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).

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.

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).

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’.]

---

56
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.

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 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, 1998). 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 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 than 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 the 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 from 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.

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.

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 cannot 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., 2000). 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 manage 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 streams, and were instructed to focus on the display rather than 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
situations 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 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).

![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.](image)

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.

2. Controlled attention path – to reflect the importance of motivation on attention 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.](image-url)
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
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 activities (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 (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 an 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 straightforward. 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. 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.