

Chapter 1

Constraint, Intelligence, and Control Hierarchy in Virtual Environments

Thomas B. Sheridan
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
U.S.A.

1. Introduction

This paper seeks to deal directly with the question of what makes virtual actors and objects that are experienced in virtual environments seem real. (The term *virtual reality*, while more common in public usage, is an oxymoron; therefore *virtual environment* is the preferred term in this paper).

Reality is difficult topic, treated for centuries in those sub-fields of philosophy called ontology—“of or relating to being or existence” and epistemology—“the study of the method and grounds of knowledge, especially with reference to its limits and validity” (both from Webster’s, 1965). Advances in recent decades in the technologies of computers, sensors and graphics software have permitted human users to feel *present* or experience *immersion* in computer-generated virtual environments. This has motivated a keen interest in probing this phenomenon of presence and immersion not only philosophically but also psychologically and physiologically in terms of the parameters of the senses and sensory stimulation that correlate with the experience (Ellis, 1991). The pages of the journal *Presence: Teleoperators and Virtual Environments* have seen much discussion of what makes virtual environments seem real (see, e.g., Slater, 1999; Slater et al. 1994; Sheridan, 1992, 2000).

Stephen Ellis, when organizing the meeting that motivated this paper, suggested to invited authors that “We may adopt as an organizing principle for the meeting that the genesis of apparently intelligent interaction arises from an upwelling of constraints determined by a hierarchy of lower levels of behavioral interaction.” My first reaction was “huh?” and my second was “yeah, that seems to make sense.” Accordingly the paper seeks to explain from the author’s viewpoint, why Ellis’s hypothesis makes sense. What is the connection of “presence” or “immersion” of an observer in a virtual environment, to “constraints” and what types of constraints. What of “intelligent interaction,” and is it the intelligence of the observer or the intelligence of the environment (whatever the latter may mean) that is salient? And finally, what might be relevant about “upwelling” of constraints as determined by a hierarchy of levels of interaction?

2. Optimization as Related to Constraint

In theory, greatest goodness, happiness, degree or success, or, in technical parlance, utility, are a matter of simultaneous solution of two kinds of equations:

1) *Objective (utility) function:*

Goodness of solution = an explicit function of salient variables, e.g., dollars, time, accuracy, etc.)

2) Certain *given constraint equations* apply, for example:

- The dollars spent must be less than some total budget
- The time spent must occur before some deadline
- The performance = some explicit function of all the salient variables (e.g., dollars, time, constants.)

Formally, the optimal solution is found by maximizing goodness (utility) under the given constraints. This operation is standard practice in control engineering, in operations research, in econometrics, in engineering design, and in any endeavor where it is possible to express the objective function and the constraints in mathematical form.

While we cannot be so explicit about the mathematics and therefore declare optimality, efforts to design good environmental interactions for humans must take explicit account of constraints in much the same manner as above. In fact it will become apparent that what we call “intelligent” and “realistic” is precisely because certain environmental interactions that we have evolved and refined (i.e., optimized, at least in the sense of *satisfied* (made good enough for practical present use in our everyday life) conform to certain constraints. Examples will be cited in:

- Natural language
 - Music
 - Body movement (dance and athletics)
 - Computer programming and supervisory control
 - Computer displays
- ...and finally, by analogy,
- Virtual environments

2.1 Constraints of Spoken and Written Natural Language

Think about natural language. Random words spoken or written are just babble, and communicate nothing, even though with randomness their Shannon information measure ($\sum p_i \log p_i$, where p_i is the probability of speaking any particular word) is maximum. Spoken language makes “sense” to us precisely because it conforms to rules of syntax. Chomsky has shown, in fact, that these rules – constraints – are to a great extent programmed into our genes.

This is true of both spoken speech and writing. In fact we have evolved many additional constraints in order that we may communicate “meaning” to one another (grammar, spelling, presentation format). Without adherence to accepted constraints we simply do not understand the message, and even small errors detract and are regarded as foreign or unintelligent. Language is a coding scheme shared by the sender and the recipient, and without compliance to the constraints of the code there is no way for the recipient to select the meaning that the sender intends.

2.2 Constraints of Music

There is nothing much rational about music. “Good” music is what people like—esthetically. Random sound is not interesting. It is just noise. Music must adhere to constraints in frequency range (pitch) and sound intensity (loudness) in order to be heard at all and not damage our organs of hearing. These ranges are constrained to only a few octaves only a few dB respectively. The constraint of harmony makes it interesting (disharmonies are used only sparingly). Perhaps less obvious is the constraint of tempo (or rhythm or beat) of different notes: between 1 and 5 Hz. Slower or faster tempos would not even be considered music.

2.3 Constraints of Body Movement: Dance and Athletic Games

To be called dance body movements must be “graceful” (meaning smooth continuous motion of extended limbs). It must conform to a natural tempo of the body, one corresponding to natural frequencies of flaccid limbs (again between 1 and 5 Hz). Otherwise it is boringly slow (maybe Tai Chi goes to 0.2 Hz) or at too “forced” a pace.

Athletes, in order to make strong and precise movements (whether in running, jumping, throwing, hitting with a bat or racquet, swimming or gymnastics) must abide by constraints of the body (limb forces and reaction times), again in the 1-5 Hz range. Of course they are also bound by the constraining rules of the game (established to be consistent with the body’s own constraints on what is achievable).

2.4 Constraints in Design of Computer and Graphic Displays

Many studies have shown that when too much information is presented in a computer display much less information is communicated than if there were less information (i.e., fewer words or graphics on the same page or screen, less clutter). We all recognize the tendency to overdo the PowerPoint slide. It is apparently done because we can: The application program makes it so easy to add great variety in color, line widths, shading, logos, etc. This addition of what Tufte (1982, 1997) aptly calls “chart-junk” serves no good purpose, usually confuses the observer and actually reduces the effectiveness of communicating the essential message.

The chaos of trying to interpret the alarm tiles in a nuclear power plant control room during a emergency is often cited as an example of too much information too fast. In a test conducted at a nuclear plant simulator at one minute after a coolant major coolant pipe break I counted over 1000 alarm tiles lit up, and after a second minute 800 more. The nuclear plant operators claim that under such circumstances they have to rely on what they call “pattern recognition” to make any sense of what is going on. Clearly too much information, probably much of it inappropriate to the situation, is being presented for the constraints of human perception/recognition (mostly because too little thought has gone into integrating the information). (Traditionally each alarm tile is independently engineered to specify abnormality of every measured variable, and there is inadequate coordination among suppliers).

In a simple laboratory experiment in selecting one of a small number of control actions in view of uncertain information and time pressure, Roseborough (1988) showed that it was much better to display only point estimates than two-dimensional probability density functions.

Moral: Less can be more; constraints can be helpful.

2.5 Constraints of Computer Programming and Supervisory Control

Computer programming is accomplished in conformance to a highly constrained software language and operating system. The slightest discrepancy in programming disables the program with respect to its intended purpose, often showing up late in the process as a “bug.”

Supervisory control (Sheridan, 1992) means a human communicates task goals and constraints to a computer that in turn works within already programmed task goals and constraints (as well as constraints of a dynamic electromechanical system) to perform the given task. This form of control is characteristic of telerobots, automated factories, automated vehicles such as modern commercial aircraft, and many other modern systems. Feedback is via a highly constrained (and usually abstract) translation of events. The paradigm of supervisory control is shown in Figure 1 as a four level hierarchy of control loops. Each block is the controller for that block immediately below. The downward arrows from any block are control commands for the block immediately below, and the upward arrows from the lower block are the feedback to the upper block. The properties (capabilities) of each lower block are the constraints on the block immediately above. The computer block is “intelligent” to the human operator insofar as it does a good job of controlling what is subordinate to it and provides feedback that conforms to the intentions and expectations of the human operator. We might say that the actuator agent appears “intelligent” to the computer insofar as it does a good job of controlling the task and provides feedback that conforms to its intentions and expectations.

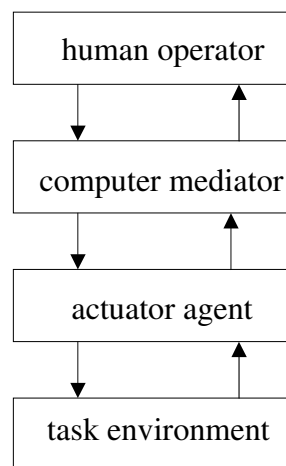


Figure 1. Hierarchy of supervisory control. (Each of computer, actuator and task environment imposes its own constraints. Constraints “well up” from below.)

Such a hierarchical paradigm applies not only to the human supervision of computer-mediated systems such as those mentioned above, but also to other forms of human interaction with the environment, where the blocks subordinate to the human are of other forms. For example in a natural language conversation between persons A and B (Figure 2), the second block down might represent the speaking/hearing constraints imposed on the cognition of the speaker A, which in turn feeds the hearing and gets feedback from the speaking constraints of B. From A's speaking perspective each lower block represents subsidiary constraints.

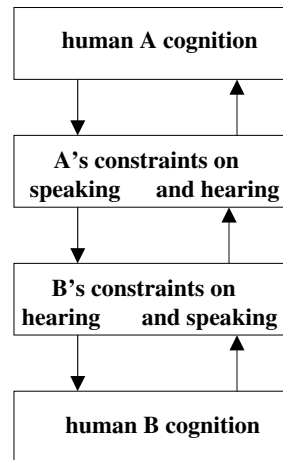


Figure 2. Constraints of interaction in social communication.

Having considered how human interaction with an environment, whether inanimate or social, succeeds as a function of conformance to the existing constraints, let us turn now to virtual environments.

3. Constraints That Apply in Virtual Environments

Seven types of constraints that apply in virtual environments are:

1. Sensory range and resolution of observer (absolute and differential thresholds)
2. Observation point consistency in space and time
3. Continuity of kinematics in space and time
4. Cause and effect
5. Mechanical impedance interaction with the observer/user
6. Symbolic interchange (words, gestures)
7. Etiquette (application of Grice's four maxims)

This is surely not the only credible taxonomy nor does it purport to be a complete set of constraints. There are many other factors that constrain what a virtual environment must be, given today's technology and today's human observers.

3.1. Constraints on Sensory Range and Resolution

This is the psychophysical set of constraints. A “sensorium” such as is depicted in Figure 3 is a region in the space of salient stimulus variables (in whichever of the five usual sensory modes or the much larger number of modes classified by Boring (1950) that can be sensed). While stimuli are commonly specified by intensity and frequency, there can be many other dimensions depending on the sensory mode, such as spatial distribution on the retina or skin, chemical makeup of the olfactory and gustatory stimulus, and temporal patterns for all the senses.

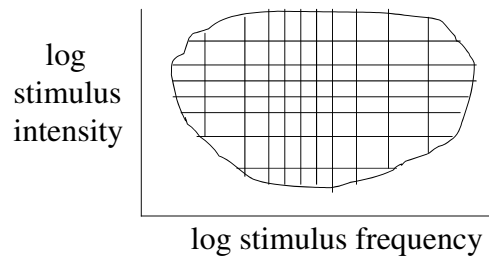


Figure 3. Stimulus sensorium in two of the (possibly many) dimensions.

3.2 Observation Point Consistency in Space and Time

The observer of a virtual environment often changes the viewing distance and angle of viewing. Indeed, to achieve a sense of presence and immersion relative to an object, such a change is often necessary. This is provided, of course, that the object size and orientation and lighting relative to the light sources remain geometrically correct during and following such a change. Achieving that correctness is a key task of the virtual environment designer.

If the observer is wearing a head-mounted display (HMD) it is important that the seen image correspond immediately to head position. Older HMD virtual environment systems exhibited a very noticeable time delay between when the observer suddenly turned his head and when the proper image appeared. Even current systems have small delays due to delays in the head-tracking hardware and image generation software.

3.3 Constraints on Continuity of Kinematics in Space and Time

Kinematics (relative motion of body or object segments) must be continuous in space and time and true to what is being simulated. Because the real world is continuous, virtual visual images that appear jerky will be discounted by the observer immediately. An infinitely high resolution in space (infinite number of pixels and polygons) and time (infinitely high speed) is obviously not possible and not necessary, but these resolutions must approach thresholds of discrimination in order to appear real. We are not there yet with today’s computers, but getting close. Achieving a sufficient degree of continuity is probably the greatest challenge with respect to computer hardware.

3.4 Constraints on Cause and Effect

Dynamic interaction between the observer and sensed objects, as well as such interactions between two or more elements in the virtual environment, should play out with a cause-and-effect relations consistent with the known laws of physics. For example, if moving object A collides with fixed object B (either because the observer is controlling A or A is seen to be moving on its own), then B should either move or be dented and A should bounce off. If a liquid spills it should splash, and if an object is dropped into a liquid it should make proper waves, etc.

3.5 Constraints on Mechanical Impedance Interaction with the Observer

Newton's law says that for any action there is an equal and opposite reaction, and there are other well known laws of mechanics. Accelerating a mass, sliding an object along a friction surface, or compressing a spring—all result in a force imposed back on the limb or real object doing the moving. A realistic virtual environment would provide force feedback in a corresponding manner—both resolved kinesthetic force to the muscles and joints (as provided by a master manipulator arm) and tactile patterns (forces distributed in time and space, as provided by a hand worn tactile display such as the Dataglove). Current virtual environment systems may provide one or the other, but not both in combination.

3.6 Constraints on Symbolic Communication with an “Intelligent” Entity: Etiquette

An advanced virtual environment allows for meaningful symbolic dialog with a virtual “intelligent being”—either by speech or gesture or written language—much as in a Turing test (where the challenge is to have a computer-based conversation with a human such that the human cannot distinguish the computer from another human). This moves us beyond the typical image of a virtual environment as one of a static physical space and into one that includes other intelligent entities that can interact linguistically with the user/observer. These entities may appear to be humans of a known and present culture, they could be humans of a different culture from the past, or they could be imagined intelligent robots from outer space in a science fiction virtual environment.

Beyond the kinematic/dynamic constraints on gesture, or the grammar constraints on speech or written language, the intelligent being must comply with the constraints of etiquette—the social conventions that have evolved in civilized societies over centuries to enhance cooperation and smooth the communication interactions between people.

Grice (1975) postulated four maxims for cooperation in human-human conversation:

1. Maxim of quantity: Say what serves the present purpose but not more.
2. Maxim of quality: Say what you know to be true based on sufficient evidence.
3. Maxim of relation: Be relevant, to advance the current conversation.
4. Maxim of manner: Avoid obscurity of expression, wordiness, ambiguity, and disorder.

Miller (2000) has applied Grice's axioms to human-computer cooperation, especially adaptive user interfaces. His rules, in abbreviated form are:

1. Make many conversational moves for every error made.
2. Make it very easy override and correct any errors.
3. Know when you are wrong, mostly by letting the human tell you.
4. Don't make the same mistake twice.
5. Don't show off. Just because you can do something does not mean you should.
6. Talk explicitly about what you are doing and why. (Your human counterparts spend a lot of time in such meta-communication.)
7. Use multiple modalities and information channels redundantly.
8. Don't assume every user is the same; be sensitive and adapt to individual, cultural, social, and contextual differences.
9. Be aware what your user knows, especially what you just conveyed (i.e., don't repeat yourself).
10. Be cute only to the extent that it furthers your conversational goals.

Grice surely had in mind sentient human beings of the same culture, while Miller seems to have had in mind a computer decision aid, where "presence" or "reality" have more to do with the intelligence of the computer than with its appearance. My purpose here is to make a distinction between appearance of reality (immersion) and intelligence. A virtual brick can appear real if we can view it from any desired angle and distance, handle it, etc, but the brick is hardly intelligent. A virtual computer that looks like a computer and responds like a computer will not seem to be other than a computer. But if the computer looks like a computer but responds like a human being, a user will be drawn in to a conversation as if there were a human on the other side. In other words there can be "presence" and "immersion" based on the apparent intelligence of the other entity, rather than physical appearance

4. Computer Assistance in Designing Virtual Environments

Designing a virtual environment is a problem in hierarchical control and in design. In the abstract, control and design are the same: one has goals that need to be specified, and constraints that need to be fulfilled. Specifying the goal really amounts to specifying the objective function—the tradeoff between salient variables. Specifying the constraints amounts to thinking hard about the limitations on the physics involved as well as the physiological and psychological properties of the users. Unfortunately, rarely can these equations be specified mathematically—as is required to derive a unique optimal solution. This still does not relieve the designer of the need to struggle with explication of the tradeoffs between objectives and the constraints—typically many and diverse, as has been suggested in the foregoing paragraphs.

Often the constraints on a virtual environment can be stated as single variable limits: a minimum number of polygons, a minimum refresh rate, a restricted range of apparent distances and viewing angles and rate of change of viewpoint within this space, a few key physical interaction phenomena between elements of the virtual scene, a restricted stimulus (e.g., vision, tactile

sensing but no sound) which is expected to provide sufficient sense of reality, etc. Sometimes there will also be known relations between two or a few variables, e.g., a tradeoff between number of polygons and refresh rate. These constraints can then be put into a computer, and the constraining relations will bound a space of hopefully not too many variables. However, as with any real design problem, the number of variables will be variables but usually significantly greater than three – so that visualization of the bounded (feasible) solution space is not possible by an ordinary mortal (whereas the computer has no problem).

Such a hyperspace will be characterized by many bounding hyperplanes, some perpendicular to the relevant variable, others at angles when there is a linear tradeoff relation between variables (or a surface where the relation is nonlinear). There will be many “corners” of this feasible solution space where planes and surfaces intersect. If all constraints are constants (planes perpendicular to axes) or linear relations (planes at angles) this is essentially the solution space of linear programming. Figure 4 shows a very simple example (in two variables) of how one or another of the corners tends to intersect the best objective function (tradeoff) curve of indifference between number of polygons and refresh rate. The computer can easily find this optimal solution in a much more complex hyperdimensional space—assuming there are a large number of alternatives available to select from in the feasible solution space and the tradeoff indifference lines are known.

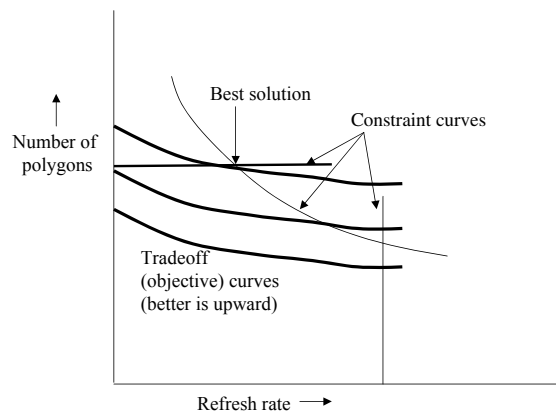


Figure 4. Best solution at intersection of constraint lines

Let us assume, however, there is only a small discrete number of solutions alternatives available (others perhaps being rejected by constraints) and one must select among these (again given that the tradeoff indifference lines are available). The computer can again be of assistance. Figure 5 shows this situation, in a very simple form. The four shapes represent four alternative designs. The triangle can be rejected immediately because it is dominated by the square (the latter being better in number of polygons and the equal in refresh rate). The cross, square and circle form a so-called Pareto optimal (non-dominated) set. However the square is the best because it is at the highest utility (the highest tradeoff indifference curve).

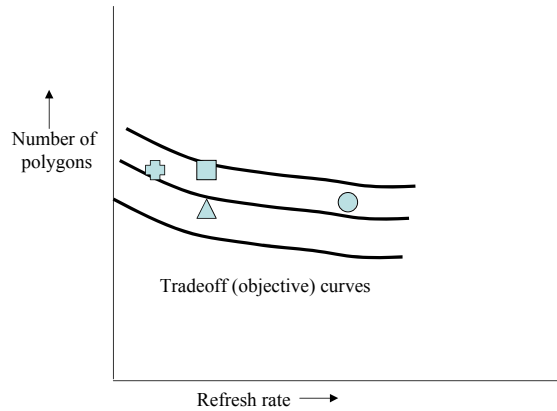


Figure 5. Best solution as the Pareto optimal alternative with the greatest utility.

Finally, if the tradeoff indifference curves are not known, the computer can at least throw out the alternatives that are dominated and leave the designer to choose among the alternatives on the Pareto frontier. In a complex design problem where the solution space cannot be visualized this alone can be of great help. Charny and Sheridan (1989) demonstrated these techniques for interacting with the designer in a five-dimensional solution space.

5. Conclusion

If the many expected constraints are not adhered to, a virtual environment (as with many other forms of human interaction) does not appear real or even intelligent. Explication of constraints as well as objective function (in the form of tradeoff indifference curves for salient variables) allows the computer to assist the designer in selecting a best design.

References

- Boring, E.G. (1950). A history of experimental psychology. NY: Appleton-Century-Crofts.
- Charny, L. and Sheridan, T.B. (1989) Adaptive goal setting in tasks with multiple criteria. In Proc. IEEE 1989 International Conference on Cybernetics and Society.
- Ellis, S.R., Kaiser, K.K., and Grunwald, A.J. (eds.) (1991). *Pictorial communication in virtual and real environments 2nd Ed.*, London: Taylor and Francis.
- Grice, H.P. (1975). Logic and conversation, In P. Cole and J. Morgan, *Syntax and Semantics: Speech Acts*, vol. 3. NY: Academic Press.
- Miller, C.A.(2000). Rules of etiquette, or how a mannerly AUI should comport itself to gain social acceptance and be perceived as gracious and well behaved in polite society. Unpublished working notes of AAAI Spring Symposium on Adaptive User Interfaces (March 20-22, 2000).
- Roseborough, J. (1988). Aiding human operators with state estimates. PhD Thesis, Cambridge, MA, MIT.
- Sheridan, T.B. (1992). Musings on telepresence and virtual presence. *Presence*, vol. 1, pp. 120-125.
- Sheridan, T.B. (1992). Telerobotics, automation and human supervisory control. Cambridge, MA: MIT Press.
- Sheridan, T.B. (2000). Descartes, Heidegger, Gibson and God: Toward an eclectic ontology of presence. *Presence*, vol. 8, no. 5, pp. 549-557.
- Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence*, vol. 8, pp. 460-565.
- Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in virtual environments. *Presence*, vol. 3, pp. 130-144.
- Tufte, E.R (1997). Visual explanations. Cheshire, CT: Graphics Press.
- Tufte, E.R. (1983). The visual display of quantitative information. Cheshire, CT: Graphics Press
- Webster's Third new International Dictionary (1965). Springfield, MA: G. & C. Merriam Co.

