Dimensions of Congruity in Immersive Virtual Environments:

A Framework for the Schematic Processing of Multimodal Sensory Cues

Tiernan J. Cahill

Boston University

Author Note

Address for correspondence:

Boston University
College of Communication
640 Commonwealth Avenue
Boston, MA 02215
United States
1-617-221-3824
tjcahill@bu.edu

Abstract

Content features of immersive virtual environments have previously been found to play a significant role in affecting users' sense of presence – a role that will grow only more important as technology develops and the fidelity of simulations improves. Seeking to understand the unique challenges associated with designing experiences that are both compelling and immersive, this short paper offers an explication and typology for the concept of *congruity*, consisting of three dimensions: *sensory*, *environmental*, and *thematic*. Congruity is grounded in the understanding that sensory stimuli in virtual environments will inevitably activate internalized schemata that relate sensory inputs from real environments. Each dimension of congruity addresses a different level of cognitive association within and between virtual objects. Underlying the concept is a pragmatic framework for articulating how multimodal cues interact to affect the perceived plausibility of virtual environments. This may serve as a basis for future empirical research examining presence in relation to platform-independent content, as well as a more comprehensive view of the phenomenological differences between immersive virtual environments and other mediated experiences.

Keywords: immersive virtual environments; virtual reality; sensory cues; cognitive schema; genre; presence

Dimensions of Congruity in Immersive Virtual Environments: A Framework for the Schematic Processing of Multimodal Sensory Cues

As the technological capabilities of virtual, augmented, and mixed reality platforms continue to improve, the distinctions between those experiences that are successful in eliciting presence and those that are not is likely to become more dependent on content features. In other words, *what* is being simulated has the potential to be just as theoretically significant to the study of presence as the technological fidelity of the simulation. In order to better articulate and address the challenges that arise when developing content for immersive virtual environments (IVEs), this paper offers a new theoretical framework for understanding the cognitive associations between real and virtual objects in the context of multimodal sensory input.

Background

The appeal of virtual, augmented and mixed reality is deeply rooted in the potential ability of the technology to produce in its users a sense of presence in the virtual environment (Bowman & McMahan, 2007). To date, there has been extensive empirical investigation focussing on the technological features and affordances of virtual reality hardware, including the latency of rendering, resolution of head-mounted displays, and degrees of freedom allowed by motion tracking systems, and the effect that these and other features have on the sense of presence experienced by the user (Cummings & Bailenson, 2016). In contrast to the current state of understanding regarding the role of technological factors, there remain many questions surrounding the relationship between content features of the virtual environment and the experience of presence.

Content Features

From the research that has been conducted in this area, it is apparent that environmental content does matter: A study by Baños et al. (2004) found that an environment designed to induce sadness[1] was associated with higher user-reported levels of engagement, naturalism, believability, and reality than a comparable neutral environment. Similarly, Bouchard, St-Jacques, Robillard, and Renaud (2008) found that participants immersed in an environment containing anxiety-inducing elements[2] reported higher levels of immersion than those situated in a similar environment that was absent of such elements. Background content features of IVEs, such as light, shadow, and music, have also been found to affect the experience of presence (Riva et al., 2007).

In addition to these affective factors, a user's sense of presence is presumably determined, in part, by those features of the virtual environment that support or undermine particular cognitive interpretations. It is important to emphasise, at this juncture, that cognitive processing in a truly immersive virtual environment is fundamentally different from the processing of other media stimuli. On a phenomenological level, the simulated reality of an IVE is not situated *within* an extant perceived reality in the same way as images displayed on a television or computer screen; instead, the sensory stimuli of the simulation are capable of occluding those of the external reality almost in their entirety. Even in the most advanced 3D IMAX movie theatre, the audience is still always presented with abundant sensory cues that what they see on screen is not a real environment, but rather, a series of images projected onto a screen. At least some IVEs are capable of stripping away these cues, however, leaving only those provided by the simulation.

There is an essential point of distinction between virtual reality, on the one hand, and mixed and augmented reality on the other. Whereas the former ideally presents the user with a complete, alternate set of sensory inputs with which to construct an alternative perception of their surroundings, the latter intentionally integrate elements of the external reality and, as such, the perceived environment is always at least partly anchored in the physical environment.

Plausibility Illusion

In the absence, then, of competing or external sensory information, how is an IVE cognitively processed and evaluated such that a user feels themselves to be more or less present within it? In conjunction with *place illusion*, which aligns closely with conventional definitions of presence, Slater (2009) has proposed a supporting concept of *plausibility illusion*: Whereas place illusion describes the sensation of being in a real place and, Slater argues, is primarily tied to the technological affordances of the virtual reality system, plausibility illusion expresses the sense that "what is apparently happening is really happening (even though you know for sure that it is not)" (p. 3553). Within this framework, plausibility illusion is operationally independent of the fidelity of the simulation, having more to do with the behaviour of virtual objects and actors than with the way in which they are rendered. Indeed, Slater, Khanna, Mortensen, and Yu (2009) found empirically that the absence of certain visual cues, such as shadows, negatively affected the illusion of plausibility, independently of graphical fidelity.

Schematic Association of Sensory Cues

The cognitive evaluation of the plausibility of a simulated environment is based, in large part, on normative expectations of the behavior of objects in the physical world. As illustrated in the above example, such expectations can be as fundamental as: *A solid object that is placed in front of a light source casts a shadow*. Schematic interpretations of sensory cues are internalized at a young age, as observations of commonplace phenomena are reified into beliefs about causality and expectations of future behaviour (Piaget, 1930). The salience of schematic understandings in IVEs can be seen in the way even deeply irrational beliefs, such as extreme paranoia, have been observed to transfer easily to simulated contexts (Freeman et al., 2008).

Likewise, the impulse to conform to existing schemata may partially explain the tendency for users of IVEs to map mediated stimuli onto their own body, such as tactile sensations in response to seeing a virtual hand being stroked (IJsselsteijn, De Kort, & Haans, 2006) or the perception of physical resistance when pulling on a virtual spring (Biocca, Kim, & Choi, 2001). On this level, conceptualizations of self-presence appear to be intimately related to bodily schemata (Biocca, 1997).

There is substantial evidence to suggest that the cognitive processing of physical spaces is similarly influenced by schematic models. Common-sense understandings of how physical objects behave and interact with each other rely on heuristic, probabilistic models, grounded in past observations, which allow humans to make rapid, intuitive inferences about how objects are likely to behave in the future (Battaglia, Hamrick, & Tenenbaum, 2013).

The first step to integrating these perspectives into the current understanding of plausibility is to provide a theoretical framework for articulating how these schemata inform the processing of sensory cues on different levels of engagement with the virtual environment. The explication of conceptual structures allows for the joining of lower-order concepts with more complex ones, enabling theoretical development that is at once grounded in data and capable of attaining great explanatory and predictive power (Chaffee, 1991). Furthermore, it sets the stage for more systematic and productive empirical research, by allowing social scientists to develop consensus around terminology, measurement, and interpretation.

Congruity at Three Levels

The concept of *congruity* is essentially a function relating two sets of sensory cues within an IVE and their corresponding schematic associations. These schemata are presumed to be both internalized on the basis of experience and, with few exceptions, normative: They are organizing principles for the world *as it should be*. They are also *heuristic* criteria for evaluating sensory input and, as such, an individual is probably only aware of them when the criteria are violated, leading to a sense of unease, disorientation, or confusion. Thus, while the concept of *congruity* is more easily explicated in positive terms, instances of *incongruity* are much more likely to be consciously experienced and reported by users of IVEs.

Congruity, as a theoretical construct, may be divided into at least three dimensions on the basis of the scope of sensory cues being compared, and whether they relate to a single object, a set of objects, or the entire perceived environment, as well as the particularities of the associated schemata.

Sensory Congruity

Sensory congruity describes an adherence to the expected relationship between corresponding sensory cues of different modalities as they relate to a single virtual object. In physical reality, objects are often perceived simultaneously through multiple sensory modalities, and the relationships between different modal cues are generally stable. There may be cases where an object is only observed via one modality: a distant waterfall, for example, too far away to be heard or felt. However, in those cases where multimodal cues are available, they are presumed to all have cognitive associations in common: If one approaches the waterfall to the point where it comes within earshot, then one will expect to hear rushing water and not some other sound emanating from the same visible source.

While there may not be comprehensive suites of sensory cues associated for all familiar objects, those associations that do exist are likely to be very deeply internalized, rooted as they are in the ordinary experience of engaging with the physical world through multiple streams of sensory input. As such, the maintenance of sensory congruity is hypothesized to constitute a baseline condition for plausibility illusion. There may be some tolerance for missing sensory information; however, incongruous sensory information from an IVE, or mixed sensory cues from the virtual and real environments, may be disastrous to the user's sense of presence. For instance, while many IVEs are able to successfully maintain a sense of presence without incorporating haptic feedback, the disruptive effect of haptic cues that are incongruous with visual stimuli will be familiar to any user who has inadvertently bumped into an "invisible" wall or piece of furniture while wearing a head-mounted display.

Environmental Congruity

Environmental congruity describes, broadly, the extent to which the behaviours of objects within an IVE conform to schematic expectations of the behaviour of comparable objects in real environments. These schemata are common to all objects in the environment and, thus, constitute a heuristic representation of the laws of physics. Through observation, humans are able to develop an innate understanding of the fundamental principles of physics, which allows them to make intuitive predictions of how objects will behave in physical reality (Battaglia et al., 2013). For example, seeing a weight balanced upon a ledge, an observer requires relatively little cognitive effort to predict, based on the apparent shape and density of the object, whether or not it is likely to fall.

This is not to say that such heuristic predictions are always accurate; the observer may not be privy to all of the relevant information or may misjudge relevant variables. Yet, as with all instances of incongruity, the effect when observed objects fail to conform to schematic understandings is uncanny. In physical reality, this effect can be observed in cantilevered architectural features that appear to defy gravity. Scott (1914) writes that such buildings cause discomfort, not out of fear that the structure might actually collapse, but because they undermine the observer's own sense of internal stability, and are thus associated with weakness and uncertainty.

From a design perspective, the imperative to reduce environmental incongruity is already apparent in the substantial investments made by game development firms in high-precision computational engines for simulating the physics of virtual objects in real-time. These engines are both challenging to engineer and demanding to operate, yet they are necessary to narrow the distance between schematic understandings of the behaviour of objects in the physical world, and the behaviour of their virtual counterparts. This is also suggestive of the ways in which technological platform features may continue to act as a constraint on content and thus, indirectly influence the ability of the simulation to maintain environmental congruity.

Thematic Congruity

Finally, *thematic congruity* articulates the understanding that certain classes of objects will have normative internal associations, according to thematic schemata. The basis for these associations may be drawn from personal experience but is just as likely to be found in cultural texts or artifacts. In either case, the schemata describe sets of objects that are understood to belong together.

Thematic schemata are closely related to conceptualizations of *genre* in entertainment media. While genre is often conceived of as the property of a given text, it may also be considered as a functional designation related to the cognitive structures that the text invokes in the reader (Schmidt, 1987). Under this framework, a film might be classified as a Western, not because of any stable, deterministic trait that flags the text as such, but because of the conglomeration of objects typically grouped under this heading: Stetsons, revolvers, stagecoaches, and so on. The introduction of an object that is not typically be part of this grouping – a helicopter, for instance – would be immediately noticeable and, most likely, palpably unsettling. There is, of course, no physical reason why a stagecoach and a helicopter could not appear side-by-side in reality; yet, an awareness exists that this shouldn't happen for the simple reason that it doesn't usually happen in either the recalled experiences or the cultural representations of either object.

There are instances where authors may intentionally introduce thematically incongruous elements into a created environment. Yet this is rarely, if ever, done with the intent of maintaining plausibility. Rather, genre-bending and genre-breaking elements are typically employed to bring the composition of the underlying thematic schemata to the forefront of the audience’s awareness (Berliner, 2001). They are intended to unsettle, to deconstruct, and to disinvolve. In the context of less immersive media, such as film and television, this may allow the audience member to step back into physical reality and to consider the text before them *as* a text, rather than as a phenomenon. It is not immediately apparent whether this should also be true of users of IVEs, given the absence of a secondary set of sensory cues to "step back" into. Whether the disruption of thematic congruity would be sufficient for users to ignore other, more fundamental, sensory cues, or whether the experience would only be mildly disorienting, is a question that demands empirical study.

Discussion

The dimensions of congruity explicated above are useful to the ongoing research of presence factors in IVEs in a variety of ways: Most immediately, they provide a stable theoretical framework for manipulation of exposure conditions in experimental studies. The comparison of presence indices across environments with high and low levels of sensory, environmental, and thematic congruity could lead to a better understanding of user tolerances for incongruity, allowing designers and engineers of IVEs to better adapt to the cognitive needs of their users.

More broadly, the concept of congruity serves as a theoretical bridge between three perspectives: An IVE can be viewed as a design object or as a cultural text: a series of decisions made by an author with a particular communicative intent. Alternatively, the environment can be conceptualized as a set of sensory stimuli that are perceived and processed in parallel by a user, leading to certain cognitive and emotional responses. Finally, the environment can be addressed philosophically as an alternate reality with its own ontology: comparable, on a phenomenological level, to the physical world. By recognizing congruity as a set of normative organizing principles that originate in phenomena, persist in cognition, and inform design, the framework presented here moves towards a pragmatic synthesis of these three approaches to understanding virtual environments.

As with any new theoretical construct, the framework is not without limitations: In particular, in its current form, congruity is only applicable to IVEs that occlude entirely sensory input from the external, physical reality, which functionally limits the scope of the construct to VR platforms. That said, it is possible that the theory could be further developed to address dimensions of congruity between co-extant virtual and real objects, and thus be applied to mixed and augmented reality contexts as well.

Additionally, while congruity is discussed above in mostly normative terms, on the premise that the experience of presence is a desirable affordance in most IVEs, it is essential to recognize that this may not always be the case. If IVEs are to be treated as a medium for creative, even artistic, expression (and there is no reason they should not be), then there may be cases in which *incongruity* becomes a useful tool with which to provoke or unsettle the audience.

It is also conceivable that certain thematic schemata may be associated with lowered demand for congruity in other dimensions. Environmental congruity, in particular, is predicated upon a certain level of familiarity with the physical laws governing a given environment; this familiarity is nearly universally transferable in the everyday experience of physical reality. However, certain thematic cues, such as those associated with science fiction or fantasy, may signal to the user a break from the familiar, and thus, a greater tolerance for incongruity.

Finally, it must be acknowledged that, while the conceptual explication laid out above hints at some practical example of the three dimensions of congruity, the actual business of creating complete operational definitions has been left to future scholarship. This decision was motivated by an awareness that the conceptual definitions, as outlined, each admit a broad range of potential operationalisations. The choice of how to operationalise each dimension, in practice, will depend mainly on the research context and the logistical constraints of developing immersive stimulus content.

References

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and emotion: Their impact on the sense of presence. *CyberPsychology & Behavior*, *7*(6), 734–741.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–32. https://doi.org/10.1073/pnas.1306572110

Berliner, T. (2001). The genre film as booby trap: 1970s genre bending and The French Connection. *Cinema Journal*, *40*(3), 25–46. https://doi.org/10.1353/cj.2001.0006

Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, *3*(2). https://doi.org/10.1111/j.1083-6101.1997.tb00070.x

Biocca, F., Kim, J., & Choi, Y. (2001). Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence*, *10*(3), 247–265. https://doi.org/10.1162/105474601300343595

Bouchard, S., St-Jacques, J., Robillard, G., & Renaud, P. (2008). Anxiety increases the feeling of presence in virtual reality. *Presence: Teleoperators and Virtual Environments*, *17*(4), 376–391. https://doi.org/10.1162/pres.17.4.376

Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: How much immersion is enough? *Computer*, *40*(7), 36–43. https://doi.org/10.1109/MC.2007.257

Chaffee, S. H. (1991). *Communication concepts 1: Explication*. Newbury Park, CA: SAGE.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272–309.

https://doi.org/10.1080/15213269.2015.1015740

Freeman, D., Pugh, K., Antley, A., Slater, M., Bebbington, P., Gittins, M., … Garety, P. (2008). Virtual reality study of paranoid thinking in the general population. *The British Journal of Psychiatry*, *192*(4), 258–63. https://doi.org/10.1192/bjp.bp.107.044677

IJsselsteijn, W. A., De Kort, Y. A. W., & Haans, A. (2006). Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed relaity. *Presence*, *15*(4), 455–464.

Piaget, J. (1930). *The child's conception of physical causality*. (M. Gabain, Trans.). London, UK: Kegan Paul, Trench, Trubner & Co., Ltd.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., … Alcañiz, M. (2007). Affective interactions using virtual reality: The link between presence and emotions. *CyberPsychology & Behavior*, *10*(1), 45–56. https://doi.org/10.1089/cpb.2006.9993

Schmidt, S. J. (1987). Towards a constructivist theory of media genre. *Poetics*, *16*(5), 371–395. https://doi.org/10.1016/0304-422X(87)90028-3

Scott, G. (1914). *The architecture of humanism: A study in the history of taste*. Boston, MA: Houghton Mifflin.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of The Royal Society B*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Khanna, P., Mortensen, J., & Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *IEEE Computer Graphics and Applications*, *29*, 76–84. https://doi.org/10.1109/MCG.2009.55

Footnotes

[1] Baños et al. used a combination of music, narrative, Velten self-statements, pictures, movies, and autobiographical recall, all embedded in different areas of a virtual park, in order to induce sadness in participants. Comparable, yet less affectively charged, elements were embedded in a similar environment for the neutral condition.

[2] Participants in the study suffered from a specific phobia of snakes (ophidiophobia). Anxiety was thus induced in the study by informing them that the virtual environment contained a number of poisonous and aggressive snakes.