Realistic Virtual Reality and Perception

John Dilworth

[*Philosophical Psychology* 23(2010), pp. 23-42]

Realistic uses of Virtual Reality (VR) technology closely integrate user training on virtual objects with VR-assisted user interactions with real objects. This paper shows how the Interactive Theory of Perception (ITP) may be extended to cover such cases. Virtual objects are explained as concrete models (CMs) that have an inner generation mechanism, and the ITP is used to explain how VR users can both perceive such local CMs, and perceptually represent remote real objects. Also, concepts of modeling and representation are distinguished. The paper concludes with suggestions as to how the ITP methodology developed here could be extended to iconic external representations and models generally.

Keywords: Concrete Modeling; Perception; Perceptual Representation; Representation; Virtual Reality

John Dilworth is a Professor in the Department of Philosophy, Western Michigan University. Correspondence to: John Dilworth, Department of Philosophy, 3004 Moore Hall, Western Michigan University, Kalamazoo, MI 49008, USA. Email: john.dilworth@wmich.edu

This paper involves an attempt to extend a particular causal theory of perception--the *interactive theory of perception* or ITP (Dilworth 2004, 2005a-c, 2006, 2008, 2009; see section 3 for motivation and a summary)--to cover some more specialized perceptual phenomena associated with *virtual reality* (VR) technologies. Generally speaking, causal theories of perception, such as the Dretske/Fodor nomic covariance-based approaches (Dretske, 1981; Fodor, 1990), attempt to explain perceptual activities in purely causal terms. Such comprehensive causal theories may also attempt to naturalize all semantic phenomena on such a perceptual base (e.g., Dretske/Fodor or the ITP (Dilworth, 2008, 2009).

But a standard hurdle for such comprehensive causal theories is that of explaining concepts of *representation* and *representational content* within their causal framework--i.e., the project of attempting to *naturalize content.* Three strategies suggest themselves: reduction, elimination or abstraction. The Dretske/Fodor approach is one well-known reductive account. Churchland (1979) and Stich (1983) attempt to eliminate propositional attitudes and propositional content from their accounts of cognition. As for the third category of abstract approaches, Dilworth (2008, 2009; see also Loar, 1981) has recently argued that all semantic concepts, such those of representation and content, are purely abstract. On this abstract approach, such concepts function as part of the broadly logical or mathematical metalevel of all cognitive sciences. In this framework, the project of naturalizing content is a project of finding explanatorily useful scientific correlations between causally based cognitive structures and abstract semantic structures such as propositions. For example, according to the ITP, to *perceptually represent* an object x as being red is to be disposed to classify it as red, and that cognitive state correlates with, or is indexed by, the abstract proposition 'x is red'.

However, in order to be fully plausible, any theory of representation must also distinguish its analysis of the concept from that of related concepts that are not fully representational. As a recent case in point, Ramsey (2007) has shown that isomorphism-based accounts of representation, such as that of Cummins (1996), are fundamentally inadequate (see also Millikan, 2004). So one significant challenge for supporters of the ITP is to articulate a valid distinction between its classificatory approach to genuine representation on the one hand, from cognitive or external putative structures C that are merely isomorphic with, rather than representing, worldly structures W on the other hand. Hence a significant part of this paper will be taken up with distinguishing VR-related cognitive and external *models*--that are isomorphic to the world--from genuine representations of the world. Fortunately, it will turn out that the operation of VR equipment can be adequately explained in terms of *concrete models*, rather than abstract models or representational content--a perhaps unexpected result that potentially could be applied to other standard problems concerning representation as well.

Previously the ITP has been defended in various ways. Dilworth introduced it (2004, 2005a-c--initially entitled the 'reflexive' theory of perception) as a general-purpose causally based perceptual theory that can apply to organisms at any level of evolution as well as to humans, and which is also justifiable on evolutionary grounds. Consequently, the basic cases dealt with by the ITP are cases of normal unaided perception for any species. However, subsequently, the ITP was applied in two broader ways. First, to cover cases of purely mechanistic, non-biological perception (Dilworth, 2006). And second, to cover cases involving *prosthetic extensions* of human perception, such as cases in which blind persons use a sensory transducer to 'see' via a pattern of touch-based stimulations on their back or tongue, or cases in which scientists can see the surface of a distant planet by looking at a photograph of it. (Dilworth, 2005b) This current article provides some independent support for the ITP treatment of the general concept of prosthetically extended human perception, by extending it to include cases involving VR technologies. Nevertheless, though the discussion concentrates on this extension to the

ITP, the methods used to raise and resolve issues are potentially relevant to the explanatory powers of causal theories of perception in general.

As for virtual Reality (VR) technology, it is a group of loosely related technologies that enable users of the equipment to interact with various objects, whether virtual objects or real objects. In real-object cases, VR equipment would enable a user to interact remotely with some real object. For example, a surgeon could use VR equipment to operate on the heart of a patient located some distance away (McCloy and Stone, 2001; da Vinci Surgery, 2007). In virtual-object cases, the object interacted with via the VR equipment, such as a virtual heart operated on by a surgeon in training (Suzuki et al, 2005; Weaver and Steffes, 2007), would be causally unrelated to any remote real object.

In cases of uses of VR equipment to interact with real objects, there are two basic VR components or factors involved. One basic component is a *remote perceptual* facility, which enables a VR user U to obtain perceptual data about a remote real object X in ways other than by merely using her unassisted sense organs. For example, a remote video camera can obtain visual data about a real object X, which can be accessed by user U via a screen, or a binocular headset etc. Another basic VR factor is a *remote presence* facility, which enables actual objects to be remotely manipulated by the user of the equipment. The user has one or more control devices which, when directly manipulated by a user from his own position, would cause equivalent movements to take place in a remote piece of robotic equipment, such as a camera moving around an object X, or a robotic arm that interacts with the object (Sherman and Craig, 2003). In these ways the user can interact with the remote real object X, and also obtain perceptual feedback via the remote perceptual facility.

Of particular interest are cases in which VR equipment is used by surgeons both in a virtual training mode and during real operations. Appropriate concepts of realistic VR technology, modeling, representation and perception must be interwoven in any adequate explanation of such mixed VR cases. This paper will provide such an explanation. One basic concept to be introduced is that of *concrete modeling*--which concept seems to be breaking new ground.[i]

As for the concept of *realistic* VR technology, examples such as the VR surgery cases cited above are realistic in two closely related ways that are relevant to present concerns. First, the purpose of the relevant virtual object, such as a virtual heart, is to simulate or mimic, as closely as is needed for the current application, the properties of a real object of the same kind, such as a real heart. And second, when using the VR equipment to interact with a real object, the information provided to the user should be realistic enough to adequately carry out the intended task, such as successfully operating on a real heart. Hence our concern with realistic VR technology is strictly limited to a functionally adequate kind of realism or verisimilitude, in VR interactions either with virtual or real objects. In contrast, what is not needed is a fully immersive or fully realistic VR experience for the user, of a kind so that e.g. the user could not distinguish her VR-assisted seeing of a virtual heart from unaided seeing of a real heart. Success-based functional adequacy in both virtual and real cases is all that is needed for present purposes.

By limiting investigation to such broadly utilitarian kinds of realistic VR cases, all of the other possible uses of VR technology, including entertainment-based uses in videogames, and other broadly fictional kinds of content, can be bypassed during this initial investigation. This is all to the good, because there are a host of thorny issues--concerning representation of non-existent fictional objects, whether we can really perceive, or just pretend to perceive, fictional objects, and so on--that are best avoided if possible in an initial treatment of our realistic cases. As a consequence, a rigorous concentration on the relevant realistic cases enables us to avoid having to claim that VR equipment *represents virtual objects*, or that we are *perceptually related to* any such represented virtual objects in using such equipment. Indeed, at least initially the concept of representation--whether or not virtual objects are involved--can be avoided altogether in the concrete modeling approach to be proposed.

A *concrete modeling* approach to realistic uses of VR technology treats the technology, in such uses (assumed from now on) as being nothing more than a technologically advanced kind of *real-time generator of a concrete interactive model*. A *concrete model* (CM) is a concrete object--as opposed to a represented model or a virtual model--that can be used to model or simulate the properties of some real object (section 2 distinguishes modeling from representation). For example, model cars can range from toy cars for children to full-size replicas of famous racing cars that have almost all of the functionality of the real vehicle. Such familiar cases are usually producing using industrial equipment that, on completion of the model, can be entirely separated from the resultant model. Consequently, the child who plays with a model car is not normally causally related to the manufacturing equipment which produced that model car.

However, in the case of VR equipment, my claim is that the equipment functions both as a device to manufacture, produce or generate a concrete model--i.e., a concrete realization of an internally generated model--*and* as a device to immediately display the model it has thus produced. VR equipment both generates a CM, and displays it for the user. Indeed, these functions are so closely integrated in VR cases, that the generation process can be considered as providing the *inner mechanism* for the generated model--i.e., providing the means by which the CM maintains its identity and becomes perceptually available for interactive uses.

For example, a VR system might have as its CM display unit a viewing screen, whose concrete phosphor or light-emitting diode (LED) surface structure displays one particular concrete aspect of the relevant CM. If the CM is a model car, typically a user then could--via cursor or mouse movements, touching the screen display, and so on--examine different aspects of the same model car, open its doors, and so on. Also, what makes all of these concrete screen aspects aspects of the same CM is that the user interaction causally interacts with the *same inner generation mechanism* of the current CM, which provides the current concrete viewable aspect of this CM.

The most salient difference between CMs having such an inner generation mechanism, and those simple CMs that lack it, is as follows. A simple CM is a model that simultaneously models all of the concrete features of that kind of model. By contrast, a complex CM having an inner generation mechanism permits a user to interact with the mechanism so as to produce, in real time, *any desired compatible subset* of the concretely modeled features of that kind of model, without also producing at the same time all other possible subsets of the relevant concrete features. Consequently, complex CMs, as in VR equipment cases, permit a distinction between the complex model itself, and the currently displayed set of compatible concrete features as chosen by the user of the VR equipment.So, to sum up, a particular piece of VR equipment, running a VR software package, is a concrete integrated generation and display unit that displays various physical aspects of the CM it generates, as a result of user interactions with the inner generation mechanism of the CM. Also, since any computer system running a software package is a purely physical computational system, with a physical display unit that is perceived by the user, it is indeed possible to explain all of the interactive facilities of realistic VR cases in terms of user interactions with a concrete model that is generated by the VR system. Hence--at least in realistic VR cases--what are generally described as 'virtual objects' in standard accounts of VR are here explained as CMs.

1. Concrete Modeling is Independent of the Degree of VR Immersion

One potential roadblock that should be removed before proceeding is as follows. Since the preceding discussion of the concrete modeling (CM) approach to realistic VR cases used a concrete screen surface as an example of the physical realization of a side of the model, it might be thought that the whole CM approach would have to be abandoned in full immersive VR cases, in which there is not any screen on which the model is displayed. However, this is not the case, for the following reasons.

Recall that on the CM account, the VR equipment generates the relevant concrete model in real time. Every aspect of the VR equipment is physical, including the physical memory chips and CPU that is processing the current concrete instantiation of the relevant VR program. Equally, every aspect of the mechanism that displays the model for the user is also purely physical. To be sure, any VR-generated models--whether immersive or not--may have some physical properties that are not fully perceived by the user, just as with more traditional media such as film. A projected screen image of a traditional film actually displays a sequence of 24 or more still photographs every second, whereas viewers of it instead perceive a single moving picture. Similarly, a video display screen as used in VR equipment is normally scanned using a moving electron beam at similarly high rates of image production, even though users are unaware of how the physical image is produced. But in both traditional and VR cases, a fully physical model is produced that is displayed to users.

As for more immersive forms of VR, a 3D holographic image might be formed in space that is more realistic than a screened image (van den Bosch et al, 2005). But such a 3D model is just a more sophisticated physical model, whose concrete display properties have somewhat more complex relations to the perceptual systems of users than those holding in less immersive VR cases. Similarly, a fully immersive system, in which the user sees a fully realistic display in his VR headset, is just another way to display a realistic concrete model to a user. Indeed, distinctions similar to those appealed to here have been available since the early days of microscopy. A microscopic image can either be displayed on a screen using projection equipment, or seen directly by viewing it through the microscope. But in both cases the same concrete specimen is seen. Similarly, different methods of displaying a VR image all equally give perceptual access to a concrete model that is generated by the VR equipment.

2. Distinguishing Concrete Models From Representations

A crucial issue for the current concrete modeling (CM) approach to VR is that of how it relates to *representational* concepts. In particular, if a model M is only so in virtue of its representing some independent kind of object X, such as a real car, then M is just a representation R of car X, having a representational content X'. But in that case, the properties of model M that are relevant to its being a model of a car--such as its having wheels, and doors that can be opened--would not literally be properties of model M, since on the representational view M does not literally have such properties in its own right. Instead, M would only *represent* some independent car X as having wheels and openable doors. Or in other words, M would not itself literally have wheels, but it would only have wheel-related representational content, in virtue of its representing some real car X--or possibly some virtual car, if the real car does not yet exist, as in prototype design. Consequently, the full benefits of a CM literalist view of perception of VR objects can only be obtained if some alternative, at least partly non-representational view of the parts and properties of models can be developed--such as of the *functional equivalence* kind developed below.

The following non-representational account of literal perception of concrete models has a broadly recursive structure. The base case is that of a whole model. It cannot be denied that concrete, whole model cars etc. exist that can literally be perceived, whether by perceiving a metal model car or a concrete car configuration on a VR screen. The recursive cases are derived from the base case by invoking the parthood relation. Model cars have parts, just as real cars have parts. But each of the parts of a model car is a model in its own right--whether it is a model door, a model wheel, and so on. Consequently, all of the relevant parts of a model can literally be perceived as well, because they too are models in their own right, just as is the whole model of which they are parts.

Next, consider various putative functional properties of a model. It is arguable that the model wheels of a model car can literally be made to turn, or its model doors can literally be opened, etc., for the following reasons. The strategy now is to explain the relevant functional properties in terms of their close association with relevant concrete events, of kinds involved in a user's concrete interactions with the model. Arguably these events are event-based models in their own right--namely, models of corresponding events of functional interactions with real cars. So on this account, a VR user can literally open a door of a concrete model car, because the concrete event of his interaction with the relevant concrete parts of the model car--such as the door and the door-handle--is an event-based model in its own right of some functionally similar event of opening the door of a real car.

Or, more succinctly put, the user's concrete interactions with the model car are themselves models of putative, functionally similar interactions with a real car. Consequently, both the user's VR interactions with the parts of the car, and the events of the concrete reactions of the parts of car to these interventions--which model similar reactions in a real car--can literally be perceived. Hence no representation is involved in a user's claims to perceive such objects and to investigate their real functional properties. Such claims can be straightforwardly and literally true, on the present account.

As to the nature of the relevant kind of *functional equivalence* between a CM X' and a real X, the relevant kind of functionality is of a purely instrumental, practical kind that facilitates interactions with a real X. Some feature F' of a model X' is functionally equivalent to a corresponding feature F of a real object X just in case the inclusion of feature F' in CM X' enables a user trained via interactions with X' to successfully interact with the corresponding feature F of a real object X.

On this account of the functional status of CMs, such models have a status that is broadly similar to that of other social or cultural artifacts, such as chairs--or real cars, the properties of both of which we can literally perceive. What makes such artifacts social is that they are at least partly characterized in terms of their functional utility in furthering our practical purposes (Searle, 1995; Brey, 2003).

VR-based CMs are equally legitimate social artifacts, insofar as they can be used by us in ways that are broadly *functionally equivalent* to how we would use the corresponding real items, as explained above. It is this functional equivalence, rather than representation, which explains the literally perceptual, objective social status of concrete models. Consequently, specifically representational concepts are not needed in a basic account of the social functionality of concrete models--which is not to deny, of course, that a model might be *used* to represent something.

Here is one further point in support of this non-representational functional equivalence thesis for CMs. Various writers on representation have tried to explain representation itself in terms of some kind of first or second-order isomorphism or equivalence of structure between a representation and what it represents (Palmer, 1978; Cummins, 1996). However, as critics have pointed out (Millikan, 2004; Ramsey, 2007), concepts of structural isomorphism or equivalence--including the present concept of functional equivalence--are much too weak to explain genuine representational relations. Similar points have been made in the philosophy of science literature on contrasts between models and representation, e.g. Suarez (2003) or Contessa (2007). For an item R to represent another item X, R must be *about* X--an asymmetric, referential, specifically semantic relation. But by contrast, equivalence relations, including the functional equivalence of models with real cases, are symmetric relations only, and they involve no semantic relation of aboutness or reference--indeed, they are not semantic relations of any kind. The fact that a model X can function in some ways that are roughly equivalent to the functioning of a real X does explain why we classify it as a model X, rather than a model Y, Z, etc. The functional equivalence also explains why a CM can readily be substituted for the equivalent real object in a VR simulation. But, again, no representation is involved in such cases.

3. A Naturalistic Approach to Perception and Representation

The account, in the preceding sections, of VR-assisted concrete modeling enables us to claim that the relevant concrete models (CMs) can be straightforwardly perceived by VR users, just as any other concrete objects can be perceived. But what exactly is involved in perceiving any concrete object, including CMs? Also, if a CM X' were optionally used to represent some real object X, under what cognitive conditions--if any--would a person qualify as perceiving what is thus represented by that CM X'? In order to answer questions such as these, at least the rough outlines of a relevant theory of perception must be presented or assumed. Consequently, at this stage it is necessary to be more theoretically explicit about the nature of perception, perceptual representation, and also the nature of putative representation by concrete models of corresponding real objects, for those circumstances in which it occurs.

Arguably the kind of perceptual theory that would be most consonant with the current CM approach to VR would be a causally based, broadly naturalistic one--because the CM approach is itself a broadly naturalistic approach that seeks to avoid any initial dependence on concepts such as those of representation of imaginary virtual objects. In particular, a perceptual theory that identifies perception of an object X with perceptually based kinds of *causal interaction* with X is highly desirable, since one of the most salient features of VR-generated CMs is that the VR equipment facilitates user interaction with them. So an initial specification for a desirable perceptual theory is that it should explain perception of an object X by a person P in terms of object X causing person P, via P's sense organs, to become disposed to cause changes in object X itself--a causally *interactive theory of perception* (ITP). Fortunately there is one such theory that has been defended in detail in the literature (Dilworth, 2004, 2005a-c, 2006, 2008. 2009). The ITP has been selected because it seems to be the only available causal theory that has the theoretical resources to adequately explain both the operation of VR equipment during surgical or other kinds of training, and its subsequent use in facilitating actual procedures carried out on live patients. Consequently, as proposed in the introduction, this paper can appropriately be regarded as a demonstration that the ITP has the explanatory resources to adequately account for the relevant VR cases under the current concrete model (CM) interpretation of them.

The basic claim of the ITP is that an organism Z perceives an object X just in case X causes some sense-organ Zi of Z to cause Z to acquire or activate some X-related disposition D, where a disposition is X-related just in case its manifestation would make some causal difference to X itself. For example, if a movement by predator X causes its prey Z, via its sensory system Zi, to become disposed to hide or flee from X, the manifestation of this disposition D would causally influence X by making it less likely that X would be able to eat Z than if Z had not perceived X at all. In such manifested disposition cases, a closed causal loop is set up in which X causes Z, via Z's sensory system Zi, to cause some *reflexive* causal effect on Z itself (the ITP was originally described as the 'reflexive' theory of perception for this reason). Or, from the point of view of Z as a perceiver, Z causally *interacts* with X, in that Z is caused by X, via its sense-organ Zi, to become disposed in turn to causally influence X itself.

A more specific version of the ITP would explain, not simply what it is for an organism Z to perceive an object X, but what it is for Z to perceive some fact about X, such as its perceiving X to have property F. For example, if a bull differentially reacts to a red cape but not to capes of other colors, its disposition to charge the red cape, but not the other capes, gives evidence that the redness of the red cape has caused it to perceive that redness, as evidenced by its having acquired a red-color-related disposition toward X. Or if a surgeon in training reacts to a perceived bulge in a CM kidney by attempting to surgically remove the bulge, similarly this would count as providing evidence that the surgeon had perceived, not just the CM kidney, but also its property of including an undesirable bulgy area.

There are three central theoretical aspects of the ITP as an adequate causal theory of perception that are relevant to our current VR issues. First, according to the ITP, a necessary condition for a causal process to be a genuinely perceptual one is that a *sense-organ* of the perceiving organism, such as an eye or an ear, should play a mediating causal role in the interactive causal structure. In the ITP itself this feature is initially explained in evolutionary terms: perceptual processes came to be widely prevalent in species because mediating sensory processes provide a fitness-promoting early warning system to those species that happen to employ them.

A second central aspect of the ITP is that it can explain, in purely causal terms, how perceptual reference to a particular worldly object of perception X is achieved. A standard problem for causal theories of perception is that perception involves causal chains, in which many objects and events play a causal role, from the worldly object through retinal images to internal brain processes. Why should one item in this causal chain be the object perceived, rather than some other? An adequate theory must explain why we don't perceive our retinal images, rather than a worldly object X, given that our retinal images are just as much part of the relevant causal chain as is object X itself (refs). The ITP identifies the relevant object of perception X in terms of the X-related dispositions acquired during normal perception. According to the ITP, we perceive a worldly object X, rather than our retinal images, because in normal perception we acquire dispositions to causally interact with X, but we do not acquire any dispositions to causally interact with our own retinal images (see Section 4 for further discussion of this issue).

A third central aspect of the ITP is closely related to the just-described feature of the particularity of perceptual reference. This third aspect concerns the issue of how to articulate, in purely causal terms, the difference between correctly versus incorrectly perceiving an object X and its properties--i.e., the issue of correct perceptual representation versus misrepresentation. As is well known, nomic covariance causal theories of the Dretske/Fodor kind (Dretske, 1981; Fodor, 1990) have great difficulties with the distinction because nomic covariance explains only correct representation. In contrast, the ITP is able to proceed as follows. Since the ITP can identify the particular object X being perceived--in terms of X-related dispositions--independently of whether it is perceived correctly or incorrectly, these further specifications can be explained in terms of which specific X-related dispositions are involved in a particular case of perception. For example, a person who perceptually acquires color-related dispositions from a red object X might become disposed either to correctly classify X as red, or to incorrectly classify it as some other color such as green. Consequently, the ITP is able to provide a substantive account of perceptual representation and misrepresentation that is not inherently biassed in favor of correct representation, as are nomic covariance accounts.

In addition, the ITP can explain specifically representational cases of perception in terms of a subset of interactive dispositions, namely conceptually based *classification* dispositions. On this approach, the difference between perceiving a CM simply as a physical object, versus perceiving it as a model of a particular kind--such as a model car--is that the relevant interactive dispositions would involve *car-related classification dispositions* in the latter case, but not in the former. So on this account, to perceive a CM X' as a model car is to be caused by X' to become disposed to classify it in terms of those functional properties that the model car X' shares with real cars, as discussed in the previous section. Also, such an account should be naturalistically acceptable, in that there can be scientific evidence as to whether or not a perceiver possesses such car-related classification dispositions. A perceiver's actual interactions with a concrete model car X' could demonstrate his abilities to turn its wheels, open its doors, etc, and hence provide evidence of the perceiver's possession and exercise of the relevant car-related classification dispositions.

Perception and perceptual representation having been interactively explained as above, an initial explanation can now be provided of what is involved when a concrete model X' is optionally recruited to *represent* some real object X, such as when a VR-based system displaying that CM X' is used to support VR-based interactions with a real object X. Fortunately, the ITP could potentially provide a straightforward explanation of such cases of external representation in purely perceptual terms. On such an account, a VR-generated CM X' is being used to represent a real object X by a perceiver P just in case the CM X' causes person P to acquire *interactive classificatory dispositions* (ICDs) toward the real object X. (The dispositions are interactive in the sense provided by the ITP itself). Also, arguably in such cases the relevant, VR-mediated causal links between real object X and the CM X' are such that X itself counts as causing P to acquire the relevant ICDs toward X, so that this would also count as a case of perception of X by person P. Hence, to simplify the explanation, a VR-generated CM X' *represents* a real object X for perceiver P just in case X causes the CM X', via VR-mediated causal links, to cause perceiver P to perceive X.

Also, in some such cases, person P may also acquire ICDs toward the CM X' itself, and consequently perceive X' as well as perceiving X--in which case there would be VR-mediated cases of dual perception. The following sections will investigate these matters in greater detail.

4. Two Preliminary Cases: Optical versus Electronic Binoculars

Before discussing VR cases any further, it will be useful to initially apply the interactive theory of perception (ITP), as outlined in the previous section, to two instructive preliminary cases--one of which is unrelated to VR, and the second of which is an intermediate case.

To begin, it is usually assumed that normal, non-VR-assisted perception of a concrete object X is *direct*, in the sense that it involves a direct causal relation between object X and a perceiver P's perceptual system. Insofar as any representation is involved in such cases, it is assumed to be limited to an internal, cognitively based perceptual representation R, which is such that R directly represents object X in virtue of the obtaining of normal perceptual conditions. (Consequently, the relevant sense of 'direct' is independent of that which is operative in traditional debates about whether perception is mediated by internal representations). Or, in terms of an ITP, under such conditions the perceiver P counts as perceptually representing object X just in case X causes P to acquire or activate appropriate *interactive classificatory dispositions* (ICDs) toward X.

Next, consider the category of *perceptual prostheses,* where a perceptual prosthesis is an assistive device for perceivers, whose function is to facilitate direct perceptual access to objects. For example, reading glasses are visual prostheses that enable a person to directly see objects more clearly, while telescopes or binoculars are more complex visual prostheses that enable a person to directly see a distant object as if it were much closer to the perceiver. Arguably such prosthetic devices do not introduce any indirectness of perception, or additional kinds of representation, because their causal function is simply that of facilitating normal direct causal interactions between an object X and a perceiver P's perceptual representation R of X. Consequently, an ITP would regard the function of such prostheses as involving no more than a causally based kind of enhancement or strengthening of perceiver P's relevant ICDs toward X.

However, things become more problematic in a case such as the following. Though, as mentioned, a traditional optical pair of binoculars is a direct prosthesis, electronic versions are now evolving (Isbell and Estrera, 2003) which interpose an electronic amplification mechanism within the optical path. Such an electronic binocular device converts incoming photons of light into electrical impulses that are amplified, and then re-converted into light by their striking dual phosphor-covered screens. Consequently, it might initially seem as if what the perceiver P directly sees when using such an instrument is a binocular fused view of the images X' on the dual phosphor screens, rather than his directly seeing the real distant object X on which the binoculars are focused. So, in terms of everyday intuitions about direct perception, this might seem to be either a case of indirectly mediated perception of the real object X, via direct perception of the dual phosphor screens X', or alternatively a simpler case in which perceiver P sees only the phosphor screens X', without seeing X itself at all.

Nevertheless, from the viewpoint of an ITP, what a person P counts as perceiving or perceptually representing depends on two factors: which object or objects cause the relevant perceptual state in P, and which object or objects P acquires ICDs toward. Also, a naturalistic ITP will look for empirical evidence to support any hypotheses regarding these matters. As for object causation, both the real object X and the dual phosphor screens X' cause perceiver P to acquire relevant dispositions, in that both X and X' are links in a chain of causation leading from X to P. So the crucial element in determining which object or objects P perceives is that of which object or object P acquires ICDs toward.

That question should be settled by gathering appropriate evidence. If there is evidence that P acquires ICDs exclusively toward X, then P would perceive only X, and not the dual phosphor screens X'. If instead there is evidence that P acquires ICDs exclusively toward the dual screens X', then P would perceive only those screens X'. Potentially a mixed case is possible also, with perceiver P acquiring ICDs toward both X and X'.

For example, if perceiver P uses the electronic binoculars exactly as he would use ordinary, purely optical binoculars--namely, simply as a means of acquiring information about the distant object X--then in so doing, X causes P to acquire X-related ICDs, so that P perceives X. But unless P also has some special interest in the functioning of the electronic binoculars--such as in observing how the eyepieces could enable him to see a magnified image of the dual phosphor screens X'--he would not also acquire any ICDs toward the screens X' themselves, and consequently he would not see the screens X', even though they are a proximate cause of his seeing of the distant object X.

As evidence for the ITP-based claim that the object perceiver P counts as seeing depends on which object P acquires ICDs toward, consider an unassisted perception case. When P sees an object X without any prosthetic assistance, there is a chain of causation leading from object X, via P's eyeballs and their inner retinas, to P's processing of the visual information. A prominent factor in this causal chain is the dual images X' on P's retinas. However, of course what P sees is the object X, not the images X' on his own retinas. An interactive theory of perception (ITP) can explain this obvious fact by invoking ICDs. P sees X, not X', because P perceptually acquires ICDs toward object X rather than towards his own retinas X'. And a typical piece of evidence for this is that, if P is hungry and food is put within visual range of his eyes, he will attempt to eat the food itself, rather than attempting to eat the relevant areas of his own retinas. Since it is obvious in such a case that P does not see his own retinas, in that he has no retina-related dispositions, it should be equally unproblematic to accept that in the electronic binocular case, a typical user would see the distant object X on which the binoculars are focused, and not see the dual phosphor screens that are the proximate cause of his seeing X, because the typical user would acquire dispositions toward distant object X but not toward the proximate screens X'.

The electronic binocular case is intermediate between unassisted seeing and VR-assisted seeing because it involves a concrete image on a viewing screen or screens. In this respect it is analogous to the concrete model produced by VR equipment, which could also be implemented using a concrete image on a viewing screen or screens. Also, the electronic binocular case can serve to provide an initial proof-of-concept that VR-assisted interactions with a real object X could count as cases in which user P does perceive the real object X that he is interacting with via the causal mediation of the VR equipment--as initially postulated in the previous section.

5. Interactive Differences: VR Cases versus Electronic Binocular Cases

A significant difference of VR-assisted cases of perception of real objects X from electronic binocular cases of perception of real objects is that VR equipment supports not just viewing, but also interaction with the relevant real object X. For example, surgical VR equipment can support both remote perception of a real kidney by a surgeon, and surgical interaction with it via robotically controlled instruments that are remotely manipulated by the surgeon (Guy's 2007; da Vinci Surgery, 2007). From the perspective of an interactive theory of perception (ITP), the remote interaction provides further evidence that the surgeon does genuinely perceive the real kidney, in virtue of his acquiring and activating interactive classificatory dispositions (ICDs) toward kidney X. Also, the evidence for this dispositional activation is provided by the behavioral results, namely the surgeon's skilled interactions with the real kidney X.

Moreover, there is another significant difference from the electronic binocular case when VR technology is used in a learning mode. In such cases, the learner, such as a surgeon in training, normally interacts exclusively with a virtual organ--i.e., a VR-generated concrete model (CM) organ--rather than a real organ. By contrast, no comparable learning mode is possible with an electronic binocular, because it lacks any facilities for a user to directly interact with its phosphor screens X' in order to change the concrete image that is displayed on them. With the binocular, the only way to change the displayed image X' is to point the binocular at a different real object X. But interactive VR learning equipment does permit a learning surgeon to directly interact with a concrete model organ X', in ways that are functionally equivalent--as explained in section 2--to ways in which he could, after training, interact with a real kidney X using similar VR equipment.

However, this ability of VR-assisted trainees to interact initially with a virtual CM X', prior to their interacting with a real item X, does have the following significant perceptual consequence for VR-assisted interactions with such real items. A surgeon S who has learned how to operate on a kidney by operating on a CM kidney X' has thereby acquired a complex series of interactive classificatory dispositions (ICDs) toward the model kidney X', in virtue of which surgeon S counts as perceiving the relevant CM X'. Also, well-designed VR equipment should have as a design goal the achievement of a realistic functional equivalence between a virtual learning mode and a real application mode, as discussed in the introduction. Consequently, a surgeon using the VR equipment should not notice any distracting display differences between his training with a virtual model kidney, and his subsequent operations upon a real kidney. He should be able to operate on the real kidney X just as if it were another CM kidney X' that he operated on during training. But a consequence of such a desirable VR design is that arguably the surgeon would initially perceive *both* the CM kidney X' displayed by his VR equipment, *and* the real kidney X on which he is operating.

This dual kind of perception should hold for the following reasons. The surgeon has been trained to acquire and activate ICDs toward the CM kidney X' on which he trains. Since the design goal is to make the real operation functionally equivalent to the virtual training, the surgeon's relevant ICDs toward the CM kidney X' should continue to hold during the subsequent real operation use, hence ensuring the surgeon's perception of X'. Also, of course the main purpose of the VR training is to ensure that, during an actual kidney operation on a real kidney X, the surgeon would indeed perceive the real kidney X on which he is operating, in virtue of his acquiring or activating ICDs toward it as discussed above. Consequently, it would seem that during VR-assisted surgery, the VR-trained surgeon would normally perceive both the CM kidney X' and the real kidney X simultaneously, in virtue of the functioning of the inner mechanisms of that CM.

This possibility of VR-assisted dual perception may be an unexpected result, but it need not be a problematic one. One way in which it might be problematic is if such dual perception cases were cognitively inefficient in some way, such as in cases of simultaneously driving a car and operating a cellphone. But arguably, as already pointed out, a goal of good VR design would be the elimination of any functional discrepancies between the functioning of the equipment in virtual and in real modes. Consequently, there need not be any additional cognitive costs to the user of the VR equipment in such a dual kind of perceptual functioning.

Also, it is possible that, if the VR equipment were used often enough in real operation cases, with no intermittent retraining on purely virtual cases, eventually the surgeon's ICDs toward the CM kidney X' would fade or become extinguished, so that eventually the surgeon would qualify as perceiving only the real kidney X on which he is operating. In this respect the situation might be analogous to that of a novice user of electronic binoculars, who initially--because of their novel features--might acquire ICDs toward the monochrome greenish hues of the dual phosphor screens X', rather than toward the distant real object X.

6. Concrete Models, Iconic Representations and the ITP

This section shows how the VR-related introduction of concrete models (CMs) in the earlier sections, along with the ITP-based account of how such CMs could be used to represent actual cases of the relevant models, potentially could be generalized. The generalization would cover any cases in which concrete external models, including iconic representations such as pictures, may be used to represent real worldly objects. To begin with CMs, recall that the concept of a simple CM--of a concrete model that lacks an inner generation mechanism--is a concept familiar from model cars and airplanes, as well as from more sophisticated, scientifically useful CMs such as a physical scale model of a tidal basin. Now usually computer or VR-based models are assumed to be, on the other hand, abstract or virtual models rather than CMs. However, the earlier sections showed, with the aid of the auxiliary concept of an *inner generation mechanism* for a CM, that computerized models of a VR-related kind could also be explained in CM terms rather than in more usual abstract or virtual ways. A significant advantage of this inner generation CM approach is that the relevant functional properties of a CM become literal and directly perceivable functional properties of the relevant CM.

In addition, sections 2 and 3 showed that such inner-generated CMs could also function as *representations*, such as a case in which a VR-generated CM kidney is used to represent, for a trained surgeon, an actual patient's kidney that he is currently operating upon with with the aid of the VR technology. In such a case the relevant CM kidney also represents the relevant actual kidney, to the extent that the surgeon he is able to acquire interactive classificatory dispositions (ICDs) toward that actual kidney itself, in virtue of the VR-supplied causal linkages between the CM kidney and the actual kidney.

One further factor is involved in these VR cases--as discussed in section 5--which normally would *not* generalize to other kinds of representation. Any adequate VR system for a real-time operation on an actual kidney K would have to ensure that the results of any manipulations of K also caused appropriate alterations in the local features of the CM kidney as perceived by the surgeon, so that he could receive appropriate feedback on the results of his manipulations of the actual kidney K. But under these conditions, the relevant causal links--in which actual kidney K causes the surgeon to acquire K-related dispositions toward that actual kidney K itself--would qualify this as a genuine case of *perception* of actual kidney K by the surgeon, by the standards of the ITP, in addition to his concurrent perception of the local CM kidney. But in general, absent the special-purpose, customized causal linkages provided by VR equipment, perception of a CM--whether or not it is used as a representation--would not normally qualify as also involving perception of any actual object that the CM might be used to represent.

The relevant generalization will now be stated more explicitly. The basic idea is that the ITP-based account of VR-based CMs, and of what is involved in representational uses of them, can be generalized to apply to *any* cases of CMs (with or without inner generating mechanisms) that can also be used as representations. Arguably the relevant class of CMs is made up of those physical objects that are realizations of some, but not all, functional properties of the corresponding real objects. On this view, a print of a photograph of the Hoover Dam is a CM that models some but not all of the functional properties of the actual Hoover Dam. For example, a print of such a photograph can be seen to have some of the same visual appearance properties of the actual Hoover Dam, as seen from the same position as that from which the photograph was taken. Or, stated in more objective scientific terms, the overall structure of the color values of elements in the photograph is isomorphic to the structure of the corresponding color values in the actual dam itself. In addition, the spatial configurations and relations between visual elements in the printed photograph, such as the prominent towers toward the top of the dam, are structurally similar to the spatial configurations and relations among the actual towers on the dam itself, and so on.

Nevertheless, though the relevant printed photograph is a CM of the Hoover Dam in virtue of its possession of the relevant proper subset of functional properties of the actual Hoover Dam, we must distinguish its property of being a CM of the Hoover Dam from its potential capacity to be used to *represent* the Hoover Dam. As argued in section 2, on the ITP-based account of these concepts, CMs are not *ipso facto* representations, because structural isomorphism in functional properties is a non-semantic equivalence relation that involves no semantic, asymmetric reference or aboutness relation between the CM and the actual dam.

Instead, on the current ITP-based account, the semantic relation of representation for a CM is to be explained in terms of some normal or standard way in which the relevant CM may be *used* by people. In the case of photographs, the socially standard way to use them is as representations of their actual subjects, rather than as non-semantic CMs in their own right. As with the VR cases previously discussed, the central point is that a CM X' is used to represent an actual object X just in case perceivers of X' would normally acquire interactive classificatory dispositions (ICDs) toward the actual object X itself, rather than toward the perceived CM X'.

For instance, to see the photograph of the Hoover Dam as representing the dam is, on this account, to acquire relevant dispositions toward the actual dam itself--such as, in generic terms, becoming disposed, should one visit the actual dam, to be impressed by its features, its great size, and so on. More specifically, to see the concretely modeled towers as representing the actual towers of the dam is to become disposed to believe that the actual towers of the dam have roughly similar functional features to those that can be seen by looking at the photograph. The next section will further investigate the resulting cognitive situation.

7. External Representations May Not Be Perceived

This section investigates the perceptual status of CMs when used to represent their actual subjects in non-VR cases. To begin, as noted in the previous section, normally someone who uses a CM to represent its actual subject does not thereby perceive its actual subject. For example, someone who sees a photograph of the Hoover Dam as representing that dam does not thereby see the actual dam, because it is not normally the case that in such a situation the actual dam causes the viewer to acquire ICDs toward itself. Instead, typically it is only the photograph--i.e., the CM--that causes the viewer to acquire such actual-object ICDs. Consequently, though the viewer may perceive the CM, in virtue of concurrently acquiring IDCs toward the CM itself as well as toward the actual subject of the photograph, the represented actual subject is not itself perceived.

In addition, a theoretically challenging issue arises in the following kind of situation. Recall that a user of VR equipment may become so accustomed to using it to manipulate some actual object that he may no longer perceive the relevant CM, even though that CM remains as the proximate cause in a chain of causality proceeding from the actual object to the viewer, and back again via the viewer's ICDs toward the actual object. But an analogous situation is possible in the case of non-VR representations as well, even though such cases normally involve no perception of the represented subject.

For example, a film might be made, consisting of aerial shots of the Hoover Dam from a helicopter circling it. Now initially, a viewer of such a film may specifically perceive the concrete screen image itself--i.e.,, the local CM--in virtue of acquiring ICDs toward that CM. However, it is likely that once the viewer becomes absorbed in viewing the film of the dam, he will no longer acquire any ICDs toward the concrete screened image. But by hypothesis, he cannot be perceiving the actual dam either in this situation, since it is not the dam itself that causes his dam-related ICDs, but only the local CM. Consequently, it would seem that strictly speaking, in such a case the viewer perceives neither the concrete screened image--the local CM--nor the dam itself. So apparently the viewer of such a film, strictly speaking, perceives nothing at all in such a case.

This anomalous situation may be theoretically handled as follows. Though it is true that, strictly speaking, a viewer actually perceives nothing in thus viewing the film, nevertheless this does not rule out that it may *seem* to him as if he is perceiving something--and in particular, seem to him as if it is the actual dam itself that he is peceiving, even though that is not the case. Two distinct concepts of seeming may be relevant in such cases (Chisholm, 1957; Jackson, 1977). First, an epistemic concept may be applicable to some viewers. A naive viewer seeing a documentary film for the first time may acquire the false belief that his current, quasi-perceptual, dam-related experiences provide adequate evidence that he is actually perceiving the relevant dam. Or second, in the case of any viewer, it may *phenomenally* seem to him that he is having experiences *as of* the dam--that is, having experiences similar to those he would have if he were actually perceiving the dam.

Fortunately, a purely causal theory of perception, such as the current ITP, need not deny that such phenomenally based, quasi-perceptual experiences may occur--whether in this representational kind of case, in which no actual perception is involved, or in non-representational cases in which genuine perception is involved. Indeed, the anomalous representational cases potentially provide a significant area of common ground between the ITP and more standard philosophical discussions about perception, such as those concerning contrasts between genuine perception on the one hand, and hallucinatory cases on the other hand, in which no actual objects are genuinely perceived (Haddock&MacPherson, 2008).

Another potentially fruitful area of common ground involves broadly imaginative or make-believe approaches to issues of perception of external representations such as pictures. On Walton's well-known account, for instance, perception of a film of aerial views of the Hoover Dam would be explained in terms of a viewer imagining or making-believe that he is actually perceiving the Hoover Dam itself, even though he is not actually perceiving it (Walton, 1990). Clearly such an account has significant parallels to the current ITP-based account. Two potential advantages of the current account over that of Walton are first, that the ITP-based account is closely integrated with a general, causally based theory of perception. And second, potentially the ITP account is more theoretically economical than Walton's imagination-based account, in that no higher cognitive functions of a specifically imaginative or creative kind need to be invoked by the ITP in order to explain the relevant kinds of representational cognition. But these issues, as with the perception versus hallucination parallels, will have to be pursued elsewhere.

8. Summary

This paper has provided some independent support, as well as some further generalization, for the interactive theory of perception (ITP) via a study of perceptual phenomena associated with realistic uses of virtual reality (VR) equipment. Such cases required careful articulation and distinction of adequate concepts of modeling, representation and perception. In particular, the concept of a concrete model (CM), along with that of an inner generation mechanism, was introduced. An advantage of such concrete models is that they can be straightforwardly perceived by users, and they literally possess properties functionally equivalent to those possessed by real Xs.

Another economical methodological factor was the adoption of the purely causal ITP itself. This was particularly appropriate for the analysis of VR-assisted interactions with real objects, because arguably the goal of such VR technologies is simply to achieve a *causal equivalence* between the user's interactions with the displayed CM X', and the corresponding user interactions with a real object X. Such a goal would be achieved as long as relevant user-caused changes in a real object X were such that X, via the VR equipment, subsequently caused *equivalent functional changes* in the user's local CM X', which changes in turn caused user U to activate appropriate further interactive classificatory dispositions (ICDs) toward X itself. But these diagnostic conditions for successful VR interactions with a real object X also served to provide a sufficient ITP-based condition for user U to *perceive* real object X--in that X caused user U, via X', to acquire ICDs toward X itself. Hence, as desired, the VR-based causal equivalence of X' and X was closely integrated with the conditions for user U to qualify as perceptually interacting with real object X. Also, a side benefit of the ITP perceptual analysis was that it permitted issues of perceptual representation to be explained in terms of conceptually specific *classificatory* dispositions--a proper subset of more generic perceptual dispositions.

The rest of the paper spelled out further details and implications of this CM- and ITP-based account of realistic VR uses. Non-interactive display technologies such as electronic binoculars provided a useful intermediate case, which served to clearly distinguish the interactive features of full-fledged VR display technologies from non-VR cases. The paper concluded with a generalization of the results to cover iconic, modellike external representations in general, for which similar distinctions of concrete modeling from representation can be made. Potentially significant connections with standard issues of perception versus hallucination, and of make-believe accounts of perception of representations, were also pointed out.[ii]

References

Brey, P. (2003). The social ontology of virtual environments. *American Journal of Economics and Sociology*, 62, 269–282.

Chisholm, R. (1957). *Perceiving*. Ithaca: Cornell University Press.

Churchland, P.M. (1979). *Scientific realism and the plasticity of mind.* Cambridge: Cambridge University Press.

Contessa, G. (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science*, 74, 48-68.

Cummins, R. (1996). *Representations, targets and attitudes.* Cambridge, MA: MIT Press.

Dilworth, J. (2004). Naturalized perception without information. *The Journal of Mind and Behavior*, 25, 349–368.

Dilworth, J. (2005a). The reflexive theory of perception. *Behavior and Philosophy*, 33, 17–40.

Dilworth, J. (2005b). Perceptual causality problems reflexively resolved. *Acta Analytica,* 20, 11-31.

Dilworth, J. (2005c). A naturalistic, reflexive dispositional approach to perception. *The Southern Journal of Philosophy*, 43, 583–601.

Dilworth, J. (2006). A reflexive dispositional analysis of mechanistic perception. *Minds and Machines*, 16, 479–493.

Dilworth, J. (2008). Semantic naturalization via interactive perceptual causality. *Minds and Machines*, 18, 527-546.

Dilworth, J. (2009). Semantics naturalized: propositional indexing plus interactive perception. *Language and Communication*, 29, 1-25.

Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Fodor, J. (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.

Guy's and St Thomas' NHS Foundation Trust (2007). Trust launches the world's first kidney surgery virtual reality simulator. Report at URL=

< http://childrensallergy.org/news/newsarchive/newsarticles/kidneysurgery.aspx>

Haddock, A. & MacPherson, F. (2008). *Disjunctivism*. Oxford: Oxford University Press.

Isbell, W. and J. Estrera (2003). Wide field of view (WFOV) night vision goggle. *Proceedings of SPIE,* 5079, 208-211.

Jackson, F. (1977). *Perception: a representative theory*. Cambridge: Cambridge University Press.

Loar, B. (1981). *Mind and meaning*. Cambridge: Cambridge University Press.

McCloy, R. and R. Stone (2001). Virtual reality in surgery. *British Medical Journal* 323(7318), 912-915.

Millikan, R. (2004). *Varieties of meaning*. Cambridge, MA: MIT Press.

Palmer, S.(1978). Fundamental aspects of cognitive representation. In: E.Rosch & B.Lloyd (eds.) *Cognition and Categorization*. **Hillsdale, N.J.:** Lawrence Erlbaum, p. 259-302.

Ramsey, W. (2007). *Representation reconsidered*. Cambridge: Cambridge UP.

Schaffer, J. (2003). Overdetermining causes. *Philosophical Studies*, 114, 23-45.

Sherman, W. and A. Craig (2003). *Understanding virtual reality*. New York: Morgan Kaufman Publishers.

Searle, J. (1995). *The construction of social reality*. New York: Free Press.

Suarez, M. (2003). Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science*, 17, 225-244.

Stich, S. (1983). *From folk psychology to cognitive science*. Cambridge, MA: MIT Press.

Suzuki, S. et al (2005). Tele-surgical simulation system for training in the use of da Vinci surgery. *Studies in Health Technology and Informatics*, 111, 543-548.

van den Bosch, A. et al (2005). Dynamic 3D echocardiography in virtual reality. *Cardiovascular Ultrasound*, 3: 37.

da Vinci Surgery (2007). The da Vinci S surgical system. Video overview at URL= < http://davincisurgery.com/surgery/system/index.aspx>

Walton, K. (1990). *Mimesis as make-believe*. Cambridge, MA: Harvard UP.

Weaver, D., MD and C. Steffes, MD (2007). Virtual surgery builds better doctors. *Detroit Medical Center Simulation Lab video*, URL= < http://www.dmc.org/videolibrary/ek_simulatedsurgerylab.html?gclid=CO-f8pyW85ACFQdEFQodOEjp1A>

NOTES

---

[i] For which credit is due to my colleague Dylan Sabo, who suggested this approach to me, and convinced me of its feasibility.

[ii] My thanks to the Editor, two anonymous referees, and to Dylan Sabo, for very helpful comments on earlier versions.