *et al.*, 2001). Recently, a third version has been published with a new factor structure (Witmer *et al.*, 2005). The well-known second version of the PQ contained 32 items, which were cluster analyzed into four factors (the term factor here is used loosely by the authors, not in the sense of a psychometric factor):

1. *Involvement/control*: How involved subjects become in the VE, and how much control they perceives they have. This factor is similar to the *Engagement* factor of the ITC-SOPI (see 2.4.1.4 below), the *Attention allocation* factor of the MEC-SPQ (see 2.4.1.5 below) and the *Engagement* factor of the TPI (see 2.4.1.6 below). Sample items are "How much did the visual aspects of the environment involve you?" and "Were you involved in the experimental task to the extent that you lost track of time?"

2. *Natural*: How natural the interactions with the environment are perceived to be, and how consistent the VE seems with reality. This factor is similar to the *Realness* factor of the IPQ (see 2.4.1.3 below), the *Natural* factor of the PQ (see 2.4.1.2 above), the *Naturalness* factor of the ITC-SOPI (see 2.4.1.4 below), and the *Perceptual Realism* factor of the TPI (see 2.4.1.6 below). Sample items are "Were you able to anticipate what would happen next in response to the actions that you performed?" and "How much did your experiences in the virtual environment seem consistent with your real world experiences?"

3. *Interface quality*: The subject's ability to concentrate on performing the main task in the VE, and how much the interface detracts from this. Sample items are "How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?" and "How much did the control devices interfere with the performance of assigned tasks or with other activities?"

4. *Resolution:* The ability to observe and manipulate objects from multiple viewpoints and distances. Sample items are "How closely were you able to examine objects?" and "How well could you examine objects from multiple viewpoints?"

The Cronbach's alpha for the PQ was estimated at 0.88 (n=152) by Witmer and Singer (1998), and another reliability analysis by Nunez (2002) found a value of 0.903, (n = 101). Immersion manipulation studies (for example, Mania & Chalmers, 2001; Nunez & Blake, 2003a; for example, Sallnäs, 1999; Youngblut & Perrin, 2002) suggest a satisfactory level of validity; furthermore, the PQ has been shown to moderately correlate with the SUS in two studies (r = 0.76 and 0.56; Nunez, 2002).

The PQ has been criticized by Slater (1999) on two grounds: First, the PQ measures the *subjects' estimations of theorized causes* of presence rather than presence itself. Second, the PQ conflates properties of the subject (such as their ability to focus attention) with properties of the system (such as the quality of the interface). While this does not reduce the criterion validity of the measure (its ability to predict the effect of presence on other variables), it does reduce its construct validity and makes research with the scale considerably harder (Anastasi & Urbina, 1996).

The third version of the PQ was developed by taking the 24 items from the second version with the highest item-factor correlations, and adding eight new items (Witmer *et al.*, 2005). These 32 items were administered to 325 subjects taking part in seven separate studies, and then factor analyzed. The resulting four factors were (in decreasing eigenvalue order):

1. *Involvement:* The degree to which the medium captures subjects' attention and their involvement in the VE. This factor is similar to part of the *involvement/control* factor in the previous version of the PQ, the *Engagement* factor of the ITC-SOPI (see 2.4.1.4 below), the *Attention allocation* factor of the MEC-SPQ (see 2.4.1.5 below) and the *Engagement* factor of the TPI (see 2.4.1.6 below). Sample items are "How much were you able to control events?" and "How responsive was the environment to actions that you initiated (or performed)?"

2. *Visual fidelity:* This is the same factor as *resolution* in the second version of the PQ (it consists of the same two items).

3. *Adaptation/immersion:* How much subjects focus on the VE content and how quickly they adjust to the VE experience. This factor is similar to the *Involvement* factor of the IPQ (see 2.4.1.3 below). Sample items are "How much delay did you experience between your actions and expected outcomes?" and "How quickly did you adjust to the virtual environment experience?"

4. *Interface quality:* This is the same factor as second version of the measure.

This new version has a small improvement in reliability (alpha = 0.91), but by having been factor analyzed, allows for a more detailed interpretation of the results. It should be noted that unlike the second version of the PQ, the order of these factors is ranked by eigenvalue order (such that the first factor explains the most variance, followed by the second, etc.). Unfortunately, Slater's two major concerns (discussed above) have not been addressed, although the factor analysis eases the conflation between individual and systems factors somewhat.

The scale was designed for use in immersive or semi-immersive environments, and is not well suited for use in non-immersive media such as text (Nunez, 2002; Nunez & Blake, 2003b). Items such as *How natural was the mechanism which controlled movement through the environment* (item 6 from version 3 - Witmer *et al.*, 2005), *How well could you actively survey or search the environment using touch* (item 13 from version 3 - Witmer *et al.*, 2005) and *How easy was it to identify objects through physical interaction, like touching an object, walking over a surface, or bumping into a wall or object* (item 29 from version 3 - Witmer *et al.*, 2005) do not have much meaning for non-immersive media, and are likely to increase the error variance in the measure due to subject confusion. Unfortunately, such items are spread between the factors, so excluding a factor from the scale does not solve the problem.

2.4.1.3 The Igroup presence questionnaire (2001)

The IPQ (Schubert *et al.*, 2001), was constructed from 75 Likert-type items, some of which were developed for the scale, while others were taken from extant measures including the PQ (Witmer & Singer, 1998), the SUS (Slater *et al.*, 1994), Ellis *et al.*'s measure (1997 in Schubert *et al.*, 2001), Carlin *et al.*'s measure (1997 in Schubert *et al.*, 2001), Towell & Towell's measure (1997) and Regenbrecht *et al.*'s measure (1998 in Schubert *et al.*, 2001). During development, two sets of data were collected: one used all 74 items completed by 224 users of systems of varying degrees of immersion (Schubert *et al.*, 2001), and the other with but with a reduced set of items (immersion related items were excluded) on 269 subjects. These data were factor analyzed to extract three factors:

1. *Spatial presence:* The sense of being in a coherent virtual space. This factor is similar to that measured by the SUS (see 2.4.1.1 above), the *Sense of physical space* factor of the ITC-SOPI (see 2.4.1.4 below) the *Spatial presence (self-location)* item of the MEC-SPQ (see 2.4.1.5 below), and the *Spatial presence* factor of the TPI (see 2.4.1.6 below).

Sample items are "In the computer generated world I had a sense of 'being there'" and "I felt present in the virtual space."

2. *Involvement:* A sense of being captivated by, and of focusing attention on the VE. This factor is similar to the *Involvement* and *Adaptation/immersion* factors of the PQ (see 2.4.1.2 above), the *Engagement* factor of the ITC-SOPI (see 2.4.1.4 below) and the *Attention allocation* factor of the MEC-SPQ (see 2.4.1.5 below). Sample items are "I had a sense of acting in the virtual space, rather than operating something from outside." and "I felt like I was just perceiving pictures." (reversed item).

3. *Realness:* How real the VE seems, and how consistent the experience in the VE is with their experiences in the real world. This factor is highly similar to the *Natural* factor of the PQ (see 2.4.1.2 above), the *Naturalness* factor of the ITC-SOPI (see 2.4.1.4 below), and the *Perceptual Realism* factor of the TPI (see 2.4.1.6 below). Sample items are "How real did the virtual world seem to you?" and "The virtual world seemed more realistic than the real world."

The current IPQ contains 29 items, and is available in German, Dutch and English; however, only the German version (the first developed) has been psychometrically evaluated, leaving the psychometric properties of the other two versions unknown (igroup.org, 2004). The reliability of the IPQ (from development data - Schubert *et al.*, 2001), is adequate, with Cronbach's alpha scores in the low to mid range (around 0.70). As two data sets were collected during development, it was possible to conduct confirmatory factor analysis using the second data set on the factors extracted from the first (Schubert *et al.*, 2001). The similarity of factor structure and factor loadings indicates the factors are stable. There is less evidence of the scale's validity. Although the IPQ is freely available, it has seen limited use. Brown *et al* (2003) found differences in the expected direction with an immersion manipulation. Similarly, Waterworth & Waterworth (2003) found higher IPQ scores for subjects experiencing concrete stimuli than abstract stimuli. Finally, Schulte-Pelkum *et al.* (2005) found that the sense of self-induced vection in an audio environment correlated positively with

IPQ scores. The scale was not designed for immersive systems, and most of the development data was collected using mostly desktop VR systems, and a few HMD and cave users (Schubert *et al.*, 2001). It is not known how the scale fares with non-immersive media (books, etc).

2.4.1.4 The Independent Television Commission's Sense of Presence Inventory (2001)

The ITC-SOPI was designed to be a cross-media measure, and has been thoroughly psychometrically evaluated (Lessiter *et al.*, 2001). Development began with a literature review which identified thirteen constructs of importance, and 63 Likert-type items were derived from these (Lessiter *et al.*, 2001). Data from six different immersion conditions (n=604) was factor analyzed to extract a four factor structure, which explains slightly more than a third of the total variance (Lessiter *et al.*, 2001). The four factors are (in decreasing eigenvalue order):

1. *Sense of physical space:* A sense of being in the VE space, and that objects and characters occupy the space. This factor is similar to that measured by the SUS (see 2.4.1.1 above), the *Spatial presence (self-location)* item of the MEC-SPQ (see 2.4.1.5 below), and the *Spatial Presence* factor of the TPI (see 2.4.1.6 below). Sample items are "I had a sense of being in the scenes displayed" and "I felt that all my senses were stimulated at the same time."

2. *Engagement:* A sense of psychological involvement with and enjoyment of the VE content. This factor is similar to the *Involvement* and *Adaptation/immersion* factors of the PQ (see 2.4.1.2 above), the *involvement* factor of the IPQ (see 2.4.1.3 above), the *Attention allocation* factor of the MEC-SPQ (see 2.4.1.5 below) and the *Engagement* factor of the TPI (see 2.4.1.6 below). Sample items are "I had a sense that I had returned from a journey" and "My experience was intense."

3. *Naturalness (Ecological validity):* The sense that the VE and its content are lifelike or realistic. This factor is highly similar to the *Realness*

factor of the PQ (see 2.4.1.2 above), *Natural* factor of the PQ (see 2.4.1.2 above), and the *Perceptual Realism* factor of the TPI (see 2.4.1.6 below). Sample items are "The content seemed believable to me." and "I had a strong sense that the characters and objects were solid."

4. *Negative effects:* Negative physiological effects (such as dizziness and eyestrain) – this factor is negatively correlated with the other three factors. Sample items are "I felt tired." and "I felt nauseous."

The final form of the scale retained 44 items over the four factors (physical space: 19 items; engagement: 13 items; naturalness: 5 items; negative effects: 6 items). The four factors are conceptually independent, so no single presence score can be computed; rather, each administration produces four independent scores which measure separate aspects of the experience (Lessiter *et al.*, 2001; Nunez & Blake, 2006).

The ITC-SOPI factor structure was verified by factor analyzing two random sub-samples of the original dataset. Cronbach's alpha for all factors were acceptable (physical space = 0.94, engagement = 0.89, naturalness = 0.76, negative effects = 0.77). The scale also has a good degree of validity, having being used in a number of published studies. In most cases, the scale is sensitive to immersion manipulations in the expected directions (Dillon *et al.*, 2001; Freeman *et al.*, 2004; Lessiter *et al.*, 2001). It is also sensitive to variation in VE content (Dillon *et al.*, 2001). The large sample used to develop the scale, and the large number of published studies using the it make this arguably the best psychometrically understood measure available.

2.4.1.5 The MEC Spatial Presence Questionnaire (2004)

The MEC-SPQ (Vorderer *et al.*, 2004), is unique in that it includes three forms: long (eight items per subscale), medium (six items per subscale) and short (four items per subscale). The scale consists of Likert-type items measuring eight constructs derived from the MEC model of presence (see 3.3.4 in chapter 3):

1. *Attention allocation:* How much attention the subjects devote to the VE. This factor is similar to the *Involvement* factor of the IPQ (see 2.4.1.3 above), the *Involvement* and *Adaptation/immersion* factors of the PQ (see

2.4.1.2 above), the *Engagement* factor of the ITC-SOPI (see 2.4.1.4 above), and the *Engagement* factor of the TPI (see 2.4.1.6 below). Sample items are "I devoted my whole attention to the [medium]." and "The [medium] captured my senses."

2. *Spatial situation model:* The subject's confidence in the accuracy of their SSM (see 3.3.4.1 in chapter 3 for a description). Sample items are "I was able to imagine the arrangement of the spaces presented in the [medium] very well." And "Even now, I still have a concrete mental image of the spatial environment."

3. *Spatial presence (self location):* The sense of having been transported to the VE. This factor is similar to that measured by the SUS (see 2.4.1.1 above), the *Spatial presence* factor of the IPQ (see 2.4.1.3 above), the *Sense of physical space* factor of the ITC-SOPI (see 2.4.1.4 above), and the *Spatial Presence* factor of the TPI (see 2.4.1.6 below). Sample items are "I had the feeling that I was in the middle of the action rather than merely observing." and "It was as though my true location had shifted into the environment in the presentation."

4. *Spatial presence (possible actions):* The sense of being able to interact with and move around the VE. Sample items are "I had the impression that I could be active in the environment of the presentation." and "I had the impression that I could reach for the objects in the presentation

5. *Higher cognitive involvement:* The degree to which the subject thought more more deeply about the VE content (i.e. beyond spatial properties). Sample items are "I imagined precisely what it must be like to further explore the world presented in the [medium]." and "The [medium] presentation activated my thinking."

6. *Suspension of disbelief:* The degree to which subjects take a critical view of the experience and allow themselves to influenced by errors and inconsistencies in the VE. Sample items are "I concentrated on whether

there were any inconsistencies in the [medium]." (reversed item) and "It was not important for me whether the [medium] contained errors or contradictions."

7. *Domain specific interest:* The degree of interested in and knowledge of the VE content. Sample items are "I have felt a strong affinity to the theme of the [medium] for a long time." and "Things like the ones in the [medium] have often attracted my attention in the past."

8. *Visuospatial imagery:* The subject's inherent ability to form mental images and manipulate visual data mentally. Sample items are "When someone shows me a blueprint, I am able to imagine the space easily." and "I can vividly imagine how small I would seem at the foot of a high mountain."

Note that these constructs are not produced by factor analysis, but reflect the structure of the MEC model (giving the scale a high degree of construct validity in terms of that model). The constructs likely correlate, and the order in which they are presented above does not reflect any relative degree of predictive power. All three forms have been psychometrically evaluated (Vorderer *et al.*, 2004). The scale is available in English, German, Portuguese and Finnish, making it the most translated scale currently available.

The MEC-SPQ measures only experience and not system variables, as it aims to be used in various immersion conditions (Vorderer *et al.*, 2004; Wirth *et al.*, 2007). It is thus mmune to Slater's criticism of the PQ (Slater, 1999). Although it combines measures related to the specific experience (such as *spatial presence* and *suspension of disbelief*) with inherent properties of the subject (such as *domain specific interest* and *visuospatial imagery*), it avoids conflation by dividing these into subscales, so that independent investigations can be made on their impact.

Due to its tight coupling with the MEC model, the scale has seen limited use, but has shown high reliability, with Cronbach's alpha values (even for the short, 4-item form) around 0.8 in studies involving various media forms (Vorderer *et al.*, 2004). In terms of validity, the scale does well – it is sensitive to media manipulations in the appropriate direction (see for example Böcking *et al.*, 2004; Gysbers *et al.*, 2004; Laarni *et al.*, 2004; Vorderer *et al.*, 2004).

2.4.1.6 The Temple Presence Inventory (2000-2004)

The TPI is still under development by Lombard & Ditton. To develop the questionnaire, two conditions were used (n=600) – a high immersion condition (to maximize presence), and low immersion condition (Lombard *et al.*, 2000a). Subjects completed 103 items which were developed from five concepts of presence found in the literature (Lombard *et al.*, 2000a). This was factor analyzed into eight factors, and the 42 items highest loading factors were retained. The scale is interesting as it combines individual and social presence in one instrument. This review shall only consider the three factors which measure aspects of individual presence (in decreasing eigenvalue order):

1. *Spatial presence:* The sense of feeling in the VE, and that the others shared the space. This factor is similar to that measured by the SUS (see 2.4.1.1 above), the *Sense of physical space* factor of the ITC-SOPI (see 2.4.1.4 above) and the *Spatial presence (self-location)* item of the MEC-SPQ (see 2.4.1.5 above). Sample items are "How much did it seem as if the objects and people you saw/heard had come to the place you were?" and "To what extent did you experience a sense of being there inside the environment you saw/heard?"

2. *Engagement:* Psychological and sensory involvement with VE content. This factor is similar to part of the *involvement* factor of the PQ (see 2.4.1.2 above), the *Engagement* factor of the ITC-SOPI (see 2.4.1.4 above), and the *Attention allocation* factor of the MEC-SQP (see 2.4.1.5 above). Sample items are "To what extent did you feel mentally immersed in the experience?" and "To what extent did you experience a sensation of reality?"

3. *Perceptual realism:* The degree to which the VE matches real experiences, in terms of perception rather than content. This factor is highly similar to the *Realness* factor of the IPQ (see 2.4.1.3 above), the *Natural* factor of the PQ (see 2.4.1.2 above), and the *Naturalness* factor of the ITC-SOPI (see 2.4.1.4

above). Sample items are "Overall how much did touching the things and people in the environment you saw/heard feel like it would if you had experienced them directly?" and "Overall, how much did the things and people in the environment you saw/heard look they would if you had experience them directly?"

These factors are similar to the ITC-SOPI, and although no correlation studies have been conducted, it seems likely that the TPI and ITC-SOPI would match up factor by factor. The TPI is highly reliable, with Cronbach's alpha scores between 0.91 and 0.79. Although not many studies have made use of the scale, the procedure used in its construction (comparing high-immersion to low-immersion conditions as well as a strong basis in the literature) promises a degree of construct validation.

2.4.2 Behavioural observation

Some researchers observe behaviours to estimate presence. If the behaviours are consistent with the demands of the VE rather than the real world, then this is taken as an indication of presence (Freeman *et al.*, 2000; Held & Durlach, 1992). Subjects are normally observed in relation to some specific element of the VE (such as interacting with another character), and presence is inferred from this. It is therefore an instantaneous measure of presence (i.e. presence is evaluated at particular points rather than over the whole experience).

Exactly what is observed varies between researchers. No one theory drives these measures, apart from the general principle that a subject acting 'as if' they are in the VE is likely to be experiencing presence (Freeman *et al.*, 2000; Held & Durlach, 1992; Slater, 2003b); hence this notion is also called 'behavioural realism' (Freeman *et al.*, 2000). Some of these measures have been shown to correlate with self-reports, suggesting that they have some validity (see for instance Meehan, 2001; Slater *et al.*, 1995c). Determining reliability is difficult, as each measure is generally designed and used only for a specific situation. Generally speaking, three classes of behaviour have been used as presence estimators: automatic social behaviours, reflex actions and posture/sway measures.

2.4.2.1 Automatic social behaviours:

Sheridan (1992a) and IJsselsteijn *et al.* (2000) suggested that the degree to which subjects engage in customary social gestures during a VE experience (such as reaching for a handshake, making eye contact, or engaging in conversational turn-taking) could be used as an indicator of presence. Bailenson and colleagues examined the interpersonal distance kept between the subject and an animated agent during a VE experience, and found that the average interpersonal distance was more like the distance kept between real individuals when characters made eye-contact with the subject (Bailenson *et al.*, 2001). This finding was later replicated and validated by the fact that some subjects believed the characters to be controlled by other subjects (Bailenson *et al.*, 2003).

Another behavioural presence indicator is emotional response. Huang and Alessi (1999) argued that facial expressions in response to emotional environments could be used as presence indicators. Ravaja *et al.*(2004), found support for this, by noting that particular facial muscles activate in response to specific VE content (see 2.4.3 below), but due to the limits of that study, this technique should not yet be considered as validated.

2.4.2.2 Reflex actions

Among the first suggestions for measuring presence by observation are those of Held and Durlach (1992) and Loomis (1992). Held and Durlach argue that a subject's response to unexpected VE elements can be indicators of presence. So, if an unexpected event occurs in a VE, a present subject should be surprised by it, and exhibit a startle response (Held & Durlach, 1992). This method was validated by Wilson *et al.* (1997) who found that subjects did indeed show startle responses to unexpected VE content. However, this simple idea is complicated by evidence that the startle response may have a more complex relationship with presence. A break in presence is a form of startle response (Slater & Steed, 2000) made in response to an unexpected glitch or inconsistent element in the VE; but in this case, contrary to Held and Durlach's expectations, the startle response is an indicator of no presence (Brogni *et al.*, 2003). Loomis (1992) proposed a more sophisticated version of this idea. He argues that behaviour can occur in response either to the elements of the VE (*distal attributions*), or to events on the display, such as glitches (*proximal attributions*).

Presence would only be indicated by responses which have distal attributions (the startle response, for example, can be made in response to both proximal and distal attributions, so does not fit the bill). Reflex actions in response to VE elements (such as dodging from away from objects in the VE) constitute such a class of behaviours (Loomis, 1992). This idea was evaluated by Slater *et al.* (1995c), in an ingenious study in which a radio was placed both the real and virtual environments. The distance between the was then manipulated as a tone was played from the real radio (Slater *et al.*, 1995c). Subjects were tasked with pointing to the radio which was playing the sound, and it was predicted that present subjects would point to the virtual radio rather than the real one. This method was validated by a significant correlation between the angle-error on the pointing task and self-reported presence (Slater *et al.*, 1995c).

2.4.2.3 Posture and sway:

Several studies have used how subjects respond to perceived motion by swaying to compensate for virtual motion (which would be a distal attribution – Loomis, 1992). Illusory self-rotation (vection) was one of the first forms to be considered. Cohn *et al.* (1996) found that subjects who were asked to reach for objects in a VE rotating on the yaw axis compensated for virtual rotation in proportion to the speed of rotation. A more complex design was employed by Ohmi (1998), who crossed degree of rotation with display type. More immersive displays led to more sway, and sway was synchronized to the VE under all display conditions (Ohmi, 1998). Freeman *et al.* (2000) and IJsselsteijn *et al.* (2001) used driver's perspective footage of driving along a winding track displayed either using a monoscopic or stereoscopic display. Subjects swayed synchronized to the stimuli, and the stereoscopic condition produced more sway. Furthermore, the degree of sway was correlated with presence self-reports (Freeman *et al.*, 2000).

A more sophisticated body sway metric was used by Emoto *et al.* (2004). They reasoned that looking at scenes with a small field of view would affect the equilibrium system, as the information provided does not match experience, and the subject would try to compensate with their posture (Emoto *et al.*, 2004). Therefore, as the field of view tends towards a natural value, equilibrium should return to normal and result in less posture compensation and sway. They indeed found that a wider fields-of-view led to less sway. This finding, as with Slater & Steed's use of the startle response in breaks in presence (Slater & Steed, 2000), illustrates the difficulty of interpreting behavioural metrics as simple direct correlates of presence.

2.4.3 Physiological measures

Several physiological variables are considered correlates of presence, although as with behavioural observation, no standard yet exists. An early study by Jorgensen *et al.* (1997) noted that VEs can produce measurable changes in physiology (in particular heart rate and galvanic skin response), and soon after efforts were underway to estimate presence this way. The most successful examples of this method are associated with Meehan (Meehan, 2001; Meehan *et al.*, 2002; Meehan *et al.*, 2003). Subjects were placed into a virtual environment in which they must navigate around a deep pit in the floor. Not surprisingly, taking part in the task is a stressful experience (Meehan *et al.*, 2002). The VE was implemented in various immersion conditions, from a four-wall cave (Meehan, 2001) to an HMD based, body-tracked system with passive haptic feedback (Meehan *et al.*, 2002; Meehan *et al.*, 2003). The studies examined changes in mean heart rate, skin temperature and galvanic skin response, and were validated both by behaviour observation, and by the SUS scale. The results are fairly impressive: change in mean heart rate and galvanic skin response were significant predictors of SUS scores (Meehan, 2001; Meehan *et al.*, 2002). Change in mean heart rate was also sensitive to changes in simulation display update rate, albeit non-linearly (Meehan *et al.*, 2003). Others such as Wiederhold *et al.* (2002), Preston (1998) and Zimmons and Panter (2003) have found similar effects, while Dillon *et al.* (2001) found a difference in heart rate across content conditions (exciting content/calm content). Meehan concludes that such measures can be used to evaluate presence, but only for stressful environments, because one needs in the environment a situation which should raise heart rate and lower galvanic skin response for present subjects (Meehan *et al.*, 2003); in effect, something to allow a distal attribution. Slater (2002) has criticized this type of measure for requiring artifacts to be inserted into the VE (such as the pit); he argues that it is contrived and impractical for general purpose applications (Slater, 2002).

Other physiological variables have produced similarly limited results. A study by Ravaja *et al* (2004) used facial electromyography (EMG), galvanic skin response (GSR) and change in heart inter-beat intervals (IBI) as well as the ITC-SOPI self-report scale. EMG activity related to experiencing a positive mood state (i.e. increased *zygomaticus* and *orbicularis oculi* activity and reduced *corrugator supercilii* activity) predicted spatial presence. Similarly, spatial presence was predicted by GSR activity which indicates increased arousal. They found no effects on the IBI (which contradicts Meehan *et al.*, 2002). Although these findings are encouraging, they cannot be generalized due to confounds in the design. The specific EMG indicators, for instance, show positive mood, but this is hardly surprising as they occur at the achievement of a goal. Although this activity is correlated with presence, it is likely that more present subjects became more involved in the game and thus felt more satisfaction (and hence positive affect) on completing the goal. It is unlikely (as conceded by Ravaja *et al.*, 2004) that this is a general purpose presence measure. GSR is more promising, but increased arousal is known to follow changes in focused attention. Although attention has been strongly associated with presence (Schubert *et al.*, 2001; Wirth *et al.*, 2007; Witmer *et al.*, 2005), such measures could only explain a small amount of presence variance, due to the large number of other factors involved.

2.4.4 Brain imaging

Brain imaging has recently been considered as a possible measure of presence. Unfortunately, both VR and brain imaging require bulky, sensitive equipment (both often relying on detection of magnetic fields), which may interfere with each other. Furthermore, brain imaging requires subjects to keep still to reduce measurement error due to muscular activity (Hoffman *et al.*, 2003). One study used used fiber-optic goggles to provide stereoscopic display to subjects inside an fMRI scanner (Hoffman *et al.*, 2003). Preliminary work showed the display could produce differing immersion levels measurable as changes in presence (Hoffman *et al.*, 2003), effectively opening the door to fairly immersive environments within the confines of the fMRI magnet.

EEG has given some interesting findings. Bischof and Boulanger (2003) placed subjects in a complex maze while monitoring theta band oscillations, which have been linked to hippocampal activity during navigation in rodents (O'Keefe & Reece, 1993). They found theta activity predicted specific behaviours in the maze (particularly making navigation errors). Although they did not measure presence, their results indicate that EEF measurement inside is VE is possible. Mikropolous *et al.* (2004) used EEG to examine decreases in alpha-band oscillations and increases in gamma-band oscillations, which are associated with shifts in attention and with visual awareness (Mikropoulos *et al.*, 2004). Measures were taken on seven subjects navigating four VE conditions, increasing in complexity and fidelity. As expected, EEG readings showed a decrease in alpha-band oscillations, and complementary increase on gamma-band oscillation. Mikropolous *et al.* conclude that because the EEG activity is sensitive to manipulations of scene fidelity, these wavelengths could be used as an objective measure for presence. However, this reasoning is flawed: First, complex environments (with more data to process) will lead to more attention focused on the VE, as habituation will take longer (Kosslyn & Thompson, 2003). Given the large manipulations used in the study (e.g. one condition was texture mapped while another was not - Mikropoulos *et al.*, 2004) the changes in EEG may only be indicating this phenomenon rather than presence; or, at best, the measure is only sensitive to broad-brush manipulations of immersion. Second, Mikropoulos *et al.* (2004) used only the EEG measure without independent validation measure, making the conclusion about presence essentially post-hoc.

2.4.5 Breaks in presence (BIPs)

This technique attempts to get around the problems of having subjects scale their own experiences (as is done in self-report measures). In a BIPS measure, subjects report on the occurrence events which are easy to detect, and because only their occurrence is reported on, problems of individual differences in relative scaling are overcome. Subjects are asked to report on "breaks in presence", which is the sensation of a sudden change or shift in attention which occurs when one is suddenly moved away from being present in a VE. Before the experience, subjects are instructed to report each time they experience a "transition to reality", which is defined for them as the awareness that they are in the laboratory where the experience is happening. Subjects report either by simply calling out "Now" (Slater & Steed, 2000), or by calling out loudly the reason for the break, for instance "Cable pull" (Brogni *et al.*, 2003). A large number of BIPs is associated with a low presence experience. For validation, Slater and Steed used a Markov chain model to predict presence from the distribution of BIPs over time (Slater & Steed, 2000), and the model was then validated using SUS data. Brogni *et al.* (2003) subsequently showed that a simple count of BIPs is a good predictor of SUS scores. The technique is therefore generally validated, although only by a few studies.

One problem with BIPs is the issue of false negatives. It is fairly obvious that a false positive cannot occur, because even if the subject did not truly experience a break and calls it, the calling of the break will itself induce a BIP. However, it is possible that a break could occur and the subject does not report it, and furthermore, the probability of a subject correctly reporting a break is probably inversely proportional to the difficulty of the experimental task. So, for example, if a subject is engaged in a difficult task, a BIP might go unreported as the subject allocates most resources to the task and forgets to call out the BIPs; this is a problem with all measures which give the subject a dual task (see Freeman *et al.*, 2000 for another example of a dual-task measure). Note however that this is a difficulty with the reliability and sensitivity of the measure, not with its validity.

2.4.6 Questionnaires versus other types of measures

All measures must serve two masters: reliability (measuring the construct with minimal error), and construct validity (measuring the actual construct it aims not, and no other) (Anastasi & Urbina, 1996). Showing that a scale is reliable is simple – well established techniques exist for doing so (Cronbach's alpha coefficient, factor analysis, test-retest techniques, etc.). Showing that a scale has construct validity is much harder (Anastasi & Urbina, 1996), and it is here that the measured debate has centered.

One side of the measurement debate opposes self-report scales on the basis that they are not able to measure presence at all (Slater, 2004; Usoh *et al.*, 2000). In an interesting study, Usoh *et al.* (2000) created a virtual environment of their laboratory, and placed 10 subjects into that VE and 10 into the actual laboratory with the same task, and then took SUS and PQ measurements. The prediction was that the real laboratory should produce more presence than the VE. The results were not as predicted. No significant difference between environments was found on either scale, although there was a significant difference in the expected direction when the number of high scorers (scoring 6 or 7) on the SUS was compared (Usoh *et al.*, 2000). Usoh *et al.* reached two conclusions: first, subjects will interpret questions to make sense in the given context, and respond on that basis. Second, presence questionnaires are not suitable for cross-environment comparisons due to their lack of sensitivity.

The study, although interesting, has some significant flaws. First, as discussed by Usoh *et al*, the power of the study is low due to the small sample. They argue that they expected a large effect, so a small sample should have been sufficient to show the effect. However, given that there is an indication of a difference (both from the significant difference between high scorers, and from the fact that the mean differences, although not significant, are in the predicted direction), it would have made sense for them to continue collecting data to see if the difference disappeared with a larger sample. The effect may thus simply be that it is smaller than expected, particularly given the fact that presence questionnaires had, at the time of that paper's publication, already shown a number of success at picking out differences between immersion conditions (e.g. Hendrix & Barfield, 1996a; Prothero & Hoffman, 1995; Slater *et al.*, 1995c). A second weakness in the study comes from the selection of criterion. While it is reasonable to expect subjects in the real world to feel more presence on average than subjects in a VE, it is not necessarily the case that self-report scales will work for real environments. Usoh *et al* argue that the questions from the questionnaires used make sense for real environments, but this proposition is hard to defend. Consider these items, form the perspective of a subject in the real world condition: "To what extent were there times during the experience when the office space was the reality for you?" and "During the time of the experience, did you often think to yourself that you were actually in the office space?" (from the SUS) or "Overall, how much did the you focus on using the display and control devices instead of the experience and experimental tasks?" and "How natural was the mechanism which controlled movement through the environment?" (from the PQ). Some subjects may have understood the intention of the item correctly, but many could have been confused and responded with an answer in the center of the scale to be safe. Regardless of what the subjects actually did do, it is safe to assume that this confusion translated into a degree of random answering, which would in turn increase the error variance, and make it harder to detect small differences. In this regard, it is telling that for both the SUS and PQ scores, the real world condition produced higher standard deviations, indicating more randomness (and possibly confusion) in response. The conclusion that subjects interpret questions in some context and respond on that basis is true (indeed, it is a fundamental principle of psychometrics – Anastasi & Urbina, 1996), but the significance of that conclusion is not that subjects are unable to report correctly on their experiences, but rather that there will always be a degree of idiosyncrasy to a subject's response. Some of it can be controlled for (such as by ensuring that the context in which they respond is meaningful and not confusing), but the rest must be dealt with by making use of large enough samples when making comparisons, to overcome the error variance.

The problems raised by Usoh *et al* are nonetheless worrying, because they do highlight some difficulties of working with questionnaire data. Another interesting paper from Slater (2004), presents the central problem in presence measurement: What is it exactly that we are measuring? Slater notes that a presence questionnaires should at the very least be able to distinguish between reality and a virtual environment, citing Usoh *et al* (2000). This is true in principle, but one cannot simply expect a questionnaire to work in contexts beyond those for which it was designed (if anything, Usoh *et al*'s argument is one against the particular questionnaires used, and not against questionnaires in general). A second objection raised by Slater (2004) is that presence cannot be measured by questionnaires, as these are suited to measuring mental states, and no presence researcher has yet identified the mental state which makes presence (Slater, 2004). While it is true that presence has not been defined as a unique mental state, it is not true that it does not exist as a well documented, independent phenomenon. Unlike "colorfulness", the straw-man concept ingeniously conjured up by Slater to show the dangers of reification, there is evidence that presence exists as a concept separate to the act of studying it. Several independent researchers (Freeman *et al.*, 2000; IJsselsteijn, 2004; Meehan, 2001) have reported that subjects placed in virtual environments respond to the stimuli presented to them as if they were real. This is an indication that presence indeed is a real phenomenon, which can presumably be measured somehow. Furthermore, researchers from other fields have independently identified and documented phenomena very similar to presence, such as placeness (Relph, 1976), Goffman frames (Rettie, 2004) and the transportation imagery model (Bracken, 2005a). Finally, Slater suggests that there are many ways to evaluate presence (including physiological measures), and it should not be the case that one method of evaluation should be used to the exclusion of others. This is of course entirely true, but somewhat misleading. Currently questionnaires are the preferred method of measurement because they are better understood than the alternatives. Although questionnaires are subjective and they do have flaws, there is a large body of literature which details how to develop, asses and refine both the reliability and validity of questionnaires (Anastasi & Urbina, 1996). A thoroughly evaluated questionnaire has a known degree of error, and a known degree of ability to correctly measure the construct of interest. Physiological measures however, while having the potential for enviably small degrees of error, currently have almost no validity at all – it is simply not clear what physiological variables are related to presence, or if there exists such a thing as an identifiable means of estimating presence using physiological or neural measures (although there is also no evidence suggesting that such a thing is *not* possible). Once physiological measures have been validated (a process which is slowly underway – see Brogni *et al.*, 2003; Friedman *et al.*, 2005), they may begin to see wider use.

2.5 The relationship between concepts and measures

Although there is debate around the relative value of various definitions and measures of presence, little attention has been focused on the relationship between concepts and measures of presence. For those engaged in empirical work, this relationship is critical, as it represents the link between research hypotheses and the interpretation of evidence (Turner & Roth, 2003). The definition one picks is that about which one wants to make a finding. The wide variety of presence definitions allows one to be quite specific about what the finding will describe. However, when the study is actually run, it is the measure chosen that defines what the effect is about: a choice of BIPs as the measure effectively means one is focusing on attention and perception related aspects of presence, while a choice of the IPQ effectively means one is focusing on spatial presence. Care must therefore be taken to find definitions and measures which are compatible. The biggest concern in such a situation is nullifying the construct validity of the measure which is used. In most cases, developers of measures go to great pains to collect and report evidence for the construct validity of their measures (see Lessiter *et al.*, 2001; Vorderer *et al.*, 2004 for examples of this process in the presence field), because a measure without construct validity measures some unknown construct and is therefore useless for drawing conclusions. In practical terms, this means that the choice of presence concept and measure are not independent. When engaged in empirical work, the choice of measure is particularly important, so it may be a good strategy to narrow the choice of presence definitions by considering only those which are implemented by measures of high validity and high reliability. Fortunately, as this review shows, the range of high quality presence measures is reasonably wide, and allows one some freedom to pick a measure not only in terms of its psychometric properties, but also of the specific presence concept required.

2.6 The concept and measure of presence used in this work

For reasons of expediency, this dissertation will deal only with individual forms of presence; that is, social forms such as co-presence and social presence will be excluded. The model which will be presented will be evaluated only for individual processing situations, and although it may hold true for social forms of presence, it has not been validated for those situations. The approach of this dissertation, as the following chapters will reveal, will be strongly cognitive. At the same time, this work is highly empirical, so only concepts which are associated with highly reliable and valid measures are practical choices. First and most important, this project requires a concept and measure which are able to capture surface processing (perceptual phenomena), as well as depth processing (semantic phenomena). This is necessary to ensure that the conclusions can be made about cognitive processes rather than simply perceptual processes. It should be made clear the distinction between "perceptual" and "semantic" effects made here. Although perception involves a large degree of semantic effects (such as semantic priming, etc – see Pinker, 1995), the distinction here refers to the origin of the data being processed cognitively. Perception is a process which begins with a sensory stimulus and ends with an activated concept in working memory (Plotkin, 1998); while semantic processing largely operates on information stored in semantic or conceptual form in declarative memory (Plotkin, 1998). Therefore, perceptual effects in presence are taken in this work to mean those effects which arise from the VE, and are subject to immersion effects, while semantic effects arise in the mind of the subject, and are thus subject to experiential, but not immersion effects.

Second, as suggested by Lee's evolutionary argument and Biocca's principle of evolutionary primacy, the concept chosen should allow one to make claims about various types of media (not simply immersive media), as presence is a phenomenon which pre-dates immersive technology (Biocca, 2003; K. M. Lee, 2004). Third, in order to remain as generally applicable as possible, the only concepts of presence to be considered will be the unifying principles; and finally, the concept must be capable of being operationalized for empirical work. Using these criteria, Lee's concept is excluded, as it does not lend itself well to operationalization (this is because Lee's proposal is closer to describing a mechanism for presence than a concept of presence). From the remaining two, Lombard and Ditton's definition (1998) is preferred over Nunez and Blake's (2001) because, although the latter is more directly specified in cognitive terms, the former has achieved more widespread use, and will therefore allow better comparison to existing research, and therefore a more solid foundation.

It should be noted that by 'presence' this dissertation specifically refers to *virtual presence*, that is, telepresence for the special case where the remote site is virtual (following Sheridan, 1992). Although it is reasonable to suggest that the findings here will apply to telepresence also, none of the supporting presented here studies use real remote sites, and for the sake of parsimony no such inferences will be made. However, in line with the evolutionary arguments presence by Biocca (2003) and Lee (2004), this dissertation considers the mechanisms which lead to virtual presence can also lead to presence in the real world, and therefore presence in the real environment is not considered as a unique case.

Choosing Lombard and Ditton's definition excludes a number of notable measures. Specifically, it excludes all physiological and brain imaging techniques, as there is no expectation of what changes one would see when a subject experiences mediation or the illusion of non-mediation. It is true that an approximation to this can be reached following Meehan's (2001) or Slater *et al.*'s (1999) approach of adding to the environment features which produce a known response under conditions of non-mediation, but this unduly trades off reliability for validity as discussed in 2.4.6 above. Of the remaining measures, the BIPs technique (Slater & Steed, 2000) is excluded as it operates only at the surface level (attention and perception), and gives no indication of deeper processing. The best remaining candidates are the PQ, the MEC-SPQ and the ITC-SOPI, as these have been through the most rigorous psychometric evaluations. Of these, the PQ is excluded, as it has had the least use with non-immersive environments. From the final two (MEC-SPQ and ITC-SOPI), the ITC-SOPI is preferred, due to its longer track record in research among various independent research groups (suggesting higher validity), as well as having had a larger sample during development, which undoubtedly increased its reliability (Anastasi & Urbina, 1996).

Chapter 3: A critical review of current significant models of presence

This chapter reviews the most significant current models of presence. Four major model families are identifiable in the literature: The two-pole and environment selection models associated with Slater; the three-pole model associated with Biocca; the focus-locus-sensus and layers of presence models associated with Waterworth, and the measures, effects, conditions model associated with Wirth and colleagues. Each of these will be examined using a standard template to allow direct comparisons between models. This template includes:

1. *Description of the model* – a summary of the model structure and history.

2. *Presence in the model* – a discussion of how the model views the state of presence, and how presence exists in the model structure.

3. *Summary of empirical evidence* – a critical summary of the most important empirical work supporting the model.

4. *Critical discussion of the model* – a critical examination of the model, including comparisons to other models, the evidence in favour of and opposed to the model. This discussion will consider the plausibility of the structures, and whether the model creates contradictions.

5. *How the model explains the five problems* – An examination of the model's power to explain the five problems for presence (these are detailed in 3.1 below).

A summary table of this review can be found in Appendix G. The models are presented in chronological order of publication, but due to ongoing development, there is considerable overlap. This review constrains itself to models of presence as a phenomenon, due to this dissertation's aim of evolving a cognitive model of presence.

This excludes models of how presence is related to other constructs, such as Zeltzer's Autonomy-Interaction-Presence model (Zeltzer, 1994) or the Immersion, Presence and Performance model (Bystrom *et al.*, 1999b).

3.1 Five problems for presence theory

The study of presence presents a few phenomena which are counterintuitive and thorny for models to explain. The literature contains five such problems of interest. For the purposes of this review, these problems will be presented as phenomena free from theory or explanation. The following sections define and describe each of the five in turn, and describe their importance to presence.

3.1.1 The book problem (Biocca, 2003)

A book stimulates only a single modality, with very low fidelity and with a low bandwidth stream of information. Furthermore, the reader has almost no control over their movement or interactions within the mediated space (Biocca, 2003). A book therefore represents an extremely low immersion system, and should, according to many current presence models, produce low presence experiences (Slater *et al.*, 1996). However, presence experiences while reading books are possible (Gerrig, 1993; Nunez & Blake, 2003b; Towell & Towell, 1997). A theory of presence must therefore be able to explain why this is possible (unless such a theory were to explicitly limit itself to being a theory of presence within highly immersive environments).

3.1.2 The physical reality problem (Biocca, 2003)

This problem can be considered the opposite of the book problem. The real world presents a continuous, high bandwidth, multimodal stream of information, and allows complete control over movement and interaction in the environment. Nevertheless, people sometimes experience no presence in the real world, due to daydreaming or being lost in thought (Biocca, 2003).

3.1.3 The dream state problem (Biocca, 2003)

The final three of Biocca's proposed problems is that presented by dreaming. Dreams can result in intense presence experiences, and yet the subject is receiving almost no stimuli from the real world, and there is no mediated environment involved in the experience either. Where then are they present? As with the book problem, one can argue that what Biocca calls presence in a dream is not presence at all, but again, one would have to define presence in a narrow, media specific form to exclude this type of experience.

Another interesting problem with discussing dreams in presence is that following Lombard & Ditton's (1997) definition of presence as the 'illusion of nonmediation', dreams should not be considered, as no mediation occurs. However, if one understands "nonmediation" in a more general sense to mean the case where *perception of an environment leads to a sense that nothing intercedes between the subject and the environment*, then dreams provide a particularly interesting category of experiences. During a dream, normal perception is circumvented, and rather than having the visual (or auditory, etc.) cortices stimulated via the sensory organs as during normal perception, they are stimulated directly by the reticular activating system (Hobson *et al.*, 2000). From this perspective, a dream is like the much vaunted science fiction situation where the VE display is done by direct stimulation of the brain. It is therefore interesting to consider the reticular activating system as a 'mediating system' during dreaming, as it may give hints of how advanced media systems may lead to presence experiences.

3.1.4 The virtual stimuli problem (Nunez, 2004a)

From a physiological point of view, a person only ever experiences one stream of external stimuli. For example, all light, regardless of whether it arises from a VE display or from the sun outside the laboratory window is received by the retina in the same way. All stimuli which encode a VE are converted into physical stimuli (light, vibrations in the air, etc) in order to reach the subject and for presence to occur (Nunez, 2004a). All stimuli, regardless of origin, are in fact real; they are all just energy arriving at the sense organs. Stimuli which arise from virtual sources (which can be called "virtual stimuli") are not tagged as being virtual and belonging to a special subset of stimuli. Virtual stimuli can share physical properties which will mark them as being different from the other stimuli. For instance, all stimuli related to the virtual environment might come from a small area of space, and may thus begin processing from a small set of adjacent retinal cells (Craver-Lemley & Reeves, 1992), or the virtual stimuli might be of a higher average intensity (Jonides, 1981). However, these properties (spatial location, relative intensity, etc.) must be inferred during perception, and can thus only be grouped together after they have been partly processed (Ungerleider & Haxby, 1994). It is therefore incorrect to state that there is a discrete number of environments or stimulus sets for a subject to choose from (such a "real environment" and "virtual environment"). There exists only one stream of stimuli from which a user can infer any number of environments. Because this process is inferential, top-down effects will play a major role (Wirth *et al.*, 2007). The problem of virtual stimuli is thus: How are certain stimuli recognized as encoding one coherent virtual environment, and how are stimuli outside of this set excluded from the presence experience?

3.1.5 The inverse presence problem (Timmins & Lombard, 2005)

Inverse presence was first described by Timmins and Lombard (2005). It occurs when real events are experienced as if they were mediated. This is most likely to happen when the events are unusual or emotionally intense, such as during the perception of great beauty or when being the victim of a crime. It is not clear how common inverse presence experiences are (the interview method used by Timmins and Lombard does not allow that inference), but the documented existence of 97 cases (Timmins & Lombard, 2005) suggests it needs to be considered by presence theory.

One can argue that inverse presence need not be explained by a theory of presence, as it is not experienced during mediation. Timmins and Lombard (2005) argue that inverse presence involves one class of experience (real) being confused for another (mediated), which is, according to Lombard and Ditton's (1998) definition of presence, the essence of the presence experience. From a psychological perspective, it is a very interesting phenomenon, as it indicates that the experience of mediation is not specifically tied to a particular class of stimuli, but can be freely associated with other sets of stimuli. Uncovering the conditions under which this sense of mediation is activated might shed light on how the converse (a feeling of non-mediation) occurs.

3.2 The five problems as a yardstick of model power

If one agrees that the five problems present important phenomena in presence, then one can judge the relative value of a presence model by how well it explains the five problems. In this review, each of the current major models will be compared in terms of their response to all five problems, to gain a comparative benchmark of their power, and to illustrate their strengths and shortcomings. This approach has been used in a limited way by Biocca, who used the book, physical reality and dream state problems as a measure of the increase of power of the three-pole over the two-pole model (Biocca, 2003), and by Waterworth who used the book and dream states problem to show the relative power of the focus-locus-sensus model of presence (Waterworth & Waterworth, 2001).

3.3 Current significant models in the presence literature

3.3.1 The two-pole / environment selection model

3.3.1.1 Description of the model

Although presented as a unified model, this is actually a composite review of a number of separate ideas, which do not formally exist as a model in the literature. The term 'two-pole model' was coined by Biocca (2003), and the term 'environment selection' was used by Slater and Steed (2000) to refer to the same family of concepts (this review prefers the latter term, as it better reflects the current sophistication of this model). This model is an evolution of the classic telepresence model, where an operator experiences being present at the remote worksite (Biocca, 2003). Sheridan (1992a) suggested that virtual presence could be understood simply as telepresence for the special case where the remote worksite is virtual rather than real. This idea was extremely persuasive, and led much of the research during the 1990s. This review will not focus on that early work, but rather on the more recent developments of that concept which have been informed by considerable empirical evidence.

Early versions – the "two-pole" model

In the "two-pole" model (a term coined by Biocca, 2003) subjects exist in one of two states (or poles) – either present in the virtual environment, or present in reality (see Figure 3.1 below). There is debate as to whether presence occurs by degrees (e.g. Wirth *et al.*, 2007; Witmer *et al.*, 2005), or whether it is binary (Slater, 2002) - see 2.3 in chapter 2. Some confusion surrounds Slater's binary position, as his questionnaire (the SUS - Slater *et al.*, 1994), provides a continuous score. However, it should be noted that Slater's practice when administering the SUS was to quantify scores such that only those scoring 6 or 7 would be considered as 'present', while the others would be considered as 'not present' (Slater *et al.*, 1994). To add to the confusion, Slater admits that continuous presence may indeed exist (Slater, 2002). Nonetheless, the two-pole model generally works for either binary or continuous concepts of presence, because at its core the model posits a simple one-dimensional progression between 'present' and 'not present' in the VE (Biocca, 2003). Generally speaking, 'not present' is not well defined, although having another system or environment which interferes with the VE of interest is central, as this model views presence as comparative (Slater, 2003a). The interesting questions arising are how a subject moves from the 'not present' to the 'present' pole (see Sadowski & Stanney, 2002), and what factors interrupt presence (Slater & Steed, 2000). A major aim of this research is to identify factors which affect and mediate presence, such that a 'presence equation' of factors and their relative contributions can be constructed (Kalawsky, 2000; Sas & O'Hare, 2001). These factors are categorized as being internal or external (Sadowski & Stanney, 2002); Internal factors are associated with the subject, for example culture (Fontaine, 1992), a tendency to become immersed (Witmer & Singer, 1998), or age and personality (Heeter, 1992). External factors (which are sometimes referred to as immersion factors - Slater *et al.*, 1995b) include display field of view (Hendrix & Barfield, 1996a), pictorial fidelity (Welch *et al.*, 1996), scene detail (Slater & Wilbur, 1997) and display resolution (Bracken & Skalski, 2006).

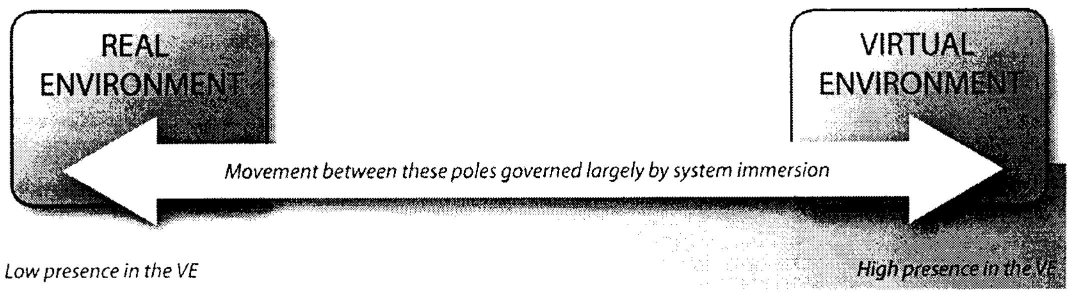

Figure 3.1: The two-pole model. The subject moves between the "real environment" pole ("not present") to the "virtual environment" pole ("present"). The dynamic shift is determined largely by the degree of immersion of the system, although mediation by internal factors is thought to play a small role.

In many ways, the two-pole model (particularly the early versions) is not a model of presence as a phenomenon, but of presence as a desirable outcome for interactive systems. The dichotomy between 'present in the virtual environment' and 'present somewhere else' (often appearing as 'present in the *real* environment') seems to develop naturally from this perspective – did the system manage to produce the desired effect (making users present) or not? And from that question, quite naturally, follow specific questions such as "what can be done to increase the likelihood of the desired result?" The current version of the two-pole model, which is better referred to as the *environment selection model*, is more sophisticated.

Current status - Environment selection theory

Environment selection theory assumes that subjects can only respond to and act in a single environment, even though they may be presented with several (Slater & Steed, 2000). This limit is imposed by their direction of gaze, limits of attention allocation, and other inherent factors (Slater & Steed, 2000). While a subject is present in an environment, it is perceived as a coherent whole (Slater and Steed use the term *Gestalt* to describe this coherence). To be present is thus to have selected one particular environment to respond to from among all competing environments (Slater & Steed, 2000). According to this model (outlined in Figure 3.2 below), a subject in a VE receives a continuous data stream from the VE, but also from other sources in the real environment (noises outside, temperature changes, etc) and from rendering errors in the VE system (Slater & Steed, 2000). Subjects in a VE thus always face two environments, and must choose one in which to be present. It is not clear whether only bottom-up data is able to force the selection of one environment over the other, or if top-down data has a role in this process (although this would seem a natural place for volitional processes such as 'the suspension of disbelief' – Slater & Usoh, 1993b).

This simple but convincing model has been subsequently refined by Slater (2002) by incorporating top-down and expectation based processing. As subjects interact in the VE, they form hypotheses about the VE (expectations for future data), and sensory inputs are tested under this expectation. If data which is consistent with the expectation arrives through other sensory channels (such as the addition of haptic feedback used by Meehan *et al.*, 2002), the VE becomes supported as a viable hypothesis. However, conflicting data (such as the tug of a cable or a rendering glitch), can cause the real environment to be selected, resulting in a break in presence (Slater, 2002).

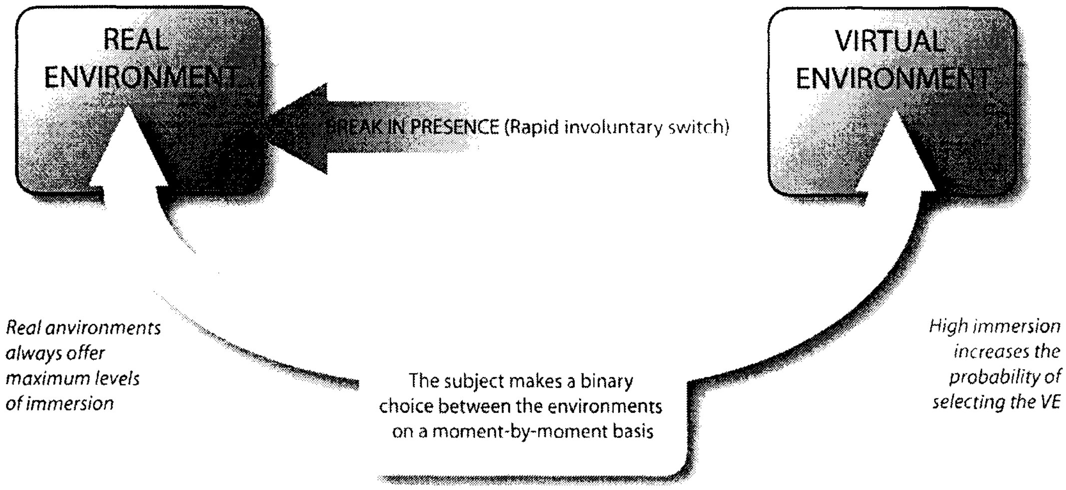

Figure 3.2: The environment selection model. The subject is either present in the real environment, or in the virtual environment. The probability of the subject selecting the VE depends largely on the level of immersion. Should a contradictory stimulus occur (such as a rendering artifact), the subject will rapidly switch back to being present in the real environment – a 'break in presence'.

The relationship between the amount conflicting data and the probability of a break in presence is not linear. Brogni *et al.* (2003) suggest that conflicting information could be attended out, incorporated into the current hypothesis, or lead to a break in presence. It is also not clear whether this will occur based only on the amount of conflicting sensory information, or if the content of the information has an effect. Depending on which environment which is selected, the subject will act differently; this allows the evaluation of presence by observing behaviour.

The process by which subjects switch between environments seems to be automatic, as There is no discussion of the possibility that a subject may willingly switch from one environment to another. In the reversible figure illusions which Slater uses to illustrate the Gestalt-like elements of the model (such as the young woman/old crone illusion and the duck/rabbit illusion – Slater, 2002; Slater & Steed, 2000) subjects can easily switch from one interpretation of the image to the other at will once they become aware that two alternatives are possible (Girgus *et al.*, 1977). However, this is probably not the case with presence. It seems that it is easier to stop being present in a

VE that it is to begin (although no empirical evidence to this effect exists). If this is the case, then it would be interesting to note the differences are between the reversible figure illusions and presence. It may be that, unlike reversible figures, there is little ambiguity in the environmental stimuli processed by the subject. The classic reversible figures are carefully constructed so that each line in the image is evidence supporting both hypotheses equally, and the small amount of data required to tip the decision in favour of one interpretation can be supplied by the subject willingly. In the presence situation however, it may be that the sensory stimuli tend to favour the "I am in the real world" hypothesis far more strongly than the alternative "I am in the VE" hypothesis; it may be impossible to provide the amount of top-down evidence required by will alone. This is because the suspension of disbelief argument assumes that all the relevant cognitive processes are mutable by the exertion of will; however, even for the case where a subject is able to exert a heavy degree of suspension of disbelief, the existence of automatic processing structures in the cognitive system (such as the automatic activation of the folk-psychology module – Baron-Cohen, 1995 – and the automatic association of active concepts in declarative memory – Atkinson & Shifrin 1968) may in fact prevent this from occurring, as some processes simply may not have any input from volition.

3.3.1.2 Presence in the model

Both forms of the model discussed above follow a consistent view of presence: Presence is the subjective sense that one is in the VE of interest, rather than in any other place (Slater *et al.*, 1994). Essentially, a subject exists on a continuum between the two poles of "virtual environment" or "real environment" (Biocca, 2003). The more recent reformulation by Slater and Steed (Slater & Steed, 2000) removes the emphasis of being present in either the virtual environment of interest or some other specific place, to a more generic sense that the subject is either present in the environment of interest, or in some other place, which is not defined. This improves the definition in some sense, as it simplifies operationalization; one need no longer worry about where in particular the subject felt present, in some sense obviating the need for a "third pole" as proposed by Biocca (2003) – see 3.3.2 below.

3.3.1.3 Summary of empirical evidence

Evidence for the two-pole model

The two-pole model simply predicts that particular factors (especially immersion factors) will increase the probability of the subject feeling present in the environment of interest rather than anywhere else. It is perhaps not so much a model as a loose framework upon which a program of raw empiricism has been based (see Kalawsky, 2000; Sas & O'Hare, 2001 for a discussion of the structure of this program). Much work has been done into uncovering factors important to presence, and much of it has been replicated. As discussed in 3.3.1.1 above, much of the impetus in this research is driven by practical concerns of creating "presence producing systems".

This review will not consider studies which take into account internal factors (that is, subject related factors), for two reasons: First, there exists no psychological or physiological model of how the individual factors interact with immersion factors in the tradition of the two-pole model, and therefore no firm predictions related to internal factors (Biocca, 2002; Sheridan, 1992a). Second, studies which consider internal factors almost always have a more complex model to test, and are therefore are better discussed in the context of those models (see further discussions in 3.3.2, 3.3.3 and 3.3.4 below). Due to the large volume of studies in this area, this review will not go into detail, but rather categorize them by type of factor investigated.

Graphical display parameters

The largest body of work has considered graphical display parameters. Generally, these studies have a simple experimental design, where the factor of interest is manipulated, and its effect on presence is measured using a self-report scale (and in a few cases with behavioral observation or by counting breaks in presence). The general finding is that increasing the quality of the display by making its artifacts and limitations imperceptible increases presence (Lombard & Ditton, 1997). Some of the factors found to have an effect are:

- **Animation**. Animated scenes (where objects which would be animated in reality are animated in the VE) lead to more presence (Cho *et al.*, 2003)
- **Colour depth.** More colour depth leads to more presence (Barfield & Weghorst, 1993)

- **Display type**. Reality almost always out-performs any type of display (Allen & Singer, 2001; Hullfish, 1996; Mania & Chalmers, 2001) with a few rare exceptions (Usoh *et al.*, 2000). Multi-wall cave systems produce more presence than monitors (Axelsson *et al.*, 2001; Schroeder *et al.*, 2001). Head mounted displays generally produce more presence than monitors (Nichols *et al.*, 2000; Slater *et al.*, 1996; Slater *et al.*, 2000; Youngblut & Perrin, 2002), but this seems a weak effect, as numerous studies have failed to replicate the effect, perhaps due to the weight and discomfort associated with wearing such a display (Slater *et al.*, 1995a; Slater *et al.*, 1999; Youngblut & Perrin, 2002).
- **Display size**. Larger displays tend to lead to higher levels of presence (Lombard *et al.*, 2000a; IJsselsteijn *et al.*, 2001; Bracken & Botta, 2002; Bracken, 2005)
- **Display update rate.** Faster updates lead to more presence (Barfield *et al.*, 1998; Barfield & Hendrix, 1995; Meehan *et al.*, 2003; Snow, 1996). Note that this effect is likely non-linear, as Meehan (2002) found no difference between update rates of 10 and 15 Hertz.
- **Geometric field of view**. Wider displays lead to more presence (Hendrix & Barfield, 1996a; Prothero & Hoffman, 1995) although within limits - Allen and Singer (2001) showed maximal presence when using a natural FOV.
- **Level of detail**. More detailed, realistic scenes lead to more presence (Cho *et al.*, 2003; Shim & Kim, 2001; Slater *et al.*, 1995c; Welch *et al.*, 1996), although this effect may be weak, as other studies have failed to replicate it (Dinh *et al.*, 1999; Snow, 1996)
- **Resolution**. Higher resolution leads to more presence (Snow, 1996)
- **Stereopsis**. Stereo-enabled displays generally produce more presence than mono displays (Cho *et al.*, 2003; Hendrix & Barfield, 1996a; Snow, 1996)
- **Texture mapping**. Texture mapped scenes lead to more presence (Cho *et al.*, 2003; Snow, 1996)

Multimodality

A second clear finding is that systems which stimulate multiple modalities simultaneously lead to more presence that systems which stimulate a single modality. The following modalities have been examined:

- **Audio.** With very few exceptions, the addition of audio increases presence (Darken *et al.*, 1999; Dinh *et al.*, 1999; Hendrix & Barfield, 1996a, , 1996b; Nichols *et al.*, 2000; Sallnäs, 1999; Snow, 1996; Welch *et al.*, 1996).
- **Haptics.** Several studies have found that the addition of haptics or touch cues increases presence (Dinh *et al.*, 1999; Meehan, 2001; Sallnäs, 1999). However, a number of studies have failed to find an effect, suggesting that this is a weak factor (Insko, 2001; Lok *et al.*, 2003; Meehan, 2001)
- **Olfactory.** This factor has not received much attention (perhaps due to the engineering difficulties associated with implementing an olfactory renderer). The results are mixed – one study (Dinh *et al.*, 1999) found no increase in presence, but another (Hoffman *et al.*, 1999) found a small gain. Due to the small number of studies available, it is very difficult to draw a conclusion on this factor at this time.
- **Proprioception.** This modality can be implemented in immersive systems by the use of body tracking. This seems to be a strong effect, as it is replicated in almost all studies (Bystrom & Barfield, 1999; Hendrix & Barfield, 1996a; Snow, 1996)

System interface and interactivity

The two-pole model predicts that factors which provide cues to the subject that they are using a VR system could reduce or interrupt presence (Usoh *et al.*, 1999). Although VE interfaces are often examined as invariant system factors, the subject's proficiency with the interface will probably interact with the interface type. The following interface related factors have been examined:

- **Interactivity.** The more possibilities for interaction provided by the system, the more presence it generates (Snow, 1996); also, active roles in the VE lead to more presence than passive roles (Larsson *et al.*, 2001; Preston, 1998).
- **Movement.** Moving in the VE produces more presence than being stationary (Cho *et al.*, 2003); although this effect may be due to increased interactivity rather than increased navigation. The more natural the method of movement, the more presence reported by subjects; real walking generates more presence than passive motion or mouse control (Slater *et al.*, 1995a; Usoh *et al.*, 1999; Witmer & Singer, 1998).

Evidence for the environment selection model

The central tenet of this model is that the subject selects between environments to be present in; a corollary is that when a change in that selection occurs, it is experienced as a break in presence. Strictly speaking, there is no empirical evidence that subjects do select between environments, but there is evidence to show that certain distractions, particularly those associated with stimuli outside the VE, do lead to breaks in presence, which can be reported, and are associated with self-reports of presence. The first study to show this used SUS scores to predict the reported breaks in presence (Slater & Steed, 2000). Another similar study found a negative correlation between number of breaks in presence and SUS scores obtained during six separate immersive VE experiences (Brogni *et al.*, 2003). Finally, Vinayagamoorthy *et al.* (2004) also found a negative slope when regressing number of breaks in presence on presence questionnaire data. It is important to note that these studies do not show that environment selection takes place; it is evidence that the end of presence is a reportable experience. One can argue that the fact that during post-VE interviews some subjects did report experience a sensation of "switching" between environments supports the switch. However, these reports should be considered contaminated by the instructions given to subjects on how to report a break in presence (see 2.4.5 in chapter 2). Nevertheless, this evidence strongly supports the notion that the number of interruptions during the VE experience can inhibit the subject's presence.

3.3.1.4 Critical discussion of the model

The two-pole model has already been thoroughly discussed and criticized in the literature; in particular Biocca (2003) has outlined important weaknesses in that model (see 3.3.2.1 below). These criticisms revolves around the central assumption that a subject must either be present in the virtual environment of interest, or in some other environment. Biocca argues the one can be present in *no physical environment*, such as when one is lost in one's thoughts (Biocca, 2003). Furthermore, because the two-pole model emphasizes the immersion-presence relationship, it does not allow for subjects becoming present in non-immersive media such as books (Biocca, 2003). These two criticisms are correct, but they cannot overcome the fact that the two-pole model has more supporting evidence for its central notion than any other presence model currently available; it is almost impossible to argue against the immersion-presence relationship.

However, it is important to consider this evidence within its limits. None of the available evidence shows that immersion is either necessary or sufficient for presence; all it shows is that one path to presence (among an unknown number of paths) is through immersion. A case in point is the role of content factors in presence. Almost all the studies reviewed in 3.3.1.3 above (indeed, most of the studies reviewed in this chapter) ignore VE content as unimportant to presence. From a purely methodological point of view, this is correct, as content is usually held constant across conditions. But this is not the case when comparing *across* studies, which may have vastly different content. Slater has explicitly stated that content is not an important factor in presence (2003a), based on evidence such as that cited in 3.3.1.3 above. However, in order to make such a claim, a similar body of evidence would have to show that non-immersion factors (such as content) have no effect on presence. The lack of evidence for content effects simply reflects scarcity of studies, not lack of effects.

Although the environment selection model is derived from the two-pole model, it is different enough to warrant an examination in its own right. The central notion in this model (that subjects select between competing environments as Gestalts), is supported by analogy using the reversible figures illusion (Slater, 2002). These illusions work because each line in the figure can simultaneously support one of two interpretations (e.g. duck or rabbit), such that the figure is completely ambiguous (Slater, 2002; Slater & Steed, 2000). However, when placed in a VE in a laboratory, subjects need to deal with a vast array or sensory information of varying degrees of intensity which they must form into a Gestalt. The subjects must select relevant stimuli while attending out the rest, based on their significance and task demands (Nunez, 2004a; Wirth *et al.*, 2007). Unlike a reversible figure, such a situation has no finite set of alternatives. Each subject constructs their situation in terms of its importance to themselves at that moment. A more fair analogy would be to consider the reversible figure as a picture on being on a piece of paper in a room. Some observers might see the duck, some might see the rabbit; but some, who may not be paying attention to the task, may only see the piece of paper and the experimenter; others may only see the décor in the room, and so on. It is true that all of these subjects are selecting between alternatives; but it is not the case that the number of alternatives is bounded by the stimuli manipulated by the experimenter.

The environment selection model also fails to explain why it is far easier for subjects to be present in the real world than in a VE (see for example Usoh *et al.*, 1999). If there is a selection being made, what factors lead to one environment consistently being chosen over another? One possible answer is immersion factors. The real world has higher resolution, more detail and stimulates more senses than any virtual reality system current available; therefore, it is selected more often. This is plausible, but it obviates the very model it is being used to support – it simply returns to the two-pole model ("more immersion means more presence") but with the added constraint that presence is now binary (the virtual environment of interest is either selected or not). Given that there is no theoretical position that categorically states that presence is binary (Slater himself stating that it might be continuous – Slater, 2002), this seems an untenable theoretical position. Nevertheless, the basic notion that some selection is happening during presence is interesting, because of the elegant way in which it explains the role of attention in presence (Biocca, 2003; Waterworth & Waterworth, 2001; Wirth *et al.*, 2007), and because of the break in presence experience (Slater & Steed, 2000).

3.3.1.5 How the model explains the five problems

The book, physical reality and dream state problems were defined as a reaction to the two-pole model (Biocca, 2003), so one can expect that these phenomena will not be explained well. Biocca has already discussed these three problems with regards to the two-pole model in detail (see 3.1 above), but not with regards to the environment selection model.

The book problem

At first glance, the environment selection model seems a good candidate for explaining the book problem. When reading, a subject has competing environments to select from (the environment in the book and the real world), and the subject can choose to read the book or simply look at it. However, this model still relies on the basic notion that immersion is required for presence, even if there is some recognition that cognition mediates the process (Slater *et al.*, 1994). From this perspective, information presented in non-immersive and non-embodied forms is extremely unlikely to lead to presence (Slater, 2003a), and this effectively precludes books from producing presence experiences, although reading a book can still lead to psychological engagement and enjoyment (Slater, 2003a).

The physical reality problem

In this situation, the subject is not processing external stimuli – they are lost in their own thoughts or preoccupations (Biocca, 2003). Again, the environment selection model seems a likely candidate to deal with this phenomenon: does the subject select the environment in their imagination, or the real environment around them? Due to the model's focus on immersion (which concerns only external stimuli), it is again difficult to explain this phenomenon. The model is not well equipped to deal with this problem due to its lack of an explanation as to what happens when someone attends out external stimuli (Biocca, 2003). It seems clear that when a subject experiences an environment, they process it into mental representations, some of which are mental images (Kosslyn & Thompson, 2003). Could bringing such images willingly to mind not lead to a similar (if impoverished) sense of being in the space? The environment selection model cannot respond to such questions as it lacks a coherent notion of what partial or continuous presence is and how it arises.

The dream state problem

Biocca (2003) presents this problem as similar to the physical reality problem, because at its core is the issue of presence in imagined environments. During dreaming, all external sensory stimuli are attended out (or if not, they are generally incorporated into the dream), and replaced by internally generated stimuli, some of which emulate bottom-up information (Hobson *et al.*, 2000). Although this situation can be considered a high-immersion situation (sensory stimuli have been replaced by the "virtual environment" of the dream), generally only a few modalities are stimulated, and dreams often contain a number of logical and perceptual inconsistencies. This makes it difficult to explain with the two-pole model (as discussed by Biocca, 2003). However, the environment selection model is well capable of dealing with dreams. Notice that in a dream, there is in fact only one environment (the dream), as all competing environments have been attended out. Therefore, even with the low levels of immersion generally found in dreams, the model has no problem explaining how one can feel present – the dream *must* be selected as the environment to be present in as it is the only choice available. Of course, this assumes that certain minimum criteria are met, for instance that the dream is in fact about a place and the subject experiences the space from the perspective of someone occupying that space.

The virtual stimuli problem

These models are not able to deal with the virtual stimuli problem due to their central assumption that presence is the selection between competing environments. Recall that to perceive an environment, the subject must first construct that environment as a coherent cognitive entity by selecting a particular subset of stimuli from the undifferentiated mass of stimuli arriving at the senses (Nunez, 2004b). Due to the limitations of human cognition, only a small subset of stimuli can be processed (Baddeley, 1986). Proposing that subjects are able to simultaneously construct and maintain several environments to select from violates this principle. One may counter by arguing that the model is not cognitive, but rather descriptive; an outside observer can enumerate several possible environments which can be constructed from the available stimuli, and then interpret the subject's behaviour as a choice between those environments. This is a true, but it fails to consider that subjects construct environment from both bottom-up and top-down data (Nunez, 2004a; Slater, 2002; Wirth *et al.*, 2007). Therefore, it may not possible for an external observer to predict or even describe the environment which the subject is experiencing presence in. Evidence for the importance of this comes from Nowak *et al.* (2006), who found that the presence in violent games was mediated by the degree of *perceived* violence in the game. To an outside observer, a game has a constant degree of violence; however, due to individual differences, subjects may construct the environment as being more or less violent, which in turn affects their presence experience. The environment selection model could show that the subject is present in a violent game by examining their behaviour; but the detail required to differentiate between two subjects who perceive different degrees of violence and therefore have different presence experiences could not be achieved by this model.

The inverse presence problem

The inverse problem arises when a subject mistakes the real scene for a mediated one (Timmins & Lombard, 2005). The classic two-pole model is not able to explain this phenomenon, as it defines a strong distinction between "real" and "virtual" in terms of immersion. Presence arises (almost automatically) as a function of having sufficient immersion. The two-pole model makes the sensation of being in the real world the standard against which less immersive mediated experiences are compared. It is therefore almost impossible to understand how a completely immersive situation (such as the real world) could lead to a sense of less presence. The environment selection model fares little better. In this model, one environment is selected from competing environments by the subject to feel present in. However, in the situations of inverse-presence presented by Timmins and Lombard (2005), there is usually only one environment available. The environment selection model predicts that at worst, a low immersion environment could be selected for presence, and in such a case, the presence experience would be low; but in its current form it cannot explain why a subject should experience a highly immersive environment which they have selected as a low presence experience.

3.3.2 Three-pole model

3.3.2.1 Description of the model

This psychological model sees presence as moving in a space defined by three idealized poles (see Figure 3.3). These poles represent complete presence in a physical space, complete presence in a virtual space, and complete presence in a mental imagery space. The model contains no explicit notions of immersion or display technology. In fact, such concepts have been removed from the model for two reasons:

1. Biocca argues that the inclusion of system and immersion variables is not relevant to explaining psychological states such as presence. The idea that immersion leads to presence (the 'sensorimotor immersion assumption' in Biocca's terms) was dictated by engineering expediencies rather than psychological theory (Biocca, 2003). For researchers working with an engineering hammer, presence naturally seemed like an immersion nail. Biocca argues that a general model of presence for use in various media must reconsider the role of technology in presence, rather than assuming it as a necessary condition.

2. Presence must have existed before VEs, as the psychological mechanisms involved must be evolved (the 'evolutionary primacy' principle - Biocca, 2003). The certain media lead to presence is an indication that something in those media capitalizes on particular aspects of cognition (as perceptual illusions do - Slater, 2002). It then follows that explanations of presence should be independent of media, and conversely, that any medium could potentially induce presence. It should therefore be the psychological mechanisms involved in presence which should take center stage in a presence theory, not the display (Biocca, 2003).

The three-pole model is essentially an elaboration of the two-pole model. Each of the three poles (physical, virtual and mental imagery) represents stimulus sources which can lead to presence in that space. For example, attention focused on a display encoding a VE will lead to a high degree of presence. As with the environment selection model, having attention divided between poles leads to reduced presence. These stimulus sources dynamically change and possibly compete with each other, causing presence to be an oscillating phenomenon. Here an important difference exists between this model and the environment selection model: Biocca (2003) specifically allows the possibility that cues from the three sources could interact or be integrated into each other to form a coherent presence experience (as opposed to the environment selection model that sees the two stimulus sources always interfering with each other).

The relative contribution of each pole to presence is controlled by two cognitive processes:

1. *Spatial attention:* According environmental changes and task demands, attention will shift between the three poles during the experience. A loud noise, for instance, will demand attention to itself, which will change the relative contribution of the three poles; or a difficult spatial task may lead to attention being shifted towards mental imagery space during a portion of the experience.

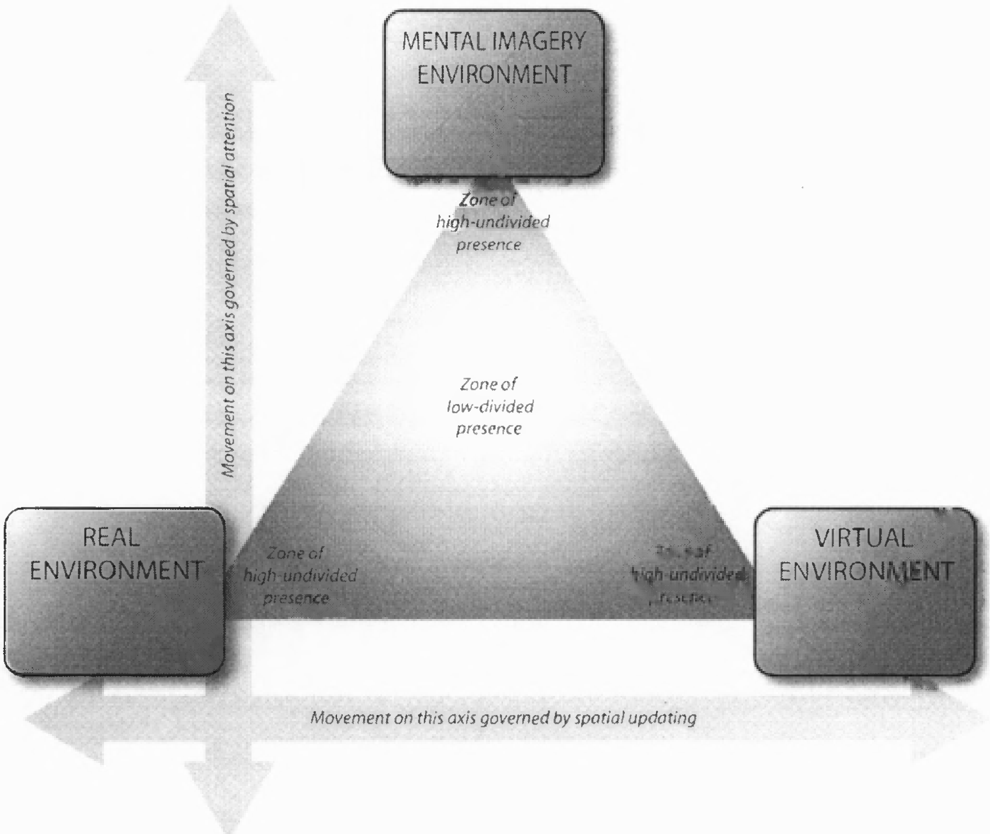

Figure 3.3: The three-pole model. The subject's experience is always located somewhere inside the green triangle; closer to the poles (real, virtual and mental imagery environments), the experience is *undivided presence*, while at the center is a *divided presence* experience. The experience moves dynamically in this space according to changes in spatial updating and spatial attention (tan arrows).

2. *Spatial updating:* As the experience progresses, the subject's model of the space and their relation to it is updated (Biocca, 2003). Such updates can lead to changes in the relative contribution of the three poles. This process is associated with bodily movement (either in terms of moving sensors within the space or affecting the space with the body), and as such is theorized to affect mostly shifts between the virtual and physical spaces.

3.3.2.2 Presence in the model

Presence is conceptualized as the processing of the cues from the three poles, with the weight of each pole being determined by spatial attention and spatial updating. Presence is thus a highly dynamic process, as the relative contributions change constantly, making presence continuous rather than binary. Subjects who have a high contribution from one of the poles experience 'undivided presence', while those who find themselves in environments where more than one pole is making a significant contribution would be said to be experiencing 'divided presence'. The undivided form of presence is considered to be 'high presence', while the divided form is considered to be 'low presence'. This model therefore considers focused, exclusive attention on one set of cues as a fundamental requirement to presence.

It is not clear how the transition happens between focusing on a stimulus set and presence in that set occurring. Although it is clear that focused attention is necessary, it is not obvious that it is sufficient. Strictly speaking, the three-pole model is not a model of presence *per se*, but rather a model of presence shifts. The model does not really need to explain how presence comes about in order to explain how subjects move from being present in a VE to being present in a real environment. If this is indeed the case, then presence exists as an ephemeral substance moved by the winds of spatial attention and spatial updating between the three poles.

3.3.2.3 Summary of empirical evidence

There has not been much empirical validation for this model. However, as it is an extension of the environment selection model, one can consider a certain subset of the evidence for that model as valid here. Specifically, all evidence which points to presence being a phenomenon which shifts continuously and dynamically between two cue sources can be applied, as can the subset of studies which finds evidence for the role of focused attention in presence (see the summary of the evidence for the environment selection model in 3.3.1.3 above for this work).

The question of whether mental imagery spaces can lead to presence, and how this can lead to presence in other types of spaces was the subject of a study by Baños *et al.* (2005). Subjects were placed into either a VR condition, where they experienced a park using a desktop VR system, or an imagery condition, where they were asked to imagine the park. Both groups were measured using the SUS at the beginning, middle and end of the experience. Averaged over thee measurements, there was no difference in presence between the VR and imagery condition, supporting the three-pole model's concept of presence in mental imagery spaces. However, when considering how presence scores changed over time, then an effect became evident: the VR group scores increased over the experience, whereas the imagery group scores decreased (Baños *et al.*, 2005). This suggests that while the model is correct about the existence of the third pole, it may not be equivalent to the other poles in terms of maintaining focus.

There is some interesting indirect evidence for the importance of spatial updating on presence. Barfield and Hendrix (1995) had subjects perform a visual search task in a VE where the simulation update rate was manipulated. Subjects in the high refresh rate conditions reported more presence than those in the low refresh rate conditions. Spatial updating can account for this difference. As subjects move through the VE, they update their mental models of the space; however, if there is not timely feedback, then spatial updating would be inhibited, as the subject is placed in a moment of doubt as to the outcome of their input. Conversely, high update rates would give almost instantaneous feedback, allowing subjects to keep their mental models synchronized with the simulation. Another study which can be interpreted as evidence for the importance spatial updating is that by Slater *et al.* (1995c). In this study, subjects performed several spatial tasks either with or without shadows rendered in the VE. Subjects in the shadows condition reported higher presence than those without (Slater *et al.*, 1995c). A shadow provides additional information about the environment (such as position of light sources, geometry of the surface on which the shadow falls, etc.) and the motion of objects relative to each other (Kersten *et al.*, 1997). Shadows in the environment would therefore allow easier updating of the subject's mental model by providing more information about changes in the space.

3.3.2.4 Critical discussion of the model

The stated purpose of this model is to deal with three particular problems in presence theory (the book, physical reality and dream state problems), and it does this very well. If one accepts that these are important problems for presence, then one can claim that the three-pole model is a useful evolution of the environment selection model. This model can thus claim almost all evidence which supports the environment selection model, and simultaneously explain Biocca's three problems, which is an impressive feat. However, the model does have some weaknesses which are worth mentioning.

The first problem is shared with the two-pole model. Presence moves dynamically between three poles, such that the sum of the contributions of the poles is constant (or, stated differently, the three poles trade off against each other). This implies that there is a finite resource available for this process. However, the model does not state why such a limitation should exist, or what the limited resource is. One can infer that the limitation is tied to spatial attention and spatial updating. For instance, attention is known to be capacity limited (for example, see Baddeley, 1986). It would be useful if the model explicitly defined the limitation, as that would improve predictions of how spatial attention and spatial updating act to move the presence between the poles.

As most presence models do, the three-pole model gives primacy to the role of attention in presence. However, this model places particular emphasis on *spatial* attention, and the selection of stimuli in terms of their position in space. It is true that the spatial origin of a stimulus can change the focus of attention (Vecera & Rizzo, 2003), and that stimuli can be aggressively filtered out based on their spatial location (Hopf *et al.*, 2006). However, attention can shift due to a number of low level factors such as stimulus intensity (Lu & Itti, 2006) as well as high level factors such as priming (Maxfield, 1997). For example, breaks in presence, which are effectively rapid shifts between two of the poles, are often triggered by changes in attention which arise from non-spatial sources (such as rendering errors or inconsistencies in the simulation - Slater & Steed, 2000). It seems then that the role of non-spatial attention may have been underestimated by the model.

Another problem associated with emphasizing spatial attention lies in considering how presence in a physical space ends. The model defines the most intense presence experiences as occurring when subjects focus all their attention on a single pole As movement between the poles is partly determined by spatial attention, this implies that if a stimulus comes from an unexpected location, it will lead to presence being drawn away towards it. When a subject is using a VE, it is simple to define what an 'unexpected location' is (outside of the experimental room, for example - Slater & Steed, 2000). However, as physical space, by definition, occupies all space, there can be no notion of an unexpected location for a stimulus to occur from. Consider this thought experiment: I am sitting, present in real a forest. A laptop is running somewhere behind me without my being aware of it, which is showing a virtual office. Suddenly, a virtual telephone rings in the laptop – all I hear is a phone ring behind me. I am likely to think it is a real phone behind me, and on turning to look, be surprised to find it is a virtual phone. Now consider another situation: I am in the office, but watching a laptop showing a virtual forest. I hear a bird sing behind me. I am probably more likely to think that it is a real bird outside my office window than a virtual bird in the VE. Even though in both scenarios the stimulus comes from an unexpected location, and is of a type which is unexpected for the environment, it is likely easier for attention to shift away from the virtual space towards the real than vice-versa. This is partly because the primary expectation is that all stimuli come from the physical space; after all, attention (spatial or otherwise) exists is to direct cognition to some stimuli in the outside world in order to alert us to sudden changes in the environment which may represent danger or other interesting events (Sperber & Hirschfeld, 1999). The three-pole model however, posits that physical, virtual and mental imagery spaces as equally important in terms of attention, which may not be the case.

The three pole structure of the model also raises some questions. It is clear that there is a set of cues from physical space which can lead to presence; Also, there is some evidence that a set of mental imagery cues that can produce presence (for example Baños *et al.*, 2005). However, it is not entirely obvious that there exists such a distinct entity as a virtual space, comparable to either physical or imagery spaces. This returns to the problem of virtual stimuli. If a VE can only be processed after its existence has been inferred from stimuli originating in the physical space, then the physical space and the virtual space are not independent poles, as proposed by the model.

3.3.2.5 How the model explains the five problems

The book problem

This model was constructed as an explicit solution to the book, physical reality and dream state problems, so it should be no surprise that these three problems are dealt with rather well. The book problem arises from the fact that subjects have presence experiences from low sensorimotor immersion environments. The solution exists in the mental imagery pole. Reading a book involves many cognitive processes including the creation of mental models of the space described in the book. Spatial attention is drawn to these mental models, and spatial updating is performed on the models, which leads to presence being dawn towards the mental imagery pole. Provided spatial attention is not drawn away from the imagery space pole, and that the space is successfully updated, the subject can sustain their presence in the book. However, because reading involves a balance between attending to the mental imagery space created by the text, and to the visual task of reading, the presence experience will usually be in a zone of low-divided presence (Biocca, 2003). This explains why reading usually produces a less intense presence experience than more immersive media (Nunez & Blake, 2003b).

The physical reality problem

Essentially, this problem asks how subjects in the physical world (a high sensorimotor immersion experience) can sometimes not be present. Again, the solution comes from a shift from the physical space pole to the mental imagery pole. If spatial attention is shifted away from the physical space to the mental imagery space (by daydreaming, for instance), then presence in the physical space will diminish. In this model, all three poles are equivalent, so such a shift is consistent with the model. If some task is being performed in the physical space, then some attention will still be devoted to processing it, and the result will be a shift into a zone of low-divided presence (Biocca, 2003).

The dream state problem

The dream state problem is seen as an extreme form of the physical reality problem. In this case there is almost no input from physical or virtual sources, but a rich set of semi-random stimuli from activations sources in the subject's brain (Hobson *et al.*, 2000). During a dream, brain activation of the parietal regions (which subserve spatial cognition) can lead to a coherent mental imagery model of some space (Hobson *et al.*, 2000). If this occurs, then this model can attract spatial attention to itself, and could potentially support spatial updating, leading to presence (particularly if the special case of lucid dreaming turns out to be true - see LaBerge, 1980). Because there is almost no competing stimulation from any other sources, attention can be focused on the mental imagery space exclusively, and a high-undivided presence experience can occur.

The virtual stimuli problem

This model is not able to deal with the virtual stimuli problem, due to the existence of the virtual space pole. The essence of the virtual stimuli problem is that there is no external distinction between stimuli which originate from the virtual source of interest and from other physical sources. Stimuli can only be inferred to have originated from a virtual source once they have been partly processed (Nunez, 2004a). This model guards itself partly form this problem by positing that the movement of presence between the poles is partly due to spatial attention. A subject experiencing a desktop VR system knows which stimuli arise from the virtual space and which arise from the physical space partly by virtue of their location in space – the virtual space stimuli arise from a particular rectangular region (the screen), and all other stimuli are from the physical space. However, this alone does not solve the problem – if for example, one of the pixels of the monitor were to become stuck on a particular colour (say red), then the stimulus of the sole red pixel would probably not be experienced as originating from the virtual space, but from the physical space. Spatial information alone is thus not enough to explain the distinction. The virtual space pole is thus not equivalent to the physical space pole or the mental imagery space pole, as these two poles provide stimuli directly (without the need inference), and the movement from the virtual pole to either of the other two cannot easily be explained by reference to shifts in spatial attention and spatial updating.

The inverse presence problem

In inverse presence, a real world phenomenon is experienced as if it were mediated or virtual. One would have to imagine a situation where attention and spatial updating are firmly focused on extreme end of the physical space pole, and yet the experience is as if focus were on the virtual space pole. This is a difficult phenomenon for this model to explain, as the model does not separate between the perception of presence and the source of the stimuli. The strong link between the source of the stimuli and the type of experience makes it almost impossible to explain inverse presence from this perspective. Also, the low cognitive level of this model (which works mostly in terms of perception and attention) makes it difficult to make use of other concepts such as memory and expectation based processing (as suggested by Timmins & Lombard, 2005) to incorporate inverse presence without major revision.

3.3.3 The Focus, Locus, Sensus / Layers of presence models

3.3.3.1 Description of the model

This model was first proposed by Waterworth & Waterworth (2001), refined in 2003, and further developed by Riva, Waterworth and Waterworth (2004). The 2001/2003 form of the model, called the Focus, Locus and Sensus (FLS) model, does not explain presence itself, but rather explains the experience of being, how this shifts between real and virtual worlds, and how subjects move from awareness to non-awareness of the external world. It proposes three orthogonal dimensions (see Figure 3.4 below):

1. *Focus:* The two extremes of this dimension are *presence* and *absence*. Presence occurs when attention is focused on perceptual processing of the concrete world outside the self. If processing becomes abstract, then attention is turned inwards, and the subject becomes absent in the world (Waterworth & Waterworth, 2001). As attention is limited (Baddeley, 1986), there is an inherent and conscious trade-off between presence and absence along this dimension. A divided state is possible, and indeed occurs in most normal experiences (E. L. Waterworth & J. A. Waterworth, 2003; Waterworth & Waterworth, 2001).

2. *Locus:* This dimension has the extremes of 'real environment' and 'virtual environment', but more specifically describes whether the subject experiences the environment directly, or mediated in some way. At the 'real environment' pole, the subject is directly embodied in the environment, while at the 'virtual environment' pole, the subject experiences the environment through an interface or some other set of hermeneutic relations (Waterworth & Waterworth, 2001). This pole captures the essence of Lombard & Ditton's (1998) notion of presence as the illusion of non-mediation. At the 'real environment' pole, there is no mediation (or at least, no perception of mediation), while at the 'virtual environment' pole, the subject experiences obvious mediation.

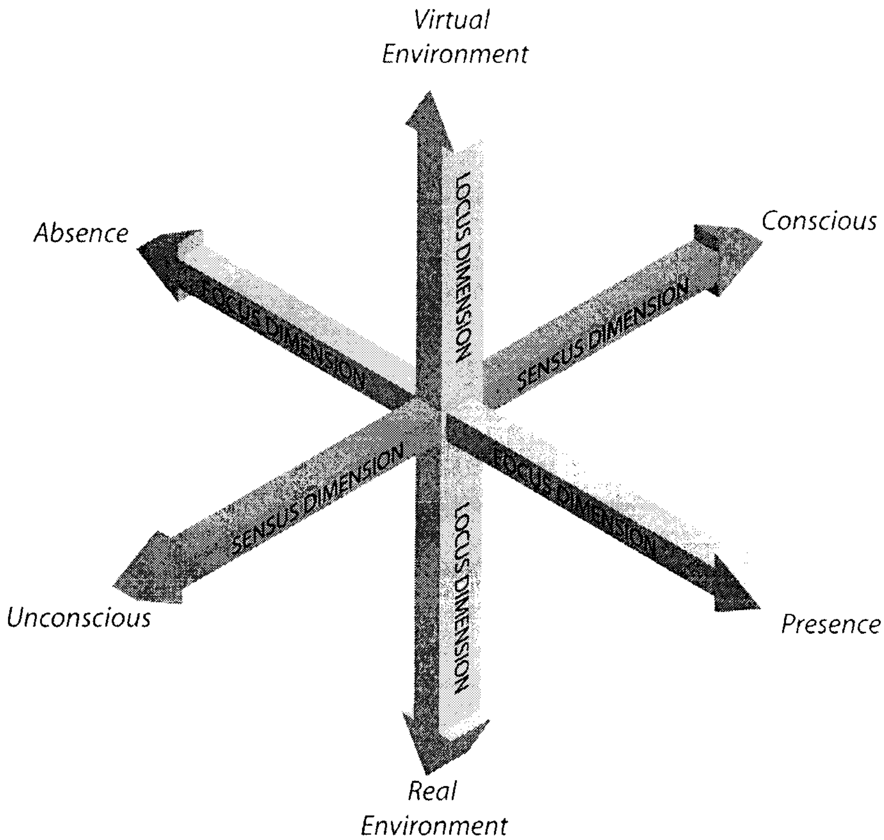

Figure 3.4: The Focus Locus Sensus (FLS) model. The subject's experience is a dynamically moving point in the space defined by the three dimensions.

3. *Sensus:* A novel aspect of this model is this dimension, which describes the subject's level of physiological arousal of the subject, between the extremes of *conscious* and *unconscious*. This dimension interacts with the focus dimension, because when arousal is high, attention tends to be directed outward for tasks such as scanning for new stimuli (Kahneman, 1973) - novel stimuli lead sensus to shift towards consciousness, and attention is focused on those stimuli. As the subject habituates, sensus shifts towards unconsciousness, and attention is freed to attend to internal processes or other external stimuli.

In this model, the position of the subject's mental state relative to the three dimensions determines the character of their experience (Waterworth & Waterworth, 2001). This state is dynamic, with shifts occurring due to a number of factors, which are not explicitly defined. However, the strong relationship between the focus and sensus dimensions, for instance, suggests how a sudden change in external stimuli could shift a subject from an unconscious state of sensus, to a conscious state while simultaneously shifting locus. Also, because the locus dimension is defined in terms of abstract or perceptual processing, it is possible to imagine that a task which demands a high degree of abstract thought would lead a subject to being absent.

The FLS model has been integrated into a complete psychological model, based on the notion that presence is strongly associated with consciousness (a conclusion also made by others such as Slater, 2002). In this Layers of Presence (LOP) model, presence functions on three separate but interactive levels of consciousness. At each level, presence is an evolved solution to some problem faced by the species during its evolutionary history (Riva *et al.*, 2004). Each of the three levels of presence acts to regulate the organism or to initiate action in the world (Riva *et al.*, 2004). The most fundamental problem which presence solves for an organism is distinguishing whether stimuli arise from inside itself, or from the environment (Waterworth & Waterworth, 2001); in humans, this sense has evolved to be significantly complex due the co-evolution of the mind, symbols and cultural artifacts (Riva & Waterworth, 2003). This has allowed presence in mediated environments.

The LOP model is defined largely in terms of neural activation patterns (no doubt derived from Damasio's habit of speaking of consciousness in the same terms). However, these neural patterns are so generally defined that it is possible to use this model as a set of psychological abstractions without specific reference to the brain. The LOP model derives from Damasio's (1999) concept of the self as having three layers, to argue that presence has three layers, each one corresponding to a layer of the self (see Figure 3.5):

1. *Proto self / proto-presence:* This unconscious part of the self contains the immediate state of the subject, including the current state of the sensory organs, as well as the internal state of the individual. Proto-presence represents the degree to which the subject can connect with the world at the most basic level – simple perception-action coupling (Riva & Waterworth, 2003). To be proto-present is thus to be effectively engaging with the world. This allows the subject to differentiate between the self and the outside environment.

2. *Core self / core-presence:* The core self is a conscious construct which is continuously updated with by both sensory information and past experience. It contains the current understanding of the subject's situation. The core self is highly dynamic, being constantly updated by changes in the external world (effectively implementing shifts in attention this way) and internal states such as mood and emotions (Riva & Waterworth, 2003). Core-presence is the outcome of focusing attention on a select subset of stimuli, to create a coherent mental picture of the current situation.

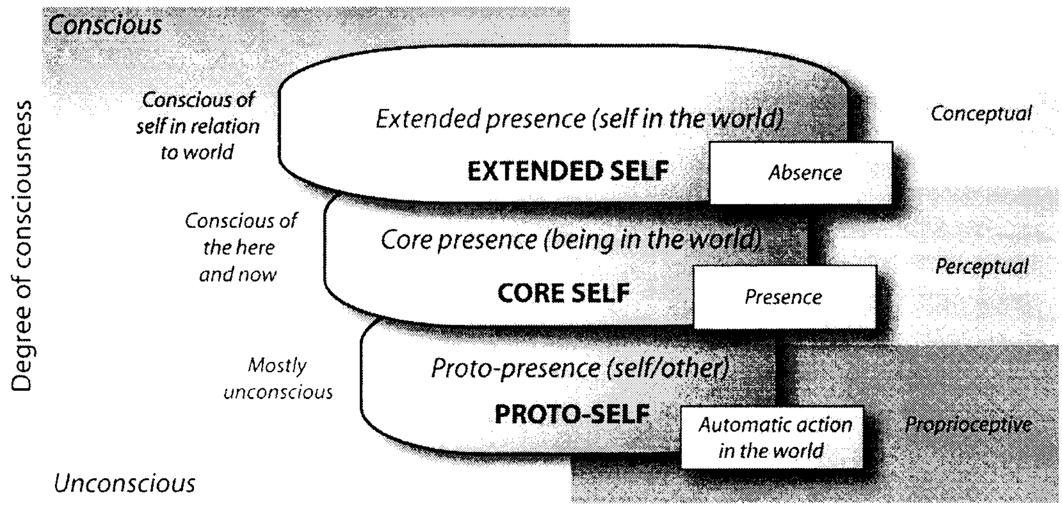

Figure 3.5: The levels of presence (LOP) model. Each of Damasio's layers of consciousness relates to a layer of presence (center ovals), the level of abstraction each responds to, and the type of behaviour which results from engaging each (yellow boxes). Maximum presence arises when all three layers are simultaneously engaged with the world.

3. *Extended self / extended presence:* This contains the most abstract processes, including invariants about the individual (such as biographical memory and personality). The extended self allows the individual to project their current state into the future, effectively making predictions and attributions not just about the individual, but about the environment also (Riva & Waterworth, 2003). The extended self plans, sets goals and creates expectations. Extended presence comes about by comparing the internal state of the extended self

(goals, predictions, etc.) with the environment's state to draw meaning from the individual's actions in the world. The feedback loop of extended presence is associated with achieving goals and extracting meaning about the environment.

3.3.3.2 Presence in the model

The three layers of presence, as with Damasio's three layers of the self, are not independent. They shift to respond to changes in the subject's internal state and to the external state of the world, and under particular conditions can achieve a high degree of integration (Riva & Waterworth, 2003). This integration (which is termed 'focused presence') is understood by the LOP model in terms of the focus, locus and sensus dimensions of the older FLS model. In the LOP model, the focus dimension (which determines whether the subject is focused on the environment or on the self) represents the degree to which the three layers of presence are aligned towards experiencing an external situation. When all three layers are integrated in this way (particularly when core presence is highly integrated with extended presence), the result is high presence; when they are not integrated, the result is absence. Presence can respond momentarily to a change in the environment (as in a break in presence) because proto-presence is highly sensitive to internal/external changes, and exists in the immediate moment. A small change in the environment or emotions can therefore trigger a change which will reduce the integration between the three layers and thus reduce presence (Riva & Waterworth, 2003). The locus dimension represents where the subject is situated experientially – in the real environment or in a mediated environment. Media provide a high degree of extended presence, as they are content-rich and abstract (E. L. Waterworth & J. A. Waterworth, 2003). Media, however, do not allow for direct perception-action coupling (they require an interface), and so proto-presence will not be engaged. The lack of integration between these levels means mediated environments will produce reduced presence when compared to real environments. Finally, the sensus dimension is related to the degree of arousal (Waterworth & Waterworth, 2001) which is passed from the proto-self (internal to the individual), through to the core and extended selves, allowing integration of the three layers and therefore high presence to occur more easily (Riva & Waterworth, 2003). Experiences which have a high degree of personal or emotional significance will begin by arousing the extended self (which understands the world at the most abstract level), and transmit downward through the core and to the proto-self, again facilitating the integration of the three layers of presence (Riva & Waterworth, 2003). The LOP model is thus capable of predicting some content effects in presence.

3.3.3.3 Summary of empirical evidence

As these are recent models, not much empirical evidence exists. The available data nonetheless seem to provide general support for the models. Unlike the three-pole model (Biocca, 2003), one cannot simply take evidence for the environment selection model as evidence for FLS/LOP, as there is no simple mapping between these models. Nonetheless, Riva *et al.* (2004) suggest how the existing corpus of findings about the immersion-presence relationship can be explained by the LOP model. At the top-most level, technology plays a very small role, as extended presence is largely internal to the subject. It involves drawing meaning and having one's predictions supported by the environment (Riva *et al.*, 2004). However, as one moves downwards towards the core layer, immersion has a larger role to play. The perceptually driven core self requires faster updates from the environment than the slower, conceptually heavy extended self. The VE must provide smooth, frequent updates to support core presence and allow its integration with the other two layers. Evidence for this comes from Meehan *et al.* (2002; , 2003) and Barfield and Hendrix (1995), who find that presence is related to update rate. Finally, extremely concrete proto-self canot operate with inputs which require decoding (Riva *et al.*, 2004). Realistic, high fidelity images are therefore preferable to iconic ones. This is supported by the bulk of evidence in support of the two-pole model (see 3.3.1.3 above).

The proto-self is also highly proprioceptive, so multi-sensory inputs which support proprioception will lead to the highest levels of proto-presence (Riva *et al.*, 2004). The evidence for this claim is also fairly clear – systems which use head tracking seem to produce higher presence (Bystrom & Barfield, 1999; Hendrix & Barfield, 1996a), as do systems which include haptic feedback (Meehan *et al.*, 2002; Sallnäs, 1999), and those which have interfaces that involve real reaching or other body movements (Schubert *et al.*, 2002; Slater *et al.*, 1995c). There is also evidence that physiological arousal (the sensus dimension of the proto-self) affects presence – Meehan *et al.* (2002) and Dillon *et al.* (2001) found that at least with for stressful and exciting VEs, change in heart rate correlates well with presence. Finally, in terms of extended presence, Riva *et al.*'s prediction is that non-immersive media will elicit lower presence than immersive media. Nunez and Blake (2003c) found that text-based VEs produce consistently lower SUS and PQ scores than desktop-based VEs; similar findings were reported by Lombard *et al.*, who found a difference in the expected direction between viewers of IMAX cinema, and viewers of small screen, black and white television (Lombard *et al.*, 2000a). Almost all these findings were published while the FLS and LOP models were under development, which makes this mostly post-hoc evidence, and does not suggest that FLS/LOP is any more powerful than the two-pole or environment selection models. Nonetheless, the theory's capacity to hold such diverse findings under a coherent framework is impressive. Two substantial tests of the FLS model have been published by Waterworth and Waterworth with colleagues (2003a; 2002).

The first of these (Waterworth *et al.*, 2002) describes observations of an interactive theater production evaluated in terms of the FLS model. The production (*Incarnation of a Divine Being*) was staged in a shared virtual space, with a chorus and chorus leader at one location, and the other actors (who are in fact also audience members) at different locations, such that most of the contact between participants was virtual. The piece was not scripted, but driven by the chorus and chorus leader who acted not only to drive the play, but also to elicit and mange the action and interactivity of the piece (Waterworth *et al.*, 2002). The VE allowed the participants (actors) to interact in the space by means of body-tracking in a stereopsis enabled large display (Waterworth *et al.*, 2002). The results of the experience were mixed, but give insight into FLS model. The chorus leader (a confederate of the researchers) had the greatest impact on subjects' experiences; many expressed surprise at the degree of their involvement. Most subjects begun the experience feeling anxious about taking part in such a public exercise, but this was replaced by a loss of self-consciousness as they began to interact in the experience (Waterworth *et al.*, 2002). In terms of the FLS model, the chorus leader increases the degree of focus, as his interactions demand the attention of the players, preventing outside stimuli from interfering in the experience. It also seems that participants experience a high degree of locus - they seem to have become part of the virtual performing group and situation, as evidenced by their loss of self-consciousness. This loss of self-consciousness also indicates a high degree of sensus, as subjects lost awareness of their own internal mental states. Although the importance of the chorus leader to the experience can be explained by other presence models, the loss of self-consciousness is more difficult for other models to address. The three-pole model has a strong emphasis on perceptual processing systems, which makes it difficult to explain changes in mental states; similarly, the two-pole model cannot explain the finding, except by reference to an 'acting as-if' explanation (Slater, 2003b), which essentially homunculizes the problem away rather than addressing it satisfactorily.

The second study which examines the FLS model directly (and the LOP indirectly) examined changes in presence and the estimation of time during two virtual experiences – one a field study, the other a laboratory study (J. A. Waterworth & E. L. Waterworth, 2003a). In keeping with the artistic sensibility generally found in the work of Waterworth and Waterworth, the VE used was unusually creative and designed to elicit a novel experience rather than to allow some particular task or practical purpose. The *interactive tent* is a low plexiglass half-tube, similar to a small tent, in which subjects lie and view back-projected images on the tent's surface. The tent also has a stereo sound system, with speakers on each side of the subject's head, and a subwoofer unit (J. A. Waterworth & E. L. Waterworth, 2003a). The subject can interact with the tent by shifting position and posture.

In the field study, the tent was used in an interactive art installation, *the Illusion of Being*. Subjects could control the form of the experience by moving their heads; left-right movements changed the experience from real to virtual (by shifting images and sounds from a realistic, filmed stream to an artificially generated stream) and up-down movements changed the experience from abstract to concrete (by shifting from images and sounds to written text and spoken words describing the scenes). The subject could therefore interactively select between four experiences (real/concrete, real/abstract, virtual/concrete and virtual/abstract). The content of the experience was constant (J. A. Waterworth & E. L. Waterworth, 2003a). Members of the public experienced the tent with no instructions or information given. After a seven minute experience, each subject was interviewed about their experience (J. A. Waterworth & E. L. Waterworth, 2003a). Most subjects did not realize the display changes were triggered by their head movements (many thought it was by means of measuring brain activity).

Most subjects reported changes in psychological state in response to changes in the form of the display, although with significant variation between subjects (J. A. Waterworth & E. L. Waterworth, 2003a). In general, subjects had a stronger sense of space during the concrete streams, and were more confused by the virtual streams. When asked how long the experience had lasted, almost all subjects underestimated the duration of their experience. Waterworth & Waterworth (2003a) conclude that the manipulation of abstractness and locus of the media form affects the character of the experience, as predicted by the FLS model.

To test these notions arising from the field study, *Illusion of Being* was adjusted for use in a laboratory study, which aimed to evaluate the effects of the experience on subjects' perceived duration of the experience. The subjective duration of an experience is related to how much mental work is done during that time: periods of high workload are experienced as longer, and periods of low workload are experienced as shorter (Waterworth & Waterworth, 2001; J. A. Waterworth & E. L. Waterworth, 2003a). The FLS model can explain this phenomenon. Experiences based on concrete stimuli (such as film) require less processing to decode, and because time estimation itself requires mental work, subjects to perceive them as taking longer. Conversely, experiences based on abstract stimuli (such as speech) will require more work to process, and will therefore be experienced as shorter. Because concrete experiences capture more focus, lead to a locus outside the body, and are more likely to stimulate sensus, they are more likely to lead to a focused presence experience (Waterworth & Waterworth, 2001). Concreteness is therefore predicted to correlate with both focused presence and length of time estimated. Time estimation can thus be used as an estimator of focused presence (J. A. Waterworth & E. L. Waterworth, 2003a).

The study used the same interactive tent, with the same four display streams, although subjects did not have control over which stream they experienced. Sixteen subjects experienced all four display stream, and were instructed to focus on the display rather than on estimating time. After each clips, subjects estimated the duration of the clip (J. A. Waterworth & E. L. Waterworth, 2003a). The subject's sense of presence during each clip was measured using eight items from the IPQ. Repetition effects were minimized by using a Latin squares design (J. A. Waterworth & E. L. Waterworth, 2003a). The IPQ scores largely matched the FLS model predictions. The concrete stream lead to more presence than the abstract stream, and the realistic stimuli lead to more presence than the artificial images. These findings are of course also predicted by the two-pole and three-pole models (as concrete and real images are essentially higher fidelity stimuli). The time estimation data were not as clear. Contrary to FLS predictions, there was no effect of media stream on time estimation. Regression analyses predicting time estimate from presence scores did show an effect, but only for one out of the four streams (the virtual/abstract stream). Waterworth & Waterworth (2003a) argue that the general direction of correlation in the other three streams suggest that the effect is general, although small. It may be that the small sample used reduced power and prevented the discovery of this small effect. As they stand the data do not support the notion that there is a relationship between estimates of experience duration and presence. Given that the one significant finding occurred in the case where some mental work was required to decode the content (the virtual/abstract stream), it could be that the relationship between mental workload and estimation is not linear, such that realistic environments require only a trivial amount of work to decode; or more likely, the decoding of realistic environments is handed off to specialized cognitive modules (Fodor, 1983), so that in effect ample mental effort is available for time estimation (Baddeley, 1986). Further work is required to resolve this issue, but as the FLS model stands, it cannot predict this lack of effect. The most convincing finding from this study in terms of the FLS model is the large degree of variability in presence scores, especially given that the tent experience was constant across all subjects. This variability may suggest that individual factors play an important role in the experience. However, given the small sample, it is possible that the variability is simple measurement error or other design artifact, and with a larger replication the effect may disappear.

3.3.3.4 Critical discussion of the model

An innovation of the FLS model is the inclusion of the sensus dimension, which provides an explicit role for physiological arousal in presence. This is important in the light of studies such as that of Meehan *et al.* (2002), which show that at in highly arousing VEs, presence varies with arousal. Also, the sensus dimension is useful in modeling the changes in arousal originating from shifts in attention, or from the arrival of a new stimulus into the perceptual field (Waterworth & Waterworth, 2001).

It is interesting that when FLS was developed into LOP, the sensus dimension was not explicitly converted into an emotion dimension, given that one of the major forces driving core presence is mood and emotion (Riva & Waterworth, 2003). Emotion cannot easily be modeled using arousal (high arousal could indicate anxiety or happiness, for example). It would be a logical step to explicitly include emotion in the LOP model rather than as a secondary force behind arousal. As the models currently stand, it is difficult to understand the exact contribution of sensus and arousal to the LOP model.

One way to understand the FLS model (and by implication much of the LOP model) is as an extension of the three-pole model. As Waterworth and Waterworth (2003b) point out, the locus dimension maps onto Biocca's physical/virtual axis, while the focus dimension maps onto the internal imagery/external stimuli axis. Waterworth and Waterworth (2003b) argue that the FLS model is more perceptual than the three-pole model, because the mental imagery pole requires conceptual processing. This argument is not convincing, because the processing involved in the mental imagery pole still involves manipulation of cognitive maps and perceptual representations of objects rather than abstract concepts. This becomes clear if one examines the underlying neural activation in direct perception as opposed to visual imagination. In functional MRI imaging studies comparing perception tasks to mental imagery tasks, several significant areas of the visual cortex activate in both tasks, and more importantly, similar shifts in activation occur when subjects change their mental images and when stimulus images are changed (Ganis *et al.*, 2004; Kosslyn & Thompson, 2003; Tong, 2002). Also, simultaneously giving a subject mental imagery and a perception tasks often leads to interference between the tasks, indicating that perception and mental imagery are functionally highly similar (Craver-Lemley & Reeves, 1992). This suggests that the three-pole model overwhelmingly emphasizes perceptual processing, while the FLS model, with its inclusion of the subject's body state in the sensus dimension takes a broader view. It is therefore expected that the FLS model would have more explanatory power than the three-pole model. This is supported somewhat by the results of Waterworth & Waterworth (2003a) discussed in 3.3.3.3 above, but the small sample size of that study limits its evidentiary weight.

A final criticism of this model is in terms of its measurement. Presence measurement is well-known to be a difficult problem (Nunez & Blake, 2003d; Singer & Witmer, 1999; Slater, 1999), and Waterworth and Waterworth recognize that this by taking a strong position that presence should be measured by objective means rather than by self report (J. A. Waterworth & E. L. Waterworth, 2003b). They suggest two methods of measuring presence (brain imaging and the time-estimate technique discussed in 3.3.3.3 above), which may indeed turn out to be valid and reliable measures. However, these suggestions raise serious complications for the FLS and LOP models, as each of these measures confound all dimensions and layers of presence into a single estimate. Given that the elements of the FLS and LOP models are internal to the psychology of the subject, they are quite difficult to manipulate. Using only one overall measure of presence (such as time estimation) it becomes quite difficult to validate the relative importance of each dimension or layer, and even harder to tease out the interactions between them. An ideal situation would include a measure of each of the three dimensions of the FLS model (or each layer of the LOP model), such that these can be isolated in studies. However, Waterworth and Waterworth do not offer a suggestion as to how this might be done; the specific validation of their models thus remains a fairly tricky proposition.

3.3.3.5 How the model explains the five problems

The book problem

Waterworth & Waterworth (2003b) explicitly state that the book problem is in fact incorrectly specified; it is not actually a problem. They argue that reading does not engage the senses, so the experience is not presence, but *"almost as if"* presence. They argue that the experience of reading is primarily engagement. This distinction recalls the four factor structure of the ITC-SOPI, TPI and PQ questionnaires (Lessiter *et al.*, 2001; Lombard & Ditton, 2004; Witmer *et al.*, 2005), which also separate *spatial presence* from *engagement*. However, as Lessiter *et al.* point out, these two factors are in most studies highly correlated (Lessiter *et al.*, 2001). This would seem to imply that the distinction, although theoretically quite clear, may not be so clean when one examines the data. Nonetheless, the FLS and LOP models are in fact capable of explaining the book problem. A book can engage the extended self quite effectively (as the reader can relate the text to their own experiences and predict how the story will develop), as well as the core self to some degree (the internal mental model created by reading the book will give the reader an idea of the present moment in the book in terms of spatial layout and states of the characters). However, a book will not engage the proto-self very effectively, as the story world exists only as mental representations which are internal to the reader. The outcome of this experience will therefore not be a particularly focused presence experience. The FLS and LOP models therefore are capable of explaining how books can lead to presence, and why they generally lead to less presence than immersive media, effectively solving the book problem.

The physical reality problem

The FLS model can explain the physical reality problem in terms of focus. The 'absence' extreme of the focus dimension describes the physical reality problem exactly. A subject who is close to absence on this dimension will not reach focused presence, regardless of their position on the other two dimensions. This carries through into the LOP model in the proto-self. A subject who is focused on internal processes will not achieve presence, regardless of the state of the other two layers of the self, as they will not be receiving input from the external environment.

The dream state problem

The FLS and LOP models are well able to deal with this phenomenon. In a dream, the locus and sensus dimensions are highly engaged (the dreamer experiences the imagined world directly, in a highly perceptual experience, and dreams are often extremely physiologically arousing). The focus dimension is also partly engaged, as dreams, although quite strange, are largely concrete. From the LOP perspective, all three layers of the self are engaged, with perhaps only the proto-self being under stimulated due to the disconnection of the motor system (E. L. Waterworth & J. A. Waterworth, 2003). A convergence of the three dimensions and three layers of the self is therefore possible, which predicts a high sense of presence for dreams.

The virtual stimuli problem

The locus dimension of the FLS model allows it to overcome the virtual stimuli problem in an elegant way. Rather than make a distinction between real and virtual environments (as the two-pole and three-pole models do), the FLS model makes the distinction between embodied environments and environments where action is mediated by hermeneutic relations, removing the need for more than one sensory information source outside the subject. The solution poses a few problems, however. Although the locus dimension explains embodied interactions well, it is not clear how it explains mediated interactions. For example, if a subject is sitting in front of two televisions (both mediated experiences), the subject can be more present in one than the other, and one can selectively switch between them (or be jerked from one to another by breaks in presence). The locus dimension does not explain such situations because it only recognizes switches between mediated and non-mediated spaces (as indeed do the two-pole and three-pole models). To completely overcome this problem, the model would need to explain how it is that from a single set of external stimuli, one can extract information to be present in many different places.

The inverse presence problem

The FLS and LOP models are able to explain this problem more fully than the two-pole and three-pole models, although they still do not provide a satisfactory explanation to this difficult problem. To explain inverse presence from a FLS or LOP perspective, one might begin by arguing that inverse presence occurs due to a lack of focused presence during a real-world situation; this may explain why subjects who experience inverse presence do not feel as if they are going through a normal presence experience. Working backwards, we would then assume that at least one of the three dimensions or layers of presence was not aligned with the others. Given that inverse presence is often associated with high anxiety or other physiological arousal (Timmins & Lombard, 2005), one would assume that the proto-self would be bombarded by demands to attend to the internal state of the organism, and similarly the focus dimension would be at the internal extreme. This might explain, from the FLS/LOPS perspective, high arousal situations lead to low levels of focused presence (even in a real environment). This fails to explain one essential aspect of the inverse presence, however – that the situation is experienced as mediated, to the degree that often the best explanation given by subjects is that it is like a movie or a television news report (Timmins & Lombard, 2005), and not like any other of a myriad forms of non-presence (such as daydreaming or plain absence). To explain this key aspect, one would require an associative link between the content of such experiences and the sense of mediation associated with such events. For example, one might expand Timmins and Lombard's (2005) memory based argument for the phenomenon to hypothesize that having experienced such events only through television, the subjective experience of mediation becomes associated with that type of content. When that content is encountered in the real world, the content becomes a retrieval cue for the subjective experience of mediation. However, no such mechanisms exist in the FLS or LOP models, essentially because the only type of learning or adjustment which they model is evolutionary. Neither the subject's previous experiences nor their media consumption history has been included in the model, and so such experiences are difficult to model.

3.3.4 The Measures, Effects, Conditions (MEC) model (Wirth *et al.*, 2007)

3.3.4.1 Description of the model

This recent model is largely cognitive, but includes some media and personality factors also. Presence occurs due to two processes: the construction of a spatial situation model, and the subsequent acceptance of that model as a viable hypothesis for interaction. (Wirth *et al.*, 2007).

The spatial situation model (SSM)

The SSM is a necessary condition for presence, and is therefore central to the MEC model. An SSM is described by Wirth *et al.* (2007) as a mental model of the space, with the following general properties:

> 1. *Completeness* – much a like a memory schema (Rumelhart & Ortony, 1977), the SSM is always complete, even at the earliest stages of exploring a new space (Wirth *et al.*, 2007), and develops more detail with exploration. Regardless of its detail, the SSM can be queried with a reasonable result ("reasonable" here is defined in terms of previous experience). On entering a new VE, a user's SSM of that apartment would be expectation heavy, based more on expectation than perceptual data. This allows even fragmented, incomplete sensory information to lead to a complete SSM. It is not stated whether semantic knowledge can make a contribution to the construction of an SSM equal to that of direct previous experience.

2. *Experiential coherence* – SSMs are derived from a combination of experiential knowledge and sensory information. They are therefore always coherent in terms of these two information sources (Wirth *et al.*, 2007). If sensory information arrives which significantly contradicts some experiential datum, the SSM is restructured to reduce the incoherence. This restructuring may be a slight change, or it may lead to the scrapping of the SSM to be replaced by an entirely new one.

3. *Idiosyncrasy* – As SSMs rely heavily on previous experience (Wirth *et al.*, 2007), it follows that they are highly personal to the subject. Although there will be commonality in SSMs constructed by different subjects (particularly for physical features such as size, colour, etc.), there will also be variation. Here a slight contradiction exists in the SSM concept – although they are defined as being idiosyncratic in this way, SSMs are also referred to as being more or less accurate than other SSMs (Wirth *et al.*, 2007). Given that SSMs are not an objective representation of a space, but a subject-centric representation of a space, it is not clear what accuracy of the SSM means, or what value there is in considering the objective mapping between the space and the SSM.

Phase I: Construction of the SSM

According to the MEC model, the first process in presence is the construction of an SSM. The first requirement for SSM construction is a subject's attention focused on the medium and its content (see Figure 3.6).

The MEC model is unusually sophisticated in its treatment of attention allocation, proposing two attention paths (Wirth *et al.*, 2007):

1. *Automatic attention allocation* – physical properties of the medium (loudness, brightness, etc) as well as unexpected or intense changes in the medium can lead to an orienting response, shifting attention to the medium (Posner, 1980). Because orienting responses occur quickly, they do not involve deep processing of the stimuli. Only physical features such as intensity and relative

position (and not content-related features) are likely to be processed (Posner & DiGirolamo, 1998). This is in line with the two-pole model which proposes that stimuli richness and multimodality are most important in eliciting presence (Slater *et al.*, 1994; Steuer, 1992). Of course, eliciting an orienting response alone cannot lead to sustained attention on the mediated environment. The stimuli must also support sustained attention by being at the very least comprehensible and moderately arousing (Posner & DiGirolamo, 1998). This agrees with Waterworth & Waterworth's (2001) notion of the importance of sensus (physical arousal) in presence.

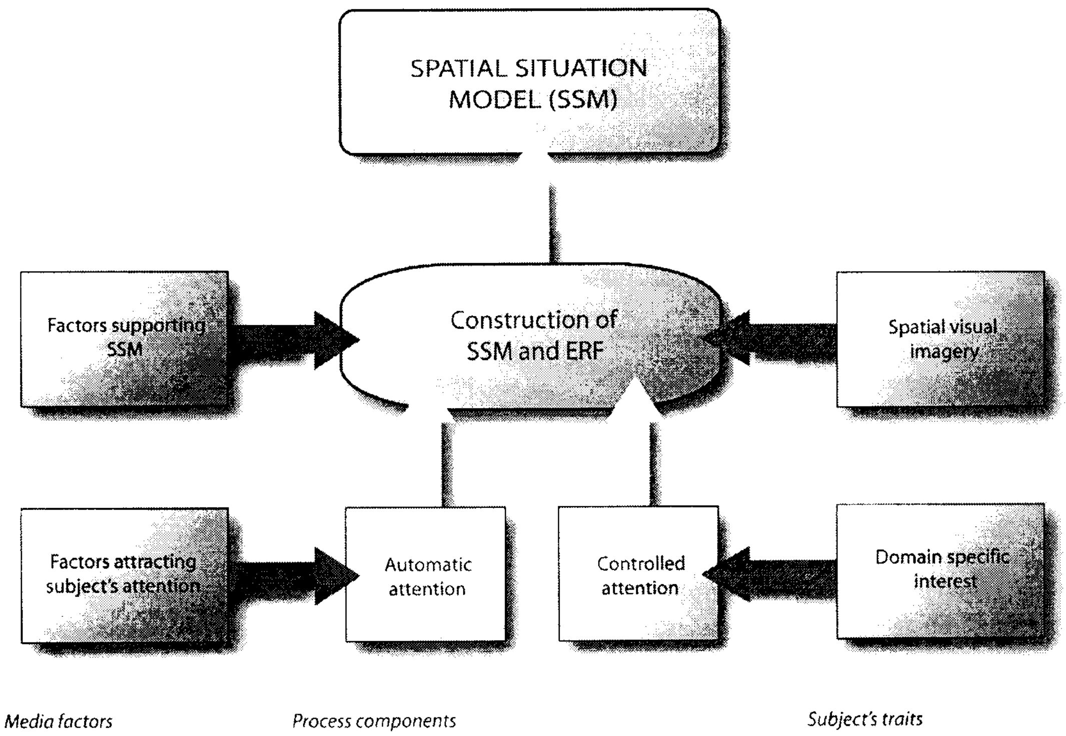

Figure 3.6: The first phase of the MEC model (SSM construction). Attention is allocated and a coherent model of the space is derived; both media factors (green boxes) and individual factors (blue boxes) affect this process.

2. *Controlled attention path* – to reflect the importance of motivation on ttention allocation (Bendiksby & Platt, 2006), the MEC model proposes several factors which explain the maintenance of attention on the medium. First of these is domain specific interest (DSI - Wirth *et al.*, 2007). DSI reflects subjects' increased motivation to attend to particular stimuli according

to their semantic content. Subjects who have their DSI engaged by a medium will find the medium interesting, and therefore willingly focus their attention on it. Apart from DSI, a number of other less significant factors affect the controlled allocation of attention, including fatigue, age, gender and emotional states (Wirth *et al.*, 2007).

These two attention allocation paths interact during processing. The form of the medium might elicit an orienting response, which would temporarily attract attention; then, based on the content of the medium, DSI and other factors might engage attention further. If during the experience a distracter competes for attention, form factors of the medium might allow for attention for remain focused. The degree to which each path contributes varies according to the medium being processed (Wirth *et al.*, 2007). For immersive media, the automatic path would be very active by virtue of the rich stimulus stream and lack of distracters; whereas for non-immersive media (such as books), the controlled attention stream would be most active, to compensate for the lack of stimuli which can produce orienting responses or ward off distracters.

Once attention is focused on the medium, an SSM will form automatically provided the medium represents a space. This process is moderated by two sets of factors (Wirth *et al.*, 2007):

1. *Spatial cues and media factors* – Spatial cues encoded in the medium are the most fundamental contributors to the construction of the SSM (Wirth *et al.*, 2007). This includes static cues (texture gradients, occlusion, spatial audio, etc.) and dynamic cues (motion parallax, stereopsis, Doppler shift, etc.). More cues lead to a more accurate SSM, although the term "accurate" in this context is not defined. These spatial cues must be presented in a coherent way (for example, with sound and visuals synchronized – Wirth *et al.*, 2007) in order for SSM construction to occur. Coherence is defined in terms of the subject's spatial knowledge of such environments (that is, the spatial cues should not obviously violate the subject's expectations for the environment). If one applies this discussion to media forms, it follows that strongly multi-modal, high fidelity systems will more easily allow for the construction of accurate

SSMs by the subject (a prediction which is largely in line with the two-pole and three-pole models).

2. *Spatial imagery and person related factors* –In order for an accurate SSM to be constructed from the available cues, the subject must have the ability to exploit the cues and cognitively process them (Wirth *et al.*, 2007). The best predictor of this ability is the subject's level of spatial visual imagery (SVI). High SVI subjects are better able to extrapolate cognitive structures from available perceptual data, even in the face of missing sensory information (Hegarty et al, 2002 in Wirth *et al.*, 2007). Interestingly, SVI has been linked to the ability to use metaphorical language (Tsur, 2002) and to the comprehension of poetical structure (Tsur & Benari, 2002). This may open an avenue for explaining the unexpected success of books at producing presence, and predicts a relationship between processing immersive and non-immersive media (Biocca, 2003).

Phase II: Selecting an ERF to be the PERF

Another important structure in the MEC model is the ego reference frame (ERF). ERFs are derived from SSMs, but encode a first-person perspective of the space defined by the SSM (Wirth *et al.*, 2007). ERFs are constantly updated as the subject moves through the space, and allows the subject to navigate or initiate action in the space (Franklin & Tversky, 1990). As ERFs are created during interactions with mediated spaces (Schneider *et al.*, 2004), Wirth *et al.* argue that it is therefore likely that subjects can maintain multiple ERFs (for example, one for the real world, and one for the mediated environment - Wirth *et al.*, 2007). Subjects will tend to switch to the ERF which is consonant with stream of stimuli which they are attending to, in order to reduce the resources required to process the environment (Wirth *et al.*, 2007). This stimulus-congruent ERF can be regarded as the primary ERF (PERF - Wirth *et al.*, 2007). When the SSM of the mediated environment is encoded as an ERF, and that ERF becomes primary, the subject experiences presence in the mediated environment (see Figure 3.7).

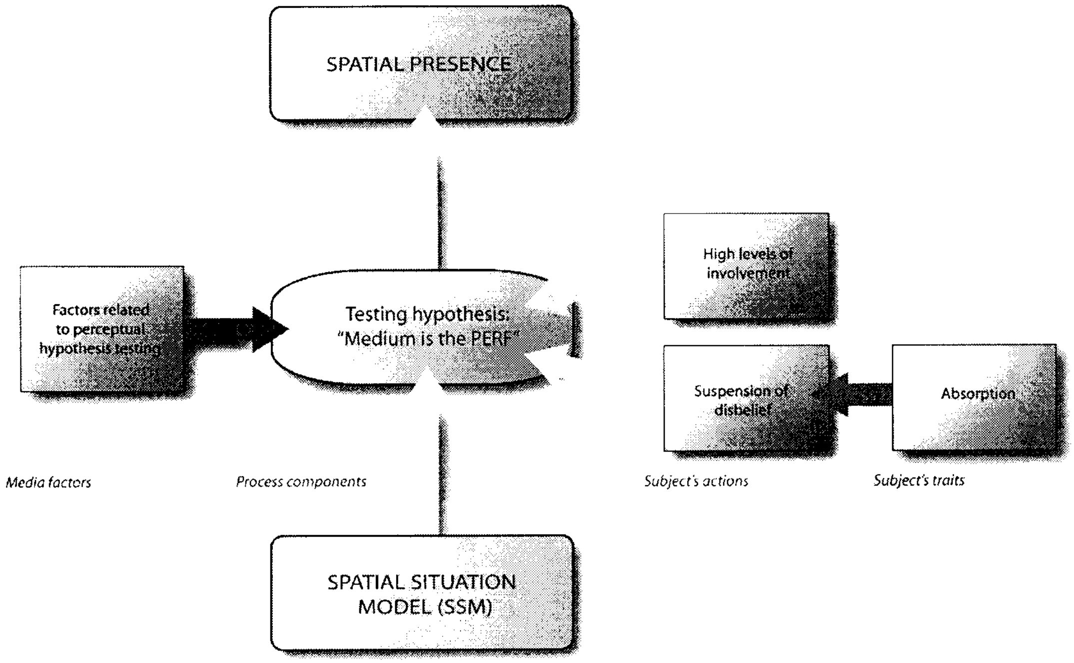

Figure 3.7: The second phase of the MEC model. Perceptual hypotheses are tested to select one ERF as primary, based on media factors and subject traits.

How one particular ERF becomes primary is explained by Wirth *et al.* by means of a hypothesis selection mechanism, based on the perceptual hypotheses theory of Lilli (Lilli, 1997; Lilli & Frey, 1993 in Wirth *et al.*, 2007). A perceiver always entertains multiple hypotheses about the scene, and perceptual information is used as evidence to confirm or disconfirm these hypotheses. The hypothesis with the most evidence is taken as true, and the perceiver behaves accordingly. As the perceptual information changes, new hypotheses may be formulated, and a different hypothesis may be selected as true.

Under this theory, it is easier to activate a hypothesis (prove it) than to deactivate it (disprove it). Hypotheses can be activated top-down (that is, by semantic priming and expectation) as well as bottom up (by perceptual data), although it is not clear what the relative contribution of each of these processes is. In the MEC model, presence defined as the state when the hypothesis "the mediated environment ERF is the PERF" is true (Wirth *et al.*, 2007). When enough evidence supports this hypothesis, and there is not an exceeding amount of contradictory evidence (as might occur during a break in presence - Slater & Steed, 2000), the hypothesis will be taken as true, and presence will occur. During hypothesis selection, the SSM is taken as a source of supporting evidence. A well defined, detailed SSM will support the hypothesis effectively, while a weak SSM (one which is consists mostly of conceptual information and is constantly contradicted by perceptual information) will not support for the hypothesis. These notions seem to be a more explicitly developed form of the environment selection model presented in 3.3.1.1 above (Slater, 2002; Slater & Steed, 2000), but the lack of reference to Slater's work in Wirth *et al.*'s paper and vice-versa suggests that the ideas were developed in parallel.

The hypothesis testing process is not entirely automatic. Two individual factors moderate the process at an abstract level: *involvement* with the medium content, and the *suspension of disbelief* by the subject. The MEC model sees involvement as non-critical acceptance of information from the mediated environment, which is strongly associated with the subject's motivation (Wirth *et al.*, 2007). Involvement has cognitive consequences (e.g. directing attention towards the medium or elaborating the stimuli to give them meaning), affective consequences (e.g. changes in mood or attitude towards the content) and behavioral consequences (e.g. selecting a particular action). As involvement is largely dependent on the content of the medium, it would seem that there is a weak relationship between involvement and spatial presence (as was argued by Slater, 2003a). Wirth *et al.* agree with Slater that involvement is not necessary for presence, but it can, under very particular situations, facilitate spatial presence for two reasons. First, a highly involved subject (who is highly motivated to experience presence) can lead to a subject willingly activate the ERF and therefore increase the probability of experiencing presence (Wirth *et al.*, 2007); second, highly involved subjects will automatically allocate more resources to processing the medium, which will leave less resources for processing competing stimuli. Less competing evidence means a higher probability that the appropriate ERF hypothesis will be selected and presence will result (Wirth *et al.*, 2007).

Suspension of disbelief is the term used by Wirth *et al.* for conscious strategies used to elicit or improve presence experiences. Suspension of disbelief is independent of involvement (Wirth *et al.*, 2007), although it seems to follow that it would be more effective if used by a highly involved subject. Wirth *et al.* see three components to suspension of disbelief: disabling the processing of contradictory or distracting stimuli; actively suppressing those contradictory stimuli which enter consciousness; and re-interpreting those stimuli which could not be suppressed as evidence in favour of the appropriate PERF hypothesis (Wirth *et al.*, 2007).

3.3.4.2 Presence in the model

As with the other models described, presence in the MEC model is a particular model state which arises under particular circumstances: when an SSM of the mediated space has been formed, and the hypothesis that the ERF encoding that SSM is the primary ERF is selected by the subject (Wirth *et al.*, 2007). Presence is a potentially fragile state, which can be interrupted by competing stimuli (as the PERF hypothesis could be selected out – Wirth *et al.*, 2007). A novel contribution of this model is the precise way in which it defines presence, which is possible due to the fact that this model limits itself to explaining spatial presence. The model does not include notions of presence as naturalness, or as engagement with the content, as suggested by Lessiter *et al.* (2001) and Lombard *et al.* (2000). It should however be noted that the relationship between involvement and spatial presence is indeed thoroughly discussed by Wirth *et al.* (2007).

3.3.4.3 Summary of empirical evidence

Although the MEC model is recent (published late in 2007), a significant amount of evidence exists suggesting its validity. Some pre-existing published work can be taken as evidence for the model, as in many ways this model expands the environment selection model by suggesting that subjects have at least two conflicting sets of stimuli, from which they select one in which to become present; and that a VE's fidelity, multimodality and the capacity to attract and hold attention predict presence (Slater, 2002; Wirth *et al.*, 2007). The MEC model also emphasizes the importance of consistency across stimuli. Evidence for this comes from Vinayagamoorthy *et al.* (2004), who simultaneously manipulated two aspects of scene realism: fidelity of characters in the scene, and fidelity of textures in the environment. The lowest presence scores were found when high fidelity characters were placed in the low fidelity scene. This suggests, as predicted by the MEC model, that presence is moderated by the fit between scene elements.

A few recent key studies have been carried out to test hypotheses which are specific to the MEC model. The first was a small study (n=26) by Gysbers *et al.* (2004), looking at the effect of number of spatial cues embedded in a text description of a space on both vividness of the SSM and spatial presence. Subjects read one of three passages describing a space: the first one contained a few spatial cues, the second contained many spatial cues, and the third contained many spatial cues plus instructions to imagine the space. The results showed, as predicted, that SSM vividness was related to number of spatial cues. The presence data, however, were inverted: more cues led to lower spatial presence. The authors interpreted this correctly to mean that SSM vividness is not a simple predictor of spatial presence. A close examination of the MEC model shows that SSM vividness is related only to the first phase. Presence will only occur if the SSM makes it into the PERF, which requires enough perceptual evidence support that hypothesis. It is likely that more spatial cues in the text *reduces* the evidence for that hypothesis, as more spatial cues in the text increases the probability of having perceptual contradictions (this is only plausible for text based environments, where the information is conceptual rather than perceptual). The study well demonstrates the complexity of the relationship between spatial cues, SSMs and presence (which one might argue justifies the complexity of the MEC model itself).

Evidence to support the distinction between the formation of an ERF from the SSM and the adoption of this ERF as PERF comes from the studies used to validate the MEC spatial presence questionnaire (MEC-SPQ) (Böcking *et al.*, 2004). 291 subjects under various media conditions completed an early version of the MEC-SPQ. A factor analysis revealed a reasonably strong three factor structure (explaining a little more than half of the total variance). The three factors were named by Böcking *et al.* as *self-location, possible actions*, and *cognitive involvement*. These three factors represent more the abstract factors in the MEC model: *self-location* is a measure of how an SSM becomes an ERF, *possible actions* measures the degree to which the mediated environment is taken on as the PERF (provided one accepts that any action is only possible if one positions oneself, hypothetically at least, in that environment), and *cognitive involvement* measures executive control over the adoption of the SSM as PERF. Although the use of factor analysis in presence theory has been criticized as a means of deriving theory (Waller & Bachmann, 2006), it should be noted that in this case the factor analysis was used not as an exploratory tool, but as a confirmatory technique (a preliminary version of the MEC model had already been published before this validation study; see Vorderer *et al.*, 2003).

An interesting feature of the MEC model is its fairly detailed account of the role of subject traits on presence. Noted that although the studies on the effect of personality variables on presence provide general support for the MEC model, they also support (although far less specifically), the LOP model. This is because the LOP model proposes that presence arises as a function of the layers of the self (Riva & Waterworth, 2003), which are presumably related to personality variables. Nonetheless, these studies provide stronger support for the MEC model than for the LOP model as the MEC model makes more specific predictions about these variables. One such study by Laarni *et al.* (2004) found that, as predicted by the MEC model, personality variables do have consistent effects on spatial presence. In particular *self-forgetfulness* had a noticeable effect. Self-forgetfulness is associated with easily losing consciousness of the self and of the passage of time when engaged in interesting activites (Kose, 2003). Interestingly, Kose (2003) uses the term "being in another world" (pp. 93) to describe high self-forgetfulness scorers, which immediately perks up a presence researcher's ears. Similarly, the reference to the loss of awareness of time is reminiscent of Waterworth & Waterworth's (2003a) study involving perception of time; it is therefore not surprising that this trait should predict spatial presence well.

A second study of interest is by Sacau *et al.* (2005), which examined three key personality traits posed by the MEC model – *domain specific interest, spatial visual imagery* and *absorption* (the first two are associated with the formation of the SSM, the third on is associated with suspension of disbelief and therefore with the adoption of the SSM as PERF). This large study (n=240 from four different countries) used four conditions: the first read linear text, the second read media-rich hypertext, the third watched a film, and the fourth navigated a three-dimensional VE. All four conditions encoded large, old buildings such as libraries or temples. Spatial presence was measured with the MEC-SPQ, which includes measures of spatial presence, domain specific interest, spatial visual imagery, and absorption (Vorderer *et al.*, 2004). The results showed domain specific interest and absorption are related to spatial presence as predicted (r=0.31 and r=0.19, respectively), but spatial visual imagery showed no relationship. When examining the effects of these three traits on the four types of media, the data showed that domain specific interest was a powerful predictor of spatial presence, only failing to predict spatial presence in the linear text condition. Absorption only predicted spatial presence in the media-rich hypertext condition. These comparisons may be somewhat blurred by the fact that the four media conditions were not randomized, but conflated with country (that is, all subjects in any one media condition came from the same country). Nonetheless, the finding is a confirmation for an important element of MEC, namely that interest in the content can enhance presence (presumably by means of allowing more control over attention - Wirth *et al.*, 2007), although a better test of this hypothesis would have been to manipulate an attention distracter during the study, as currently the path by which domain specific interest affects spatial presence is not clear. The fact absorption was a weaker predictor of spatial presence may be significant for the MEC model. While domain specific interest is thought to affect spatial presence at the stage of the formation the SSM, Absorption is thought to affect spatial presence at the stage of selecting the SSM as PERF. Recall that an accurate SSM is a necessary condition for its selection as PERF; this means that absorption affects spatial presence *after* domain specific interest has made its contribution. One would therefore expect a lower correlation between spatial presence and absorption, as it effectively functions as a moderator in the path between domain specific interest and spatial presence. This could be tested by explicitly running a path analysis with absorption as a moderator, or by having an explicit measure of suspension of disbelief (with which absorption should have a direct relationship).

3.3.4.4 Critical discussion of the model

The MEC model is well-defined with a sizeable amount of evidence supporting it. It is able to describe how perceptual data and conceptual data interact through a set of well-defined cognitive processes. Unlike the hypothesis-selection and three-pole models, the formation of the presence experience is not a spontaneous 'black box' phenomenon, but is posed as the outcome of two separate processes. These allow the model to explain failures to become present in the face of immersive media (either as a failure of the medium to attract attention at the first process, or as a failure for the SSM associated with the medium to become PERF), as well as breaks in presence (when a stimulus which is not consonant with the current SSM takes attention and either reduces the coherence of the SSM, or reduces the amount of evidence for it to be selected as the PERF). Finally, the MEC model is able to provide a lucid explanation of the interaction between medium content and spatial presence, by positing that attention allocation is moderated by domain specific interest. Even when compared to the more substantial FLS/LOP models, the MEC model is capable of generating very explicit hypotheses for empirical testing, and is able to link the higher levels of cognition (such as willing suspension of disbelief, controlled attention and domain specific interest) with very low level variables such as stimulus intensity and attention (the FLS/LOP models favour the higher level constructs, being more vague about lower level processes).

The MEC model also has some weaknesses which need to be mentioned. Firstly, the MEC model has been deliberately formulated as a model of spatial presence *in mediated environments* (Wirth *et al.*, 2007). This choice is particularly strange given the strongly psychological character of the model. From an evolutionary perspective, a model of presence must assume that whatever mechanisms lead to presence must have evolved long before the existence of mediated environments (Reeve & Nass, 1996; Biocca, 2003; K. M. Lee & Yung, 2005). A theory of presence should therefore be able to explain presence in real environments first of all. On careful reading of Wirth *et al.* (2007), it seems that the MEC model may indeed be able to adequately explain presence in real environments, if the importance of particular variables is modified. For instance, media related variables, such as the medium's ability to hold the subject's attention, would need to be reduced in importance.

A significant theoretical weakness in this model is related to the conception of presence in the model, and the use of the MEC-SPQ in studies validating it. The existence of the MEC-SPQ (which quantifies the subject's experience as a number based on a sum of responses to Likert type items) strongly suggests that the model treats presence as a continuously varying quantity. This notion is supported by how the spatial presence subscale scores are used in research (normally as covariates, or as outcomes to analysis of variance analyses – see for instance Böcking *et al.*, 2004; Sacau *et al.*, 2005). However, the model itself treats presence as a binary phenomenon, because presence is defined as the state when the SSM encoding the mediated world is *selected* as the PERF. If the scales are being used to support a binary concept of presence, one might expect a absolute score cut-off above which subjects are considered to be presence (similar to that done by Slater *et al.*, 1995c, where responses at the top end of his questionnaire as scored as 1, and all below as 0); however, the MEC-SPQ does not contain such a provision. This situation creates two fundamental problems: one, the definition and operationalization of presence are contradictory; and two, that a model which implements a binary concept of presence has been supported mostly by evidence derived from continuous measures of presence. This problem does not have a simple solution.

3.3.4.5 How the model explains the five problems

The book problem

Given that Biocca is included in the list of authors of the MEC model, one would expect the model to deal with the first three problems quite well, and this seems to be the case. In the MEC model, books and other non-immersive media can lead to presence as long as they are processed cognitively to produce an SSM (Wirth *et al.*, 2007). Recall that SSMs are created by a combination of sensory and conceptual data (hence their constant completeness). This implies that even with very little sensory input (as would occur in a book) an SSM can still occur. Also, there is no reason to think that a book cannot provide enough spatial information to allow for a detailed SSM, particularly if the subject has a high degree of spatial visual imagery. The difficulty lies in the book SSM being taken as the PERF. As books produce low levels of stimulation on a single modality, and they take considerable effort to decode, reading is easily interfered with by other stimuli. This makes it difficult for the 'book SSM as PERF' hypothesis to be maintained, and although presence can occur, it will likely be continually interrupted. Furthermore, the MEC model is capable of explaining why books generally produce presence for particular individuals better than for others. This occurs at both stages of the process – first, individuals with high spatial visual imagery will be able to construct a more accurate SSM, and second, individuals with high domain specific interest in the content of the book will be better able to control their attention though suspension of disbelief, and thus support the 'book SSM as PERF' hypothesis by eliminating support for rival hypotheses.

The physical reality problem

This problem is essentially one of mental effort being turned towards internal processing rather than to processing sensory stimuli. The MEC model includes a highly detailed description of the role of attention, so it copes with the physical reality problem reasonably well. If one assumes that the real world is encoded simply as another SSM and ERF (with extremely high levels of stimulus richness and multimodality, of course), then one can consider the amount of attention given to it and the amount of support for the 'real world SSM as PERF' hypothesis in the same was as a VE is considered. If the level of attention on the real world is low, very little support for the hypothesis will exist (due to a low degree of supporting evidence), and the hypothesis will be dropped. However, it is not clear what happens when the hypothesis is dropped, as the MEC model seems to assume that there will always be at least one other stimulus source which can lead to an SSM and ERF. This assumption exists because the MEC is a model of presence in *mediated* environments, and the real world is always assumed to exist as a replacement hypothesis. If the model is modified so that it is possible that no ERF is active, then this would explain the physical reality problem well. However, in its current form, the MEC model is not ideally suited to explaining this problem.

The dream state problem

This problem is similar to the physical reality problem, as it involves a situation in which internal processing is favoured while external stimuli are blocked out. This situation is possible under the MEC model, as discussed above, although it requires stretching the model somewhat. The interesting aspect of the dream state problem is the dream itself – it provides a source of high-level stimuli (not sensory, but perceptual) which can lead to presence. From the point of view of the MEC model, this is fairly straight forward. The dream provides a set of stimuli not unlike those produced by reading, and from these an SSM can be built if the correct cues are present. Unlike reading, individual factors such as spatial visual imagery will not play a large role, as during a dream the visual cortex is likely activated directly by the reticular activating system (Hobson *et al.*, 2000). If an SSM forms, then presence should occur if the perceptual hypothesis 'dream SSM is PERF' has sufficient evidence. Given that external stimulation is disconnected during dreaming (Waterworth & Waterworth, 2001), a great deal of supporting evidence can easily be collected for the hypothesis. However, the activation of the visual cortex during dreaming is largely random (Hobson *et al.*, 2000), so it is still possible for discordant images to appear and disprove the hypothesis.

The virtual stimuli problem

All of the models which have been discussed to this point (except perhaps the FLS/LOP models - Riva & Waterworth, 2003; Waterworth & Waterworth, 2001) have been unable to satisfactorily resolve the virtual stimuli problem, because they explicitly draw a distinction between a finite number of 'streams' or 'worlds' from which sensory data arises. The MEC model overcomes this problem completely by its use of SSMs. The SSM is an abstract structure which is composed of sensory and conceptual data which is selected by its content to be internally consistent from the sensory information available (Wirth *et al.*, 2007). There is no requirement that data arises from any particular source, or that the inclusion of some data, due to their origin, will reduce presence. Under the MEC model, every piece of spatial data (be it sensory or conceptual) is considered either as evidence in favour of or opposed to one or more SSMs and ERFs. If it fits (content wise) with the SSM or ERF, then it is evidence in favour of the perceptual hypothesis, and it increases the likelihood of presence; if it supports a different hypothesis, then it reduces the likelihood of presence. Note that this solution to the virtual stimuli problem is possible because presence is made an arbitrary state. One is said to be present if one particular ERF (of many possible ERFs) which we have designated as being of interest is supported as the PERF. In the MEC model, one is always present in some SSM (in that some SSM is always active in the current PERF); but "presence" is only counted when the SSM of the mediated environment we are investigating is the PERF.

The inverse presence problem

Although the MEC model is able to explain the other four problems reasonably well, it unfortunately does not deal with the inverse presence problem well. In the MEC model, presence is defined as simply the switching to a PERF previously defined as interesting. Mediation is considered only as a factor which reduces the quality of sensory data, and makes the subject less likely to attach and hold attention to those stimuli. This cannot capture the essentials of inverse presence which are the qualia of a mediated experience, and the triggering of particular expectations based on the content of the environment (Timmins & Lombard, 2005). This cannot be modeled unless some system of automatic, associative memory is included, such that particular features of a situation can trigger off contextually dependent expectations (e.g. fire and explosions must mean that Arnold Schwarzenegger is about to enter the scene, and I feel like eating popcorn). The MEC model does include aspects of memory and previous experience - these are the components of the library of spatial experiences which subjects use to interpret sensory stimuli as spaces. But these contain only spatial information, and are not linked to episodic memories of previous experiences in similar spaces which may lead to the expectations characteristic of inverse presence. One can defend the MEC model in this regard by stating that it is a model of *spatial presence*, and its lack of power in explaining inverse presence comes from an intentional design choice rather than an inherent model limitation. This may indeed be the case, but if it is so, the MEC model will also have problems explaining other phenomena related to the qualia associated with particular spaces (such as a feeling of awe on entering a cathedral).

3.4 Conclusion

This chapter has summarized four of the current significant families of presence models on a common framework to aid comparison (see the summary table in Appendix G). Most models have some roots in the two-pole model, and thus benefit from the large body of empirical support of that model. Generally speaking, the two-pole/environment selection and the MEC models have the most significant body of empirical evidence, and the FLS/LOP models have the least. In terms of the five problems, no one model was able to convincingly deal with all five. The two-pole/environment selection model is the weakest in this regard, dealing only with the dream-state problem. The strongest in this regard was the MEC model, which could deal (at least partly) with all except the inverse-presence problem. The following chapter will propose the CLICC model, which aims to distill the strengths of the four families of models discussed in order to deal with all five problems, while remaining consistent with previous empirical findings.

Chapter 4: The capacity limited, cognitive constructionist model of presence (CLCC)

This chapter describes the capacity-limited, cognitive constructionist model (CLCC) of virtual presence. The chapter begins by giving a broad justification for the architecture used (sections 4.1 and 4.2), and then moves to describing the components and structure of the model (section 4.3). Finally, it discusses how presence exists in the model (section 4.4), and illustrates the predictive power of the model by explaining, among other theoretical points, the presence-immersion relationship (section 4.5) and the five presence problems defined in 3.1 in chapter 3 (section 4.7).

The CLCC model contains a complex modular structure (see 4.1 below), and it models presence as a dynamic model state, which can be understood as consisting of two main effects which occur simultaneously and interactively:

1. The creation of a semantic (themed) bias which permeates the model
2. The construction of the current environmental situation in working memory

This interaction keeps the model relevant to the environmental situation, in as cognitively efficient a manner as possible (this is necessary due to the small capacity of human working memory – Baddeley, 1986). Once a particular construction or understanding of the environment has been arrived at, the model is present in that environment. The model will then try to maintain that construction by filtering out perceptual information and inhibiting the influence of top-down data which are semantically irrelevant. If it should occur that the current construction is no longer relevant to the environment (either due to the existence of a lot of contradictory data, or because of an unexpected change in the environment), a process of reconstruction will occur, which will make a new, meaningfully coherent construction of the environment. The name of the model describes this basic operation: it is *capacity limited* due to the cognitive constraints of human information processing, and *cognitive constructionist* due to the semantic construction of a meaningful, coherent mental structure of the environment by the subject.

4.1 Structural basis

As suggested by Biocca (2003) and Lee (2004), the mechanisms which give rise to the presence experience must serve an evolutionary purpose more fundamental than presence in virtual environments. These mechanisms exist to provide the subject with an up-to-date mental representation of the environment, into which are encoded learned potential courses of action. Because evolution tends to develop systems which are brutally efficient (Tooby & Cosmides, 1990), this is achieved within the constraint of highly limited capacity, such that all parts of the system are always using all available resources to process the scene and infer possible actions. Because the system has evolved to deal with stimuli which encode real environments, it is possible to fool it, provided stimuli are presented in a form which somehow stimulates the appropriate cognitive modules (Lee, 2004).

In order to the 'lean and mean' character of evolved cognition, the CLCC model of presence is compatible with current findings in cognitive science, while being able to predict and explain presence phenomena. The structure of the model largely follows the information processing models of cognitive psychology. Specifically, it follows the *stages of processing* model of Atkinson and Shiffrin (1968), and to some extent the *levels of processing* model of Craick & Lockhart (1972). Although there has been some controversy about these models of memory, a large number of empirical studies over the past thirty years (including several meta-analyses), suggests that these are robust and well supported models of semantic processing (Conway, 2002; Decoster & Claypool, 2004). In the tradition of the stages of processing model, the CLCC model proposes that the cognitive architecture which leads to presence is a set of discreet processing modules which take data as input from some nodes, transform that data, and output it to another set of nodes for further processing. Unlike classic stages of processing models, however, the CLCC model operates due to emergent properties of the structure, which arise due to the interaction of the data content, and how it flows around the entire model. Considering presence as the emergent property of a complex system is common in extant presence models - for instance, it has been used in the LOP model (Riva *et al.*, 2004) and MEC model (Wirth *et al.*, 2007) discussed chapter 3. The role of emergence in the CLCC model has also been inspired by the levels of processing model (Craik & Lockhart, 1972), which states that in a cognitive processing system, information is encoded and processed at multiple levels of abstraction, from perceptual to conceptual (Craik & Lockhart, 1972). Furthermore, the levels of processing model states that synchronous processing of data across several levels of abstraction leads to an enhancement of overall processing performance (Craik & Lockhart, 1972).

Many extant presence models are designed to only explain presence phenomena (such as the presence models of Biocca, 2003; Slater, 2002; Wirth *et al.*, 2007). The CLCC model however has been designed to closely reflect existing general purpose human information processing models, as is done by the LOP model (Riva *et al.*, 2004). This approach has two advantages; one theoretical, and one pragmatic. Theoretically, it satisfies Biocca's (2003) and Lee's (2004) evolutionary requirement that presence should not be considered as a unique response to mediated environments. More pragmatically, using a well-understood architecture as the basis for a model reduces the need to empirically validate the overall structure of the model, allowing empirical investigation to focus on specific aspects of the model which are pertinent to presence.

4.2 Cognitive constructionism

A central axiom of the the CLCC model is that each subject constructs their own unique experience of an environment by applying previous knowledge of similar situations and by interacting with the current situation (Bruner, 1990). In this way, the subject can operate in novel environments, but in an adaptive way which is constrained by previous experience. This idea is derived from educational (and particularly Marxist) theories (see for instance Luria, 1974; Vigotsky, 1978), but also exists in cognitive psychology as the idea that goal-directed behaviour can only occur when there is an interaction between bottom-up and top-down information. Bandura (1986) referred to this as *reciprocal determinism* – the subject, as an active agent, changes the environment, and is simultaneously motivated by those changes to respond in particular ways. Bandura's explanation operates at the level of the individual, but similar concepts exist at lower levels. For instance, Rumelhart and McClelland's layered, competition-based connectionist networks only reach a stable state (and thus complete computation) when activation has been resonated between the top-most and bottom-most levels of the network (Rumelhart *et al.*, 1986).

Inclusion of constructionist principles allows a presence model to simultaneously explain cognitive phenomena (such as the role of attention allocation or the importance of multimodality; see Sallnäs, 1999; Slater & Steed, 2000) and subject-environment interaction (such as the role of passive haptics or successful navigation; see Barfield & Weghorst, 1993; Meehan, 2001).

Three essential features of constructionism (Bruner, 1990) which may be useful to a presence model are:

1. *The subject is an active agent in the world* – a subject's normal mode of being in an environment involves exploration and interaction. This is explicit in some models of presence, such as the argument for embodiment present in Biocca's model of embodiment (Biocca, 1997), and the Locus dimension of the FLS model (Waterworth & Waterworth, 2001). Not all models of presence recognize the subject's agency in the environment however. Notably the earlier versions of the environment selection model (Slater, 2002), which see the subject as passively selecting and processing information about the environment, with the presence experience arising more or less automatically when the correct conditions are met (Biocca, 2003; Slater *et al.*, 1995c; Witmer & Singer, 1998). Recent developments of the environment selection model (Slater, 2002) have emphasized the subject as actively participating in their experience, suggesting that there is a fair degree of consensus about the importance of agency in presence.

2. *The subject interactively creates their experience* - under this principle, the subject's experience of the space depends on their actions and goals within that space (Bandura, 1986). This notion was first applied to telepresence by Sheridan (1992b) with his elegantly simple model of in-the-loop human-machine interaction and telepresence. Recently, it has been extensively adopted by the MEC model (Wirth *et al.*, 2007), which emphasizes the role of the SSM (an interactively constructed model of the environment). The importance of a subject's interactions in the VE on their presence experience has been an area of active inquiry. Schubert *et al.* (2002) showed that subjects' perceptions of their degree of interaction in the VE positively predicted

presence. Other studies which have examined the relationship between presence and task performance or task completion (which require interaction in the VE), have shown that, for instance, navigation (Barfield & Weghorst, 1993), self-rated ease of task (Bystrom & Barfield, 1996) and self-rated task performance (Romano *et al.*, 1998) all correlate with presence in the expected direction.

3. *The final product is meaningful to the subject* – as the experience is constructed by the subject by combining the current situation with previous experience, it is essentially idiosyncratic (Bruner, 1990). This implies that the content of a VE will interact with the subject's previous knowledge during the presence experience. Although Slater has argued against the role of content in the presence experience (Slater, 2003a), the MEC model (Wirth *et al.*, 2007) proposes that the subject's domain specific interest (which predicts their interest in and experience with the VE content) plays an indirect role in the presence experience. Similarly, the FLS model (Waterworth & Waterworth, 2001) proposes that how arousing the content of the VE is will partly determine the nature of the presence experience. There is very little empirical evidence of how the meaning given to a scene affects presence in that scene. An interesting study by Nowak *et al.* (2006) showed that for violent video games, presence was predicted not by the objective degree of violence in the game, but by the degree of violence *perceived* by players, suggesting that the meaning attributed to the scene may in fact contribute to the presence experience.

4.3 Structure and components of the model

The CLCC model has the same basic architecture as the stages of processing model (attention filter, short-term storage, long term storage – Atkinson & Shiffrin, 1968), with some important differences – see figure 4.1 below. First, unlike the stages of processing model, the data paths between components in the CLCC model are one-way – that was done simply to make more explicit the stages of data transformation. Second, the CLCC model is not only a model of memory encoding and retrieval. Although memory plays an important role in the CLCC model, it also includes aspects of perception and motor control. Finally, while the stages of processing model describes only the subject, following the arguments of Floridi (2005) and the theoretical suggestions of Biocca (1997) and Schubert *et al.* (2002) the CLCC model is epistemic - it includes aspects of the environment, and provides a mechanism for feedback via the subject's perception of their own actions in the environment.

The components of the model are as follows (described from bottom to top):

4.3.1 External stimuli via sensory cortices

Description: This is the entry of bottom-up data into the model. The data are in the form of unprocessed neural stimuli from the retina, the auditory nerves, etc. This node (which includes basic perception done in the sensory cortices) converts raw neural stimuli into objects. For example, visual stimuli are converted into a 2½ dimensional sketch (Watt, 1988), and motion is identified by optic flow on the retinal field (Cavanagh & Mather, 1989).

Inputs from: The physical environment (including all immersive systems, real events, the body itself, etc), which provides physical stimulation.

Outputs to: The stimulus attenuator for selection of relevant stimuli.

4.3.2 Stimulus attenuator (attention)

Description: The primary purpose of this node is to ensure that capacity limits of the system are not violated (Baddeley, 1986; Posner, 1980). This is done by selecting a subset of stimuli from those available at the sensory cortices. Stimuli are selected as a function of their relevance to the current semantic bias of the entire system (Lavigne & Denis, 2001; Maxfield, 1997). However, this selection is not a simple linear function. The probability of a stimulus being selected for further processing is highest at both extremes of congruency to the current system bias. Highly congruent stimuli, which support the current bias, are selected, and incongruent stimuli are excluded to maintain semantic coherence. However, in order to remain sensitive to significant changes in the environment, highly incongrious stimuli (either in terms of physical properties such as brightness, or in terms of semantic difference) are highly likely to be selected for further processing (Treisman, 1969), which would result in a violation of semantic coherence, and a reconstruction (see 4.3.11).

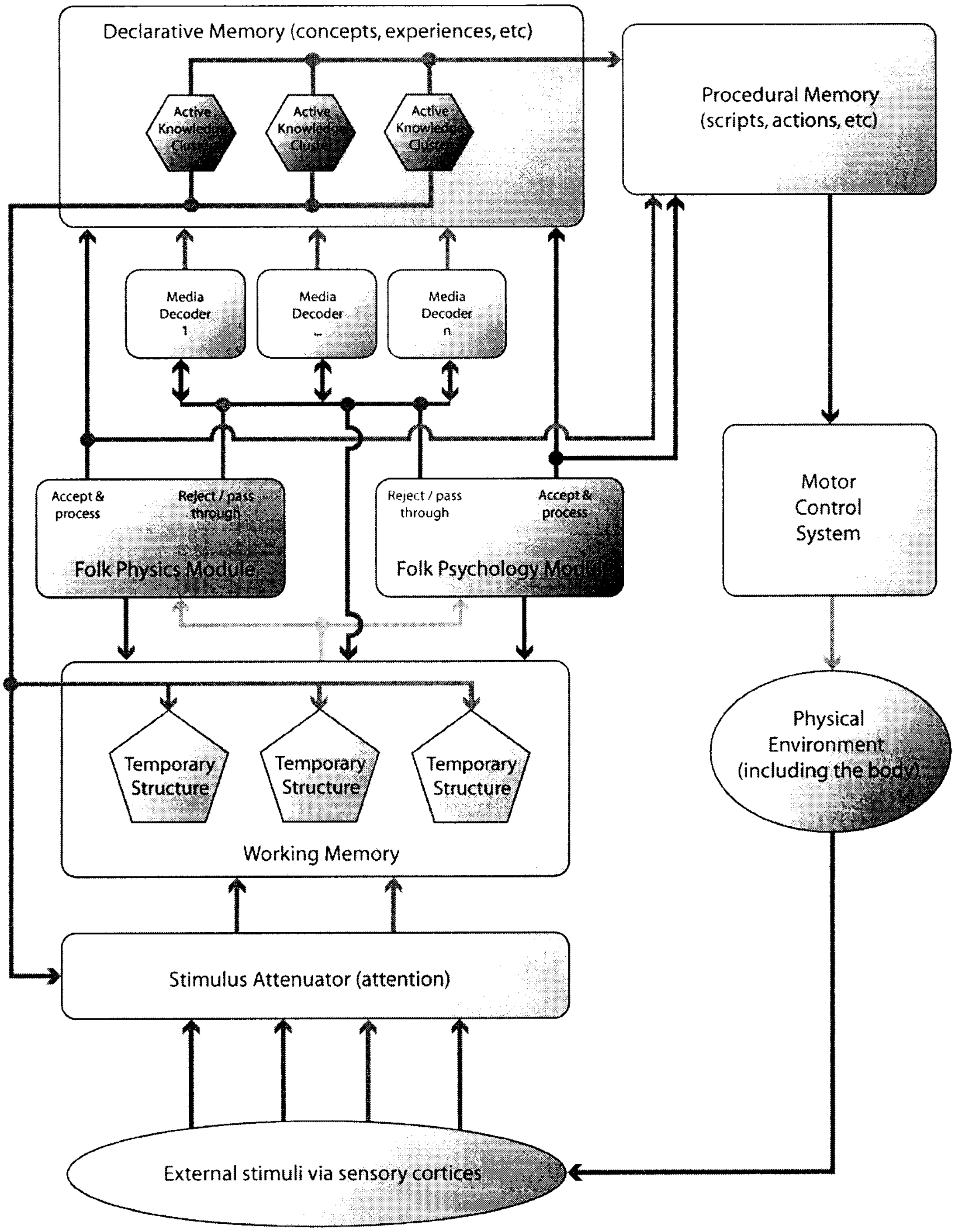

Figure 4.1: The structure of the CLCC model. Information flows only in the directions indicated by the arrows on the information paths (see section 4.3. for discussion). The three nodes in the yellow shaded region are the only nodes at which measurement is possible (all other nodes are cognitive abstractions implemented by unknown neural mechanisms).

Attention has been widely discussed in the presence literature (Riva & Waterworth, 2003; Slater & Steed, 2000; Waterworth & Waterworth, 2001; Wirth *et al.*, 2007; Witmer *et al.*, 2005). In the CLCC model, the stimulus attenuator models attention as an extremely simple mechanism which reduces the number of incoming stimuli while requiring only minimal processing to make the decision (in the tradition of Treisman, 1969). This mechanism allows the attenuator to model breaks in presence. Although the term 'break in presence' strictly applies to a subjective experience without reference to a particular cause, the most non-controversial uses of the term apply to situations where a subject suddenly experiences an expected stimulus, such as the tug of an HMD cable (Slater & Steed, 2000), a rendering glitch by the system (Brogni *et al.*, 2003) or an experimentally introduced artifact on the display (Vinayagamoorthy *et al.*, 2004). All of these situations pose that a stimulus which is incongruent to the semantic bias of the system (see 4.3.6 below), demands attention to itself and breaks the subject out of the experience. It is hardly surprising that a tug of a cable should lead to a break in presence; however, what is interesting about a break in presence is that given the current achievable levels of fidelity they do not *constantly* occur. In current VE systems, subjects must always deal with a large number of competing and conflicting stimuli. This is because the stimulus attenuator, with its automatic tendency to exclude moderately incongruent stimuli removes these artifacts from processing before they can adversely affect the experience. However, a sudden intense stimulus (especially if it is highly incongruent to the semantic bias of the system) will be selected for further processing, leading to a reconstruction and thus a break in presence. Note that not all intense stimuli will lead to a break in presence (this will only occur if that stimulus leads to a reconstruction) – if the intense stimulus is semantically related to the VE (a roaring jet passing overhead in a VE of an airport), then it may strongly reinforce the current construction.

Inputs from: The sensory cortices (from which it selects data), and from the active knowledge clusters in declarative memory which implement the semantic relevance bias.

Outputs to: Working memory, to provide perceptual (bottom-up) data for the construction of temporary structures.

4.3.3 Working memory

Description: This is a temporary buffer used during processing which maintains a mental representation of the current situation (Baddeley, 1986, , 1998). The CLCC model makes use of the working memory model of Baddeley rather than the more recent models such as that proposed by Cowan (2001) and Oberauer (2002) due to the extensive use of Baddeley's model in human factors research which has seen application in the presence field - for instance, see Bysrom et al's IPP model of performance in VEs (1999a), or Stanney *et al*'s review of human factors issues in presence research (Stanney *et al.*, 1998). The contents of working memory are accessible to consciousness (Baddeley, 1998; Rumelhart *et al.*, 1986), which makes it a potential site for presence measurement. Working memory's capacity is highly limited – only between five and seven meaningful chunks can be stored (Baddeley, 1986; Cowan, 2001). Working memory is not a unitary store – it has two known subsystems, which operate on specific modalities. The *phonological loop* is used for verbal (and auditory) information (Baddeley, 1986), and its primary purpose is processing of speech (Baddeley, 1998). It also plays an important role in reading (Carpenter & Just, 1989). The second subsystem is the *visuospatial sketchpad*, which is used to process visual and spatial data (Baddeley, 1986). It also plays an important role in the simulation of physical events such as predicting where thrown objects will fall (Cowan, 2001). Each of these systems has its own store of capacity, such that loading one will not affect tasks which make use of the other (Baddeley, 1986, , 1998). For example, a navigation task (which loads the visuospatial sketchpad) and remembering a set of numbers (which loads the phonological loop) will not interfere with each other. However, if one engages in remembering a set of numbers and reading a piece of text simultaneously, then performance in one or both of the tasks would be adversely affected, as the tasks are competing for verbal working memory capacity. Most cognitive processes require some working memory (with the exception of processes which have dedicated neural circuits, such as face recognition; see Andreasen et al., 1996).

Working memory can be usefully applied to presence, but has hardly received any attention in the literature. Baddeley's working memory model (1986; , 1998) has been used to explain phenomena which, from a micro-cognitive perspective, are similar to presence. These phenomena include reading comprehension (Carpenter & Just, 1989), and spatial navigation (Garden *et al.*, 2002). These tasks involve the processing of a subset of external stimuli so as to decode some meaning (be it spatial or otherwise) to allow further inferences about the space or about action in the space. If it is true that presence involves both perceptual and conceptual processing of an environment (Waterworth & Waterworth, 2001; Wirth *et al.*, 2007), and given that all processing requires some working memory (Baddeley, 1986), then it follows that for presence to occur, some amount of working memory will be required to process the environment.

How working memory is allocated to a task is a complex process. Simply having more information to process does not imply that more working memory will be allocated, or that more working memory will be needed. Data which are meaningfully related are automatically *chunked* into fewer, more abstract complexes, effectively freeing up working memory space (Baddeley, 1986). Thus, an immersive display which renders a large number of highly correlated variables (both in terms of perceptual synchronization and semantic organization) may require little working memory to process, as the information can be easily chunked. If a latency were to develop in the display of one of the channels (for instance, in sound rendering), then the temporal discrepancy would prevent the chunking of sound together with the other variables, and thus more working memory would be required to process the scene. This same mechanism can explain why stimuli which come from 'outside' the virtual environment (as with the radio position manipulation discussed in Slater *et al.*, 1995c) can reduce presence – due to the spatial discrepancy, these stimuli will not chunk with the stimuli 'inside' the virtual environment, and will thus require more working memory to process. If enough of these anomalies abound, then they will begin to impinge on the working memory which is necessary for successfully processing the VE. This reduction in the amount of processing focused on the VE will then lead to a reduction in the sense of presence.

Inputs from: The stimulus attenuator, which provides relevant perceptual (bottom-up) stimuli for processing. Secondary inputs are taken form the media decoders, and the folk psychology and folk physics modules, which use working memory as a scratchpad during processing.

Outputs to: To the folk psychology and folk physics modules, for semantic decoding.

4.3.4 Folk physics and folk psychology modules

Description: These modules recognize and respond to the contents of working memory. The folk physics module infers physical properties such as mass, velocity and spatial arrangement from these stimuli. It is also responsible for and accessing mental models and cognitive maps (Plotkin, 1998). The folk psychology module infers psychological properties, mental states and intentionality; it also responds to facial expressions, and is used in inferring emotion (Plotkin, 1998). As these modules are evolved for a particular purpose, they are highly specialized, and only respond to highly specific stimuli. The folk psychology module may allow for the bridging of individual forms of presence (the as presented in this form of the CLCC model) with social forms of presence.

The inclusion of these modules in the model is for two reasons: First, due to the empirical evidence from cognitive science that these modules are highly automatic and lead to important effects on many aspects of cognition including perception and memory (see Plotkin, 1998; Pinker, 2004); and second, on the suggestion on Lee (2004) who argued that these modules are involved in presence due to their tendency to automatically process even stimuli which crudely represent the objects they have evolved to process (this argument in detailed in section 2.2.2 in chapter 2).

Inputs from: Data are taken from the active temporary structures in working memory. These are tested against a minimal set of features to determine their suitability for processing in this module.

Outputs to: This module has two main output paths. The first path is used when the stimuli are not of the type processed by the module – in this case, the data is passed through to the media decoders for processing there. If the data are of the type to be processed, the data may be used to trigger a reflex action. For instance, the folk physics modules may trigger leaning into a curve during a virtual car ride (Freeman *et al.*, 2000), while the folk psychology module may trigger automatic following of eye gaze (Baron-Cohen, 1995). When an automatic response is required, the data are immediately passed to procedural memory for selection of a motor program for execution; otherwise, they are passed to declarative memory for semantic processing. A minor feedback path also exists from each module back to working memory, which allows the modules to make use of working memory as temporary storage during processing.

4.3.5 Media decoders

Description: Decoding the contents of working memory into semantic meaning is complex because the specific operations required vary depending on the medium used. Decoding a photograph of a room requires different cognitive processes to decoding a verbal description of that room, even though the final semantic products will be similar. In the CLCC model, decoding of media is done by learned cognitive modules (one per medium) whose inputs are data from working memory, and whose outputs are abstract representations of the content of the medium in declarative memory. Each of these media decoders is a collection of strategies and processes for decoding one particular medium. There are many possible decoders – a writing decoder (after evidence from Carpenter & Just, 1989), a film decoder (after evidence from Bordwell, 1989), a diagrammatic decoder (after evidence from Allmendinger, 1998), and so on. The decoder will only exist in a given subject if that person has learnt how to decode that medium. When a new set of perceptual stimuli are considered for processing, the appropriate media decoder is selected on the basis of a small set of key features (for instance, the basic shape of letters in declarative memory might trigger the writing decoder). Once the media decoder has been activated, it proceeds automatically with decoding the stimuli. If a media decoder attempts to decode the wrong type of medium (for instance, in the case of a picture being embedded in text as might occur in a magazine), then the error becomes a signal for the selection of a different decoder.

Media decoders, as processing units, require working memory. How much working memory a particular media decoder requires is a complex question. It seems reasonable to suggest that some decoders will require more working memory than others. For instance, images and video are relatively easy to decode (partly because there are dedicated neural circuits, in the visual cortex and other areas, specialized for this task - Krubitzer, 2005), whereas writing requires more resources to decode (a first a visual pass is required to decode individual letters and words, and a parallel second language decoding pass to decode the meaning of the sentences as a whole; Carpenter & Just, 1989). It also seems reasonable to suggest that some media decoders become more efficient with practice. For instance, reading requires time to learn and generally improves with practice, eventually becoming almost effortless (Carpenter & Just, 1989). Similarly, some film genres make use of conventions which must be learned at first, but are later decoded with little effort (Bordwell, 1989). From a working memory perspective, this increase in efficiency and associated sense of effortlessness come from a decrease in the amount of working memory used by the decoder as it becomes more efficient in chunking data (Baddeley, 1986).

Inputs from: Data are taken from the pass-through outputs of the folk psychology and folk physics modules.

Outputs to: The decoders have two outputs: One is to declarative memory, which is used to activate semantic meaning (the product of decoding). The second path is a feedback path to working memory used during processing. Media decoders can also shuffle data to other media decoders along a parallel bus. This will occur in two situations: when a decoder takes data but finds it cannot process it, or when decoding complex media (such as text embedded in a film), which requires two decoders to simultaneously process a scene.

4.3.6 Declarative memory

Description: This memory system stores two types of information: semantic information (including concepts and the relationships between concepts – Squire, 1999) and experiences (including the temporal organization and personal significance of experience – Squire *et al.*, 1993). Unlike working memory, declarative memory is effectively unlimited in terms of capacity (Squire, 1999). Declarative memory is used in the CLCC model in preference to the more specific subdivision into semantic and episodic memory suggested by Tulving (1995) for reasons of parsimony – there is currently not enough data to infer separate roles for semantic and episodic memory in presence.

When processing an environment, declarative memory associates semantic meaning to percepts, and contextualizes with respect to the subject's previous experiences.

Various models of declarative memory encoding exist, but the preferred concept in the CLCC model is that of schemata (Rumelhart *et al.*, 1986). A schema is a concept encoding structure which is modified by experience to contain default information (Rumelhart & Ortony, 1977). This allows the inference of missing information. For instance, if in a restaurant I see a man wearing a black bowtie and holding a notepad, I can infer that this man will take my order and I can pay the bill to him at the end of the evening (by activation of the 'waiter' schema). Schemata are connected in semantic networks, such that activation of one schema automatically leads to the activation of related schemata, thus creating semantic contexts (Rumelhart *et al.*, 1986). These small networks of activated schemata are termed *active knowledge* clusters in the CLCC model (see 4.3.10 below). The semantic context provided by active knowledge clusters propagates through the entire model: It is transferred directly to the stimulus attenuator, to act as the basis for the inclusion of stimuli by context relevance (a process termed semantic priming – Maxfield, 1997); and it is passed to working memory to allow for semantic based chunking, as well as provide meaning to the temporary constructions (partly leading to expectation based processing – Posner & Snyder, 1975). Because declarative memory contains almost exclusively learned information, it will be a source of a large degree of individual variation; however, this variation is not entirely idiosyncratic, as much semantic information is shared across populations such as cultural groups (Hirschfeld & Gelman, 1994) and professional groups (Goldman, 1986).

Inputs from: Media decoders and the folk-psychology and folk-physics modules, which provide perceptual inputs to be placed into wider semantic contexts.

Outputs to: The stimulus attenuator to implement the relevance bias; also to working memory, to ensure the formation of semantically coherent temporary structures. Finally, the system outputs to procedural memory, to give information about objects (hardness, weight, etc. as well as semantic information for speech and writing) which are required to plan and execute action in the context of the current semantic bias.

4.3.7 Procedural memory

Description: This node is responsible for storing sequences of behaviours (Squire *et al.*, 1993), including speech, walking, social responses such as greeting and maintaining eye contact, and simply conditioned reflex actions (Squire, 1999). It is included as a system separate to declarative memory in the CLCC model due to evidence that it is implemented by different brain circuits (Squire, 1999), and from clinical double-dissociation studies showing that lesions in particular brain regions negatively impacts declarative memory but not procedural memory, and vice-versa (Squire, 1999). Further evidence for the separation of these systems comes from Leeb *et al.* (2005) who showed that the *intention* to move a limb, even when actual movement is inhibited, generates a measurable EEG scalp potential. In the CLCC model, procedural memory has two roles: first, as suggested by Lee (2004), to initiate reflex reactions based on inputs from the folk psychology and folk physics modules, such as changing body posture (Freeman *et al.*, 2000), reaching for a virtual stimulus (Slater *et al.*, 1995c) or adjusting interpersonal distance in response to an agent (Bailenson *et al.*, 2001). The second role of declarative memory is the performing of goal-directed behaviours based on inputs from the folk physics and folk psychology modules. These behaviours include direct behaviours (such as reaching with an arm in response to an input to grab – see Waterworth & Waterworth, 2001) as well as symbolically mediated behaviours (such as clicking the mouse in response to an input to grab - Waterworth & Waterworth, 2001).

Inputs from: Temporary structures in working memory, which can act to inhibit as well as enhance the expression of behaviours based on the current semantic bias. A second input comes from the folk physics and folk psychology modules, to allow both goal-directed and automatic responses to objects. A final, indirect input is the set of active knowledge clusters in declarative memory, which allow for speech (by providing semantic information) as well as mediated behaviours (as when using an interface to mediate the VE).

Outputs to: The motor control system, for movement of the body or production of speech.

4.3.8 Motor control system

Description: This system contains the low-level neural and physiological controls for behaviour in the environment (including muscle movements and the motor aspects of speech). This node can be directly measured by means of physiological measures (as done by Meehan, 2001; see chapter 2 for a review of such measures). This system will also be of interest when considering behavioural presence measures (such as posture and sway – IJsselsteijn, 2004; see chapter 2).

Inputs from: Motor programs in procedural memory

Outputs to: Movement of the subject's body in the environment

4.3.9 Physical environment

Description: This is the physical environment, from which stimuli arise, and in which the subject interacts bodily. In many respects, this node represents embodiment, and underlines the importance of interaction in the world for presence (Lakoff & Johnson, 1999; Schubert *et al.*, 2002). With the exception of dreaming, where the reticular activating system circumvents normal perception (see 3.1.3), all stimuli arise from the physical environment, although, due to working memory limits, not all stimuli can be processed (Baddeley, 1986). Note that this is not the *mediated* or *virtual* environment (which has no explicit existence in this model – see 3.1.4 in chapter 3). This has two important implications: first, all stimuli (be they considered real or virtual) compete for processing on equal terms, without the benefit of semantic distinctions between them (these distinctions are only applied after a simulus has been selected for processing). Second, all motor movements made by the subject are expressed in the physical environment; movement in the virtual world occurs due to manipulation of the VE interface in the physical environment.

Inputs from: The motor control system, as well as from object to object interactions within the physical environment itself (which is a closed system in its own right).

Outputs to: Physical stimulation of the sensory organs which are converted into percepts by the sensory cortices, to enter at the stimulus attenuator.

4.3.10 Active Knowledge clusters

Description: These structures are highly transient, being created and discarded as the overall state of the model changes inside of a single experience. Although the CLCC model conceptualizes these clusters as objects, they consist of small, related clusters of active schemata in declarative memory (Rumelhart *et al.*, 1986; Rumelhart & Ortony, 1977). Due to the associative nature of declarative memory, active schemata automatically self-organize into meaningful clusters (as each concept is connected to semantically related concepts, allowing activation to spread between them - Rumelhart *et al.*, 1986). These clusters contribute significantly to the overall semantic relevance bias of the system, as they feed back to the stimulus attenuator (to filter out irrelevant stimuli), as well as to working memory, to provide semantic coherence to the current construction.

If active knowledge clusters stop receiving stimulation from working memory, their activation will gradually decay. One active cluster may also have its activation inhibited by another competing knowledge cluster, in which case it will decay more rapidly. (Rumelhart *et al.*, 1986; Rumelhart & Ortony, 1977). Regardless of how a knowledge cluster loses activation, it is not an instantaneous process. The more extensive the semantic network is which is activated, the longer it will take for the decay or inhibition to sweep across all connected schema (Rumelhart *et al.*, 1986). The amount of time taken for a cluster to lose its activation is referred to in the CLCC model as its *thematic inertia.* Clusters with higher thematic inertia are more likely to remain active and compete with other knowledge clusters, and are thus more capable of exerting a semantic bias over the system. Because thematic inertia is associated with the extensiveness of the particular schemata network, experts in particular content knowledge domains (who have more extensive knowledge networks – Ericsson & Lehmann, 1996) will have more thematic inertia for those content domains.

Inputs from: Bottom-up inputs from media decoders, the folk psychology module and the folk physics module. An active cluster can also pass activation (inhibitory or excitatory) to other, semantically similar clusters, via the networked structure of declarative memory.

Outputs to: Other knowledge clusters which are semantically related (both excitatory and inhibitory; see *inputs* above). The active knowledge clusters, as components of declarative memory, contribute to the overall semantic bias of the model (see 4.3.6 above).

4.3.11 Temporary working memory structures

Description: As with active knowledge clusters, these temporary structures are created and discarded in order to keep the overall system sensitive to changes in the environment. However, unlike the active knowledge clusters which represent activation patterns in fixed networks of knowledge, the temporary structures represent free-form clusters of data (Baddeley, 1986) which arise from the activation of declarative memory as well as from the use of working memory as temporary storage by the processing modules. They represent system state rather than structure; they represent all of the conscious information available to the subject (Rumelhart *et al.*, 1986). Due to the semantic bias transferred to working memory from declarative memory, these temporary structures are always constructed with a high degree of semantic coherence, and thus exert a great expectation bias on processing (Baddeley, 1986).

As these structures include both perceptual and semantic information, they are well suited to explaining content related effects in presence. In order for content to affect processing, the percept must first be decoded (requiring perceptual information) and then understood as meaningful objects (requiring semantic information). This is not the case with similar structures in other presence models, which incorporate very little or no semantic information (such as the SSMs of the MEC model – Wirth *et al.*, 2007). In the CLCC model, the set of temporary working memory structures is termed 'the construction of the environment' as these structures contain the conscious understanding which the subject has constructed by interacting in the environment (Bruner, 1990). It should be noted that although the temporary structures act to maintain semantic coherence, it is possible for an irrelevant stimulus to arrive from the stimulus attenuator demanding processing capacity (due to massive semantic dissimilarity). When this occurs, a sudden re-allocation of working memory occurs, and some temporary structures are discarded while others are changed drastically, leading to a new semantic interpretation of the scene – this is termed a *reconstruction*. Sudden reconstructions are associated with the break in presence experience, while gradual reconstructions (where working memory capacity is gradually allocated to tasks other than processing the environment) are associated with a *drift in presence* – see section 4.4.2 below.

Inputs from: Perceptual data arriving bottom-up from the stimulus attenuator, and conceptual data arriving top-down from the declarative memory system.

Outputs to: No direct output. These structures exist as data clusters in working memory, which are evaluated by the folk psychology module, the folk physics module or one or more media decoders.

4.4 Presence in the model

The CLCC model is strongly based on a general purpose cognitive model, as opposed to models which are specifically designed for explaining presence, such as those of Slater (2002), Biocca (2003), the FLS model of Waterworth and Waterworth (2001), and Wirth *et al.*(2007). Therefore, only the state of the model (and not the structure, which exists for processing in general) is pertinent when considering how presence occurs. In the CLCC model, presence is not a phenomenon of its own right, but rather side-effect of the model structure (as suggested by K. M. Lee, 2004). It is important to note that some of the structure of this model is inherent (such as the folk physics module, which is a species evolved trait), but other parts are acquired through learning (such as the media decoders). The state of the model is simply the distribution of activation among the structural components of the model. In this model, presence occurs mostly due to inference. At any moment, the system contains a set of active and semi-active knowledge clusters and behavioral scripts. Due to the associative nature of declarative memory, this pattern of activation allows contextualized inferences, and if any one inferential cluster receives enough activation to either lead to a temporary structure in working memory (a conscious thought) or to an overt behaviour, then one can say that that a particular cluster was 'selected' for expression (although the selection requires no more mechanism that the accumulation of activation and competition between active knowledge clusters).

The CLCC model implements presence as the perceptual illusion of non-mediation (Lombard & Ditton, 1997). This occurs when the subject is selecting a subset of stimuli (those encoding the virtual environment) from which to regulate their cognition and behaviour (this also owes something to the cognitive presence concept - Nunez & Blake, 2001). When a subject focuses on a VR display, processing resources