TRUTH TRACKING AND KNOWLEDGE FROM VIRTUAL REALITY

Billy WHEELER

ABSTRACT: Is it possible to gain knowledge about the real world based solely on experiences in virtual reality? According to one influential theory of knowledge, you cannot. Robert Nozick's truth-tracking theory requires that, in addition to a belief being true, it must also be sensitive to the truth. Yet beliefs formed in virtual reality are not sensitive: in the nearest possible world where P is false, you would have continued to believe that P. This is problematic because there is increasing awareness from philosophers and technologists that virtual reality is an important way in which we can arrive at beliefs and knowledge about the world. Here I argue that a suitably modified version of Nozick's sensitivity condition is able to account for knowledge from virtual reality.

KEYWORDS: virtual reality, truth tracking, sensitivity, reliabilism, externalism

1. Introduction

Suppose S comes to believe that P based solely on their experiences in virtual reality. Is it possible for S to know that P? According to one influential theory of knowledge, they cannot. Robert Nozick's famous truth-tracking analysis requires that, in addition to a belief being true, it must also be sensitive to the truth:[1]

> **Sensitivity**: If it had not been that case that P, then S would not have believed that P.

Now it looks like any belief formed solely on the basis of virtual reality will fail sensitivity. In the nearest possible world where P is false, S would have had the same experiences, and therefore S would have continued to have believed that P. Whilst S's belief that P might be sensitive to the *virtual* world, it will fail to be

[1] In addition to the truth of P, S believing that P, and the sensitivity principle, Nozick's full analysis requires a fourth condition called 'adherence:' if it had been the case that P, then S would have believed that P. Given that the main problem for knowledge from virtual reality stems from the sensitivity principle, this will be the main focus of the paper.

© LOGOS & EPISTEME, XI, 3 (2020): 369-388

sensitive to the *real* world. Nozick himself recognized as much when he discussed his account in connection with the famous brain-in-a-vat thought experiment:

> There remains, for example, the case of the person in the tank who is brought to believe, by direct electrical and chemical stimulation of his brain, that he is in the tank and is being brought to believe things in this way. The person in the tank does not know he is there, because his belief is not sensitive to the truth… The operators of the tank could have produced any belief, including the false belief that he wasn't in the tank; if they had, he would have believed that.[2]

Nozick does not consider this a problem because he does not believe that a person in the brain-in-a-vat scenario is capable of knowledge. Even if the scientist decided to be honest and reveal truths to the envatted person, they would still not have knowledge—if the scientist had induced the same beliefs in a world where they were false, the envatted person would have continued to believe them.

However, Nozick's refusal to permit knowledge in 'virtual worlds' is beginning to look increasingly untenable. There are two main reasons for this. Firstly, philosophers have started to recognize the value of virtual reality as a source of belief and knowledge about the world.[3] Writers such as Jon Cogburn and Mark Silcox,[4] for example, make a comparison between virtual reality and other fictional media, such as novels, movies and computer games. Although fictional, these media can contain truth, and under the right conditions can provide knowledge about the world.[5]

Secondly, and perhaps more importantly, there is growing empirical evidence for the claim that knowledge can be attained from virtual reality. For at least five decades, virtual reality has been used in some form or another to train

---

[2] Robert, Nozick, *Philosophical Explanations* (Cambridge, MA: Belknap Press, 1981), 175.

[3] See for example Jon Cogburn and Mark Silcox, "Against Brain-in-a-Vatism: On the Value of Virtual Reality," *Philosophy & Technology* 27, 3 (2014): 561-579; Eva Dadlez, "Virtual Reality and 'Knowing What It's Like:' The Epistemic Upside of Experience Machines," in *Experience Machines: The Philosophy of Virtual Worlds*, ed. Mark Silcox (Lanham: Rowman & Littlefield International, 2017), 75–86; James McBain, "Epistemic Lives and Knowing in Virtual Worlds," in *Experience Machines: The Philosophy of Virtual Worlds*, 155–168.

[4] Cogburn and Silcox, "Against Brain-in-a-Vatism," 561-579.

[5] For the wider discussion of knowledge from fiction (which does not differentiate virtual reality from other forms of fictional media) see Axel Spree, "Fiction, Truth and Knowledge" in *From Logic to Art: Themes from Nelson Goodman*, eds. Gerhard Ernst, Jakob Steinbrenner and Oliver Scholz (Paris: Walter de Gruyter, 2009), 329–344; Kathleen Stock, "Learning from Fiction and Theories of Fictional Content," *Teorema: Revista Internacional de Filosofía* 35, 3 (2016): 69–85; AsbjørnSteglich-Petersen, "Fictional Persuasion and the Nature of Belief," in *Art and Belief*, eds. Ema Sullivan-Bissett, Helen Bradley, and Paul Noordhof (Oxford: Oxford University Press, 2017), 174–193.

individuals in the skills needed to fly planes, land spacecraft, and perform certain surgical operations. More recently, so-called 'educational VR' is being touted as a replacement to traditional teacher and textbook-led instruction. The aim is to create virtual worlds that represent objects and events that are difficult to explore experientially in the real world. There is significant evidence that individuals can come to a better understanding and gain new knowledge about objects and their behaviour in the real world as a result.[6]

If knowledge is attainable from experience in virtual reality, then it would show that Nozick's sensitivity condition is not necessary for knowledge. Virtual reality would join a list other belief-forming methods, such as induction, introspection and testimony that—whilst commonly believed to provide knowledge—nonetheless fail to meet sensitivity.[7] However, just as it has been argued that the appearance of insensitivity in these methods only emerges when the sensitivity principle has been improperly applied, so I will argue much the same is true for virtual reality.[8] Although Nozick's original sensitivity condition is too strong to account for knowledge from virtual reality, a suitably modified version of it—one that takes into consideration stages of belief formation within virtual worlds—can account for the correct cases in which knowledge is attained.

---

[6] For examples see the discussion in section 2.

[7] The case from induction has been discussed by Jonathan Vogel, "Tracking, Closure, and Inductive Knowledge," in *The Possibility of Knowledge: Nozick and His Critics*, ed. Steven Luper-Foy (London: Rowman & Littlefield, 1987), 197–215; Ernest Sosa, "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999): 137–49; Duncan Pritchard, "In Defence of Modest Anti-Luck Epistemology," in *The Sensitivity Principle in Epistemology*, eds. Kelly Becker and Tim Black (Cambridge: Cambridge University Press, 2012), 173–192. The case from introspection has been given by Jonathan Vogel, "Reliabilism Leveled," *Journal of Philosophy* 97, 11 (2000): 602-623; Ernest Sosa, "Rational Intuition: Bealer on its Nature and Epistemic Status," *Philosophical Studies* 81, 3-2 (1996): 151-162; Ernest Sosa, "Tracking, Competence, and Knowledge," in *The Oxford Handbook of Epistemology*, ed. Paul Moser (Oxford: Oxford University Press, 2002), 264–287. For a discussion of testimony see Stanford Goldberg, "Sensitivity from Others," in *The Sensitivity Principle in Epistemology*, eds. Kelly Becker and Tim Black (Cambridge: Cambridge University Press, 2012), 43-65; Tristan Haze, "Two New Counterexamples to the Truth-Tracking Theory of Knowledge," *Logos & Episteme* 6, 3 (2015): 309-311.

[8] Alternative versions of the sensitivity principle that have been given to solve these and other problems can be found in Joseph Salerno, "Truth Tracking and the Problem of Reflective Knowledge," in *Knowledge and Skepticism*, eds. Joseph Campbell, Michael O'Rourke, and Harry Silverstein (Cambridge, MA: MIT Press, 2010), 72–81; Goldberg, "Sensitivity," 43-65; Fred Adams, John Barker, and Murray Clark, "Knowledge as Fact-Tracking True Belief," *Manuscrito* 40, 4 (2017): 1-30; Kevin Wallbridge, "Sensitivity, Induction, and Miracles," *Australasian Journal of Philosophy* 96, 1 (2018): 118-126.

In the next section I start by defining more clearly what I mean by virtual reality and the kinds of experiences that I am concerned with as a method of belief formation. James McBain[9] has recently proposed an account of knowledge in virtual reality as a response to what he considers are the flaws in Nozick's own theory. He uses Dretske's information-theoretic account in order to do so, but as I shall show, the central idea can be captured using a variation of sensitivity, what I call 'virtual sensitivity.' In section 4 I outline two objections to the McBain-inspired virtual sensitivity principle that show it is neither necessary nor sufficient for knowledge from virtual reality. Finally in section 5 I outline a new sensitivity principle, 'virtual sensitivity+,' that I argue overcomes the problems with McBain's account and can explain the cases in which knowledge from virtual reality is attainable.

2. Virtual Reality as a Belief-Formation Method

In the original presentation of his tracking theory Nozick realized that the sensitivity principle needed to be relativized to the method through which the belief was arrived at. He illustrates this with his well-known 'grandmother case:'

> A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset. Yet this does not mean she doesn't know he is well (or at least ambulatory) when she sees him. Clearly, we must restate our conditions to take explicit account of the ways and methods of arriving at belief.[10]

This example fails sensitivity even though intuitively the grandmother has knowledge. In the nearest possible world where P is false (where P = 'her grandson is well'), the grandmother would continue to believe that P. In this possible world the grandmother would have believed it using a different method. Instead of using perception she would have based her belief on testimony. It seems reasonable then that when judging whether or not sensitivity has been satisfied, we must keep constant the method being used to arrive at a belief.

Although I won't argue for it here, I believe there are good grounds for treating virtual reality as a distinct method by which we can come to arrive at beliefs—one that depends causally on other cognitive faculties (much like testimony does) but sits somewhere between testimony and instrument-based belief.[11] For better or worse, Nozick himself thinks we can individuate methods

[9] McBain, "Epistemic Lives", 155–168.

[10] Nozick, *Philosophical Investigations*, 179.

[11] It is feasible, for example, that a virtual world could be designed and constructed in a completely automated fashion. Imagine a space probe that scans and maps some distant planet

simply based on our experience of them and how distinct they 'feel' to us.[12] For most it will not be obvious that virtual reality is a distinct method since few have (yet) had the chance to experience it. So what I aim to do here is to give some examples of the way virtual reality is currently being used, especially in training and education contexts, and provide a general outline of the main steps and cognitive processes used in arriving at beliefs through it.

It will be useful to begin with a definition of virtual reality. Although not uncontentious, I will follow the most widely held definition given by Howard Rheingold[13] and Michael Heim[14] according to which virtual reality is a computer-generated sensory experience that is both *immersive* and *interactive*. Immersion is a difficult idea to define precisely but for present purposes we can think of it as the subjective feeling of presence inside the virtual world produced by a computer. These feelings are generated by experiencing a sensory interface that can involve a range of technologies including: head-mounted displays, virtual reality rooms, surround sound headphones, and haptic equipment (such as suits and gloves) that provide feelings of force, motion, and even temperature to the user. What also separates virtual reality from other types of immersive media (like movies) is its interactivity. By making decisions via an input device, the user can change the outcome of the experiences that are being fed to them by the computer.[15]

In what ways can virtual reality be used to arrive at new beliefs about the world? Training simulators that use virtual reality have been used for decades in industries such as the military, healthcare, and aerospace.[16] In most of these cases

---

and sends the data back to earth that is then automatically rendered into an immersive 3D virtual world for scientists to explore. This seems distinct from testimony and yet in an important sense depends on the human design of the hardware to reliably gather data and portray the distant planet. Whether or not virtual reality is a distinct method of belief-formation from testimony or instrument-based belief will not matter much for my argument. What matters is that beliefs about the real world can be generated from experiences in this way.

[12] Nozick, *Philosophical Investigations*, 184.

[13] Howard Rheingold, *Virtual Reality: The Revolutionary Technology of Computer-Generated Artificial Worlds - and How It Promises to Transform Society* (New York: Simon & Schuster, 1992).

[14] Michael Heim, *The Metaphysics of Virtual Reality* (Oxford: Oxford University Press, 1993).

[15] These conditions provide a broad definition of virtual reality that will include not only state-of-the-art forms of virtual reality that utilize head-mounted displays, but also more familiar interactive media such as video games and training simulators. More narrow definitions are possible, but in these cases one needs to specify either precisely the hardware involved or the level of immersion produced. I will follow others who write on this issue by sticking to the broad definition, even if it includes experiences we do not normally call 'virtual reality.'

[16] Derek Stanovsky, "Virtual Reality," in *The Blackwell Guide to the Philosophy of Computing*

virtual reality is used to train individuals in the skills needed to operate complex equipment, such as an aircraft or space probe. According to education theorists, what makes virtual reality so good at this is that it provides 'situated learning' opportunities that are difficult to have in the real world due to cost and safety concerns.[17] The kind of knowledge that is gained from training simulators is skills-based knowledge or 'knowledge-how.' But more recent virtual reality programs have been created that aim to provide factual knowledge or 'knowledge-that.' The developers of these programs aim to replicate the advantages of situated learning that have been found in training simulators by applying virtual reality to more factual learning outcomes. Examples of programs that have already been developed include:

- River City (medicine and epidemiology)[18]
- Supercharged! (electrostatic forces)[19]
- Virtual Cell (cell biology)[20]
- Immune Attack! (immunology)[21]
- Whyville (basic scientific concepts)[22]
- Quest Atlantis (history)[23]
- EcoMUVE (ecosystems)[24]

---

and Information, ed. Luciano Floridi (Oxford: Blackwell, 2004): 167–177.

[17] Christian Schott and Stephen Marshall, "Virtual Reality and Situated Experiential Education: A Conceptualization and Exploratory Trial," *Journal of Computer Assisted Learning* (2018): 1-10.

[18] Chris Dede, "Immersive Interfaces for Engagement and Learning," *Science* 323, 66 (2009): 66-69.

[19] Janice Anderson and Mike Barnett, "Learning Physics with Digital Game Simulations in Middle School Science," *Journal of Science Education and Technology* 22, 6 (2013) 914–926.

[20] Tassos A. Mikropoulos, Apostolos Katsikis, Eugenia Nikolou, and Panayiotis Tsakalis, "Virtual Environments in Biology Teaching," *Journal of Biological Education* 37, 4 (2003): 176-181.

[21] Melanie Stegman, "Immune attack players perform better on a test of cellular immunology and self confidence than their classmates who play a control video game," *Faraday Discuss* 169 (2014)*:* 403-423.

[22] Carlos Monroy, Yvonne Klisch, and Leslie Miller, "Emerging Contexts for Science Education: Embedding a Forensic Science Game in a Virtual World," *Proceedings of the 2011 I-Conference: Inspiration, Integrity, and Intrepidity* (New York: Association for Computing Machinery, 2014): 622-629.

[23] Sasha Barab, Tyler Dodge, HakanTuzun, Kirk Job-Sluder, Craig Jackson, Anna Rici, Laura Job-Sluder, Robert Carteaux, Jo Gilbertson, Cohan Heiselt, "The Quest Atlantis Project: A Socially-Responsive Play Space for Learning," in *The Design and Use of Simulation Computer Games in Education*, eds. Brett Shelton and David Wiley (Rotterdam: Sense Publishers, 2007): 159–186.

For the sake of illustration, let us take a closer look at the first of the programs on this list.

River City was developed by Chris Dede and his colleagues at Harvard University and produced by Acitiv worlds, Inc. It is an immersive virtual reality platform that aims to teach young people about diseases and disease transmission. Users immerse themselves in a fictional 19th century city and learn to behave like scientists. Their aim is to understand why the inhabitants of River City are getting sick and what to do in order to prevent further infection. They do this by 'talking to various residents in the simulated setting, such as children and adults who have fallen ill, hospital employees, merchants, and university scientists.' In the process users 'learn to identify problems through observation and inference, form and test hypotheses, and deduce evidence-based conclusions about underlying causes.'[25]

Research shows that students who partake in the virtual River City program have much higher rates of success in transferring what they have learnt inside the simulation to the real world. According to Dede:

> Our research results from River City show that a broader range of students gain substantial knowledge and skills in scientific inquiry through immersive simulation than through conventional instruction or equivalent learning experiences delivered via a board game. Our findings indicate that students are deeply engaged by this curriculum through actional and symbolic immersion and are developing sophisticated problem-finding skills. Compared with a similar, paper-based curriculum that included laboratory experiences, students overall were more engaged in the immersive interface and learned as much or more.[26]

Similar results have been found in other studies of educational virtual reality programs.[27] The key concept here is that of 'transfer,' where a belief or fact learnt inside a *virtual world* is upheld or turned into a belief about the *real world*.[28] This

---

[24] Tina Grotzer, Amy Kamarainen, Shari Metcalf, Shane Tutwiler, and Chris Dede, "Teaching the Systems Aspects of Epistemologically Authentic Experimentation in Ecosystems through Immersive Virtual Worlds," Paper presented at *The National Association of Research in Science Teaching (NARST) Conference*, San Antonio, TX, (April 23, 2017).

[25] Chris Dede, "Immersive Interfaces," 67.

[26] Chris Dede, "Immersive Interfaces," 67.

[27] In particular see the studies by Brian Nelson and Diane Ketelhut, "Scientific Inquiry in Educational Multi-user Virtual Environments," *Educational Psychology Review* 19, 2 (2007): 265–283; Merrilea Mayo, "Video Games: A Route to Large-Scale STEM Education?," *Science* 323, 5919 (2009): 79–82; Barney Dalgarno and Mark Lee, "What are the learning affordances of 3-D virtual environments?," *British Journal of Educational Technology* 41, 1 (2010): 10–32.

[28] Chris Dede, Jeffrey Jacobson, John Richards, "Introduction," in *Virtual, Augmented, and Mixed Realities in Education*, eds. Dejian Liu, Chris Dede, Ronghuai Huang, and John Richards (Singapore, Springer, 2017), 6.

suggests that the cognitive process or method involved in forming a belief about the real world includes at least two stages. Firstly, a set of internal cognitive methods are used, such as perception, deduction, induction, etc., in order to arrive at beliefs about the virtual world. Then the participant uses these beliefs to generate a further belief about the real world (see Fig. 1).

It is very likely that beliefs formed inside virtual reality and that are about virtual worlds will have a different meaning or semantic content to beliefs about the real world. Even if I experience objects in a virtual world that are perceptually similar to objects in the real world, such as snow, stop signs, tigers, etc., my beliefs will be about tokens of these 'virtual objects' only, and will not typically include tokens of them in the real world. If I come to believe that 'snow is white' in the virtual world, and on this basis come to form a further belief that 'snow is white' in the real world, then I will have two separate beliefs. The first is a belief about the snow in the virtual world and its property of whiteness, whereas the second is a belief about the snow in the real world and its property of whiteness.

Figure. 1: Virtual World to Real World Belief Transfer

This raises difficult questions such as: 'how do beliefs and statements represent or get to be about virtual worlds?,' 'under what conditions are statements about virtual worlds true?,' and 'what is the ontological status of virtual objects?'. I will not attempt to answer these questions in any detail here. To do so would orientate the discussion too far away from the main epistemological question I want to answer. Instead I refer the interested reader to the ongoing work that is currently being undertaken in this area.[29]

[29] For a pragmatic approach to truth in virtual worlds see Michael Heim, *Virtual Realism* (Oxford: Oxford University Press, 2000). Ilkka Niiniluoto adapts a possible world semantics for fictional worlds and applies it to virtual reality: Ilkka Niiniluoto, "Virtual Worlds, Fiction, and Reality," *Discusiones Filosóficas* 12, 19 (2011):13-28. Theories of the metaphysics of objects in virtual reality include various 'realist' accounts, such as those of David Chalmers, "The Virtual and the Real," *Disputatio* 9, 46 (2017): 309-352; Myeung-Sook Yoh, "The Reality of Virtual Reality," in *Proceedings Seventh International Conference on Virtual Systems and Multimedia*, (Berkeley: CA, 2001): 666-674; Espen Aarseth, "Doors and Perception: Fiction vs. Simulation in Games," *Intermédialités: Histoire et Théorie des Arts, des Lettres et des Techniques* 9, 34 (2007):

Nonetheless, I think there is a very intuitive sense in which a belief or a proposition about a virtual world can be separated from a belief or a proposition about the real world. Literature and other fictional media provide more familiar examples of this. Although the statement 'levitation is possible' is true in the world of *Harry Potter*, this statement is false in the real world. Following David Lewis,[30] we might suppose that the difference between them is recognized in ordinary discussion by an implicit prefix of the kind 'In the *Harry Potter* stories, …' that is attached to the first, but not to the second. Although this is far from a complete analysis of the difference in semantic content between these two statements, it provides a useful way for us to differentiate similar sounding claims about a virtual world from the real world. When an explicit distinction is called for, let us indicate this by using 'P_V' for a statement P that is made about a virtual world, and 'P_R' for a statement P is that is made about the real world.

Just like we intuitively recognize a difference between statements about virtual worlds and the real world, we also allow for some of these statements to be true and others false. It is natural, for example, when talking about the video game *Super Mario Bros* to say that 'Mario wears red overalls' is true, whereas the statement 'Mario wears green overalls' is false. Again, how one explains this depends very much on one's theory of truth for virtual worlds and the ontological status of the objects and properties that make these statements true. However, there is one difference between virtual worlds and real worlds that is worth highlighting in connection here. Unlike a work of literature, what is 'true' in a virtual world depends on more than just its program and the intentions of its original creator. It will also depend on the functioning of the hardware that runs the program and the input provided by a user. In addition, there are facts about a virtual world that might lay dormant in its programming because a user did not provide the right input in order for it to manifest. Yet in these cases, we would still want such content to be part of the virtual world, even if it is never actually experienced by a user. If there is a secret level in which Mario's overalls turn

35-44. A phenomenological approach to virtual objects is given in Philip Zhai, *Get Real: A Philosophical Adventure in Virtual Reality* (London: Rowman & Littlefield, 1998). Fictionalist approaches have been developed by Jesper Juul, *Half-Real: Video Games Between Real Rules and Fictional Worlds* (Cambridge, MA: MIT Press, 2005); Cogburn and Silcox, "Against Brain-in-a-Vatism," 561-579. A conditional or dependent realist view (where the reality of virtual objects depends on reproducing qualities of their physical counterparts in the real world) has been developed by Philip Brey, "The Physical and Social Reality of Virtual Worlds," in *The Oxford Handbook of Virtuality*, ed. Mark Grimshaw (Oxford: Oxford University Press, 2014): 42–54.

[30] David Lewis, "Truth in Fiction," *American Philosophical Quarterly* 15, 1 (1978): 37-46.

green, we want a statement asserting this fact to be true, even if nobody experiences a 'green Mario.'

Because each implementation or run of the program will produce different states of affairs, it is possible that on different occasions incompatible statements will obtain. For example, for a video game played on one occasion it might be true that 'the boss gets defeated,' whereas on another occasion the statement might be false. Does this imply that there is no consistent conception of truth that can be applied to statements about virtual worlds? I do not believe this observation is necessarily problematic. In these kinds of cases, we might suppose that the language is being used in a rather imprecise way. If I utter the statement 'Manchester United beat Chelsea,' then this is both true and false, since in their histories of competitive soccer, Manchester United has sometimes won against Chelsea and sometimes lost. To be more precise, I would need to qualify the statement to refer to the time or to the actual match played. A similar device could be used in the case of propositions about virtual worlds, where the proposition is suitably indexed to either the time or the run of the program.

3. McBain on Knowledge and Virtual Reality

The case against sensitivity on the basis of virtual reality has been made recently by James McBain. Even if a belief formed in a virtual world is true, the person who holds that belief will not have knowledge because it will fail to be sensitive to the real world:

> The person plugged in is not sensitive to that which is true of the situation—that she is being fed stimuli about the world she is experiencing. The method by which she is arriving at her believing this does not counterfactually hold. The details of the world she is experiencing could be changed by the operators. What she is sensitive to is the stimuli, not the world. Therefore, she would not have knowledge in the machine. While the person plugged into the machine will have lots of beliefs about the virtual world she is in, none of those beliefs, on Nozick's account, will constitute knowledge.[31]

If the designers of an educational VR program, such as River City, had decided to make the virtual diseases behave in ways quite unlike the real world, then the user would have believed this instead—despite it being false. Their beliefs would not track the truth. Yet intuitively, and empirically, students immersed in the River City program can gain knowledge about how diseases function in the real world.

[31] McBain, "Epistemic Lives," 159.

Whilst staying committed to a broadly externalist epistemology, McBain attempts to explain how knowledge is possible in virtual reality in a way that overcomes the shortcomings of Nozick's sensitivity principle. His solution is to move to an alternative framework, one that includes not counterfactuals, but 'reliable information flow.' Put simply, according to McBain, knowledge in virtual reality is possible provided the beliefs formed are true and are reliably connected to the content of the virtual world. In order to explicate this idea more precisely he adopts Fred Dretske's analysis[32] of knowledge as true belief caused by information. There is one important difference, however. Whereas Dretske envisioned the source of information to be a fact or event in the world itself, McBain allows for the source to be a fact or event in the virtual world.

Although McBain's account describes knowledge in terms of Dretske's theory, he suggests it is consistent with Nozick's theory of knowledge and could be reworked with the aid of a modified sensitivity condition.[33] This is what I will attempt to do in this section. First I will explain in more detail how McBain uses Dretske's theory to account for knowledge in virtual reality. Then I will reformulate its central idea in terms of counterfactuals to arrive at a version of sensitivity that can capture the advantages McBain believes his theory has over Nozick's classic tracking approach.

Dretske's original account of knowledge was meant to apply to perceptual belief, and given that perception also plays an important role in knowledge from virtual reality, it seems a suitable place to start. Dretske's central idea is that if a truth-maker for a proposition P transmits the information that 'P' along a channel, such that it is received by S and causes S to believe that P, then S knows that P.

His account relies on a number of key concepts such as 'reliable channel' and 'information'—ideas that are difficult by themselves to understand philosophically. To this end, Dretske utilizes Claude Shannon and Warren Weaver's mathematical theory of communication when thinking about reliability and information flow. According to Shannon and Weaver, information should be measured in terms of the amount of *uncertainty* reduced when a choice is made from a range of possible outcomes.[34] If the probability of each outcome has an objective value, then a numerical value can be assigned to indicate how much information each outcome provides. This quantity of information or 'entropy' is measured in terms of the number of binary digits (or bits) needed to individually encode that message.

---

[32] Fred Dretske, *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1981).

[33] McBain, "Epistemic Lives", 162.

[34] Claude Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana: The University of Illinois Press, 1949), 1-2.

If one thinks of a source of information x as an object that can be in one of many different states, then a reliable channel can be defined as one that successfully transmits the same *amount of information* along the channel as is generated at the source:

> **Reliable Channel:** In order for a signal r to reliably carry the information that a source x is in state F, the following need to be met:
>
> (i) x is in state F
>
> (ii) The signal carries the same amount of information (i.e. in bits) as would be generated if x was in state F
>
> (iii) The amount of information the signal carries about the source x is or includes the quantity generated by x being in state F (and not by x being in state G)[35]

Although this tells us when a channel reliably communicates the same amount of information (in bits), it tells us nothing about whether it reliably communicates the same message in terms of its semantic content. If the source is the rolling of a die, then it will transmit the same amount of information whether it lands on a 5 or a 6. Dretske resolves this with the following additional definition:

> **Semantic Content:** A signal r carries the information that 'x is in state F' if, and only if, the conditional probability of x being in state F given that the signal r transmits the message that 'x is in state F' is 1 (but less than 1 given the receiver's background beliefs alone).[36]

This definition claims that a successful transmission of a message occurs only when there is a lawlike connection between the fact itself and the reception of the message. In other words, it is impossible for a channel to transmit the message that 'x is F' unless x is in state F. As he says 'false information, misinformation, and disinformation are not varieties of information—any more than a decoy duck is a kind of duck.'[37] Dretske includes the thought that the probability might be less than 1 given the receiver's background beliefs as a concession to the way we ordinarily talk about information.[38] For example, even if the information 'x is in state F' is true and this is received, it might be not be informative to somebody who already knows it is the case.

Putting all this together, Dretske defines knowledge as follows:

---

[35] Modified from Dretske, "Knowledge," 63-64.

[36] Modified from Dretske, "Knowledge," 65.

[37] Fred Dretske, "Précis of Knowledge and the Flow of Information," *Behavioral and Brain Sciences* 6, 1 (1983): 57.

[38] Dretske, "Précis," 57.

> **Knowledge**: S knows that 'x is in state F' if, and only if, S's belief that 'x is in state F' is caused by the information that 'x is in state F'.[39]

As an illustration suppose that P is true and P = 'there are 5 apples in the basket.' The world, viewed as a source of information, is in one of many different states and is therefore able to transmit an objective quantity of information. This information is transmitted to the receiver through light waves to the eyes and interpreted to form a belief. If the arrangement of light waves is such that its state mirrors (in terms of number of bits) the state of the world when there are 5 apples in the basket, then it is a reliable channel. If the content of the message is such that it would not have been transmitted unless the odds of it being true are 1, then the right semantic information is also transmitted.

According to McBain, this framework for defining knowledge from ordinary perceptual belief can be extended to virtual reality in the following way:

> Once one is hooked up to the machine (or, currently, puts the headset on), the designed world will send signals to the user about that world. The signals will carry as much information about the feature, item, or event being the case in the virtual world as would be generated by that feature, item, or event being the case in the world. Once the user receives the signal, she combines that with any relevant background knowledge about the world, all the while supplementing any gaps with knowledge of the actual world. There is an objective probabilistic connection between what the interface gives and the virtual world being such and such way. If the probability of the virtual world being such and such way when the interface informs us that is 1, then we have grounds for believing the world is that way.[40]

The idea seems to be this: the source of S's belief need not be a fact or event in the real world in order to count as knowledge. Provided the fact or event that occurs in the virtual world carries the same amount of information and the same semantic content as a belief about the real world (and is true), then this can provide knowledge—despite the information not having its origin in the world itself. As McBain puts it: "contra Nozick, [knowledge] is not a matter of where the signals arise."[41] For example, if S has a perceptual experience of 5 apples in a basket whilst in virtual reality, then S can come to know that 'there are 5 apples in the basket' in the real world provided: (i) there are 5 apples in the basket, (ii) the source of information (i.e. the virtual world) is in a state such that it provides just as much information in bits as 5 apples being in a basket, and (iii) S's belief that

[39] Modified from Dretske, "Knowledge," 65.
[40] McBain, "Epistemic Lives," 166.
[41] McBain, "Epistemic Lives," 167.

'there are 5 apples in the basket' is caused by the information that 'there are 5 apples in the basket' that it received from virtual reality.[42]

There are a number of things that need to be said about McBain's proposal. Firstly, it is clearly an idealization as it currently stands. Few, if any, existing virtual reality programs can transmit the 'same amount of information' about a virtual object or event as would be received from perceiving it in the real world. Virtual worlds are just not that detailed. Having said that, it could be argued that this a short-term problem, and we can certainly imagine in the future virtual reality software and hardware capable of producing experiences that are as detailed as the real world.

Secondly, we have seen that beliefs about the real world formed on the basis of experiences in virtual reality are likely to have a different semantic content to similar beliefs held about the virtual world itself. Again, this is not necessarily a problem for McBain's proposal. What is needed is to recognize a distinction between the content of the belief about the virtual world and the content of the real world. This does, however, require McBain to restate the supposed connection that obtains between the belief in the virtual world and the real world. It is not enough to say that the belief about the virtual world have 'the same information and semantic content' as a belief about the real world. Even if the worlds are equally detailed (and so contain the same information *vis-à-vis* their entropy) they will necessarily have different content because one is about virtual objects and the other is about real objects. The connection would need to be spelled out in terms other than content identity (perhaps causation or counterfactuals), but I don't see this as being fatal to the proposal.

That McBain's account probably needs supplementing with counterfactuals or a causal connection suggests that it might be worth rephrasing it in terms more familiar to the tracking theory. If we use counterfactual dependence as our mark of reliability in belief formation rather than Dretske's reliable information channel, then McBain's account comes out as follows:

McBain-Inspired Tracking: S knows that P_R only if,

(i) P_R and P_V are true,

[42] At times, it reads as if McBain is only concerned to explain how knowledge of the virtual world is possible, rather than knowledge from virtual reality. However, if this is the case then the comparison of his view to Nozick's comments on truth tracking seem ill-placed, as when Nozick rejects knowledge in the machine, he clearly has knowledge of the real world in mind. In the rest of the paper I will use McBain's account as a foundation for explaining knowledge *from* virtual reality, even if McBain did not originally intend it to be used in this way.

(ii) S believes that P_R and S believes that P_V

(iii) If it had not been that case that P_V then S would not have believed that P_V

These conditions capture the core idea latent in McBain's theory: knowledge is attainable from virtual reality provided the user forms a belief about the real world that is true and they base this belief on a belief about the virtual world that is reliably connected to the content of that world. Does this provide an adequate solution to the problem? Can we now say precisely under what circumstances knowledge from virtual reality is attainable? Unfortunately, as I will now argue, the combined conditions (i)-(iii) are neither necessary nor sufficient for knowledge from virtual reality.

4. Problems for Virtual Sensitivity

The proposal given in section 3 on the basis of McBain's account of knowledge effectively weakens Nozick's classic sensitivity in favor of what we might call 'virtual sensitivity.' Instead of a belief formed in VR being sensitive to the facts of the real world (i.e. the facts that make it true), S can have knowledge provided their belief is based on a belief about the virtual world that is sensitive to the facts of the virtual world.

> **Virtual Sensitivity:** Where S bases their belief that P_R on their belief that P_V, if it had not been the case that P_V, then S would not have believed that P_V.

For participants inside a virtual world, clearly this condition is much easier to satisfy than classic sensitivity. A user who is using an educational VR program will continue to believe what they are experiencing even in a world where it is false. But it seems unlikely they would continue to believe it if it were not true in the virtual world itself.

Is virtual sensitivity enough to guarantee knowledge from virtual reality? One worry that immediately emerges is that beyond the condition that P_R is true, there is no reliable method or means connecting S's belief that P_V to the fact that P_R. This raises the logical possibility that a person might come to form a belief about the real world based on one that is true in the virtual world, but that nonetheless is only true by luck. Consider the following case:

> COMPUTER MALFUNCTION: A new education VR program has gone to market. In its current form it contains a falsehood about the real world. Whereas P is true in the real world, P is false in the virtual world of the education program. S buys the program and runs it on their computer. However, their computer has a malfunction that incorrectly reads not-P as P, and so when is implemented, creates a visual experience of P. On this basis S comes to believe that P is true, both in the virtual world and the real world.

In this case S has a true belief, both about the virtual world and the real world, and their belief about the virtual world is sensitive to the truths of the virtual world. Yet I would argue that in this case S does not know that P. It might be argued that S is not forming beliefs based on the 'right version' of the VR program, and if their machine had not malfunctioned, their belief would have satisfied virtual sensitivity. But who is to say what the right version is? As we saw in section 2, the virtual world is a combination of many factors: the written program, its implementation on a computing device, and the decisions made by the user. We cannot simply exclude P_V from being true in a virtual world by saying it was not the one intended by the program writer. Many features of virtual worlds were not the result of the intentions of the programmers, often because of complexity and unforeseen consequences. Sometimes this is part of the appeal, but there seems to be no way of excluding this in principle.

It looks as if any proposal along the lines of virtual sensitivity needs to be supplemented with further conditions. It might be thought that it is no surprise that virtual sensitivity (along with true belief) is insufficient for knowledge. After all, Nozick's original tracking theory had two counterfactual conditions: the classic sensitivity condition and his adherence condition. Perhaps the addition of an adherence condition in line with McBain's overall approach is what is called for.

> **Virtual Adherence:** Where S bases their belief that P_R on their belief that P_V, if it had been the case that P_V, then S would not have believed that P_V.

Unfortunately, this does little to change the outcome of the COMPUTER MALFUNCTION case. Here the nearest possible where P_V is true is the actual world, and in this world S believes that P_V. So even though this condition is satisfied, S still does not have knowledge that P_R.

Part of the problem with McBain's proposal (whether couched in terms of counterfactuals or reliable information flow) is that whilst the user's beliefs are sensitive to the facts of the virtual world, they are not sensitive to the real world. This type of sensitivity must be included in the definition to exclude lucky true beliefs based on virtual reality. In the next section I will propose a way in which this can be done that does not require the strength of Nozick's original sensitivity principle.

Before that, however, I want to consider a different case against virtual sensitivity. This case is important for motivating the view I propose in section 5 because it suggests that a user in VR does not need to have knowledge about the virtual world itself:

> SPECTRUM INVERSION: A virtual reality program has been designed that deviates systematically from the real world. Every 5 minutes once per hour the

> virtual world inverts the colors that are experienced by a user. A user S has experienced this world many times and this has caused their brain to compensate for the inverted periods. During an inverted period, the virtual world displays a blue stop sign, however S comes to believe that the stop sign in the virtual world is red. On this basis S comes to form the belief that stops signs in the real world are also red.

In this case S has a true belief about the real world but a false belief about the virtual world. During the period in which they form the belief, the stop sign is blue, and their belief that it is red is false. Yet arguably in this case, S *does know* that stop signs are red in the real world. Their belief is connected to the truth in the real world in a non-accidental way. Overall, the medium through which they come to form beliefs (the virtual world) is reliable, and when combined with their own internal compensation for the inverted periods, provides beliefs that are sensitive to the real world.

What this suggests is that even though a belief about the real world may involve, as part of its causal history, an experience and a belief about a virtual world, that belief about the virtual world does not need to be true. In fact, their belief about the virtual world does not even need to be sensitive to the truths of the virtual world. In the SPECTRUM INVERSION case, where P_V = 'stop signs are red,' in the nearest possible world where this is false (the actual world) they believe it. Their belief fails virtual sensitivity and yet somehow they have knowledge of the real world.

The upshot of this example is that although beliefs within a virtual world form part of the causal history for a belief about the real world, their connection to the truth of the virtual world is not significant for knowledge of the real world. These beliefs play a *cognitive* role (they are part of the method of belief-formation) but they do not play an *epistemic* role. In other words, knowledge of the real world based on an experience in virtual reality does not first require a person to have knowledge about the virtual world. Perhaps this should not come as a surprise. Most externalist views of perception and testimony, for example, only require that a person's belief be connected to the facts in the right way; they do not demand the stronger requirement that they first have beliefs or knowledge about the various stages in the generation of their belief. Having a true belief about the virtual world is clearly important for knowledge about the virtual world, but when thinking about using virtual reality to gain knowledge about the real world, the truth of these beliefs no longer seems necessary.

5. Fixing the Problem: Virtual Sensitivity+

In this final section I will introduce a new version of sensitivity that is stronger than the McBain-inspired virtual sensitivity principle, but weaker than Nozick's original classic sensitivity principle. The motivating idea will be that where a belief about the real world is based on one about the virtual world, a person can come to know about the real world provided there is a non-accidental, lawlike, connection between the facts of the real world and the beliefs they form inside the virtual world. In other words, their beliefs about the virtual world are sensitive to the real world.

This idea allows us to separate brain-in-a-vat scenarios where an envatted person can have knowledge from those in which they cannot. Consider the following two cases:

> EVIL SCIENTIST: S has been envatted her entire life and is fed sensory stimuli by an evil scientist. The evil scientist uses all kinds of methods to decide on the content of the world experienced by S. One day the scientist uses a random number generator to decide how many fingers S's avatar will have in the virtual world. The outcome of the random number generator is 5 and S's avatar in the virtual world comes to have 5 fingers.
>
> BENEVOLENT SCIENTIST: S has been envatted her entire life and is fed sensory stimuli by a benevolent scientist. The benevolent scientist uses the best methods they can to ensure that the content of the world experienced by S is as close as possible to the real world. Knowing that humans have 5 fingers, the scientist programs the virtual world so that S's avatar has 5 fingers.

Now let us suppose that, for one reason or another, S is given a human body and returns to the real world. Many of the beliefs she formed whilst a brain in a vat are transferred to beliefs about the real world. Do any of these constitute knowledge? A good case can be made for arguing that beliefs formed in the EVIL SCIENTIST scenario will not count as knowledge (even if true), whereas those formed in the BENEVOLENT SCIENTIST scenario will. In the BENEVOLENT SCIENTIST case, if it had been true that humans had 6 fingers, then it is probable that S would have believed this about her avatar's hands instead. In his attempt to make her experiences as close to the real world as possible, it is likely that the benevolent scientist would have included this fact in the virtual world and that S would have come to believe it.

The sensitivity that is called for here is one that is stronger than virtual sensitivity. For that reason, let us call it 'virtual sensitivity+':

> **Virtual Sensitivity+:** Where S bases their belief that P_R on their belief that P_V, if it had not been the case that P_R, then S would not have believed that P_V.

Putting it all together, on the proposal being suggested here, S can come to know that P (is true about the real world) provided P is true, S believes that P and S's belief that P meets virtual sensitivity+.

We can now explain why the individual in the COMPUTER MALFUNCTION case does not have knowledge. Even though their belief about the real world is true and is based on a belief about the virtual world, their belief in the virtual world is not sensitive to the truth of the real world. In this scenario, let us suppose that P is false. In which case the design of the educational VR program would now accurately reflect the truth of the real world. However, once again there is a malfunction and instead of the machine producing a virtual world where P is false, it produces one in which P is true. S comes to form the belief that P based on their experiences. But now we can see what went wrong. Here their belief is not sensitive to the truth of the real world: in the nearest possible world where P is false, S would continue to believe that P is true. The belief formed in COMPUTER MALFUNCTION fails virtual sensitivity+ and therefore does not amount to knowledge.

What about the SPECTRUM INVERSION case? This example suggests that beliefs about a virtual world do not need to be true in order to provide knowledge of the real world. At first glance this looks puzzling and even counterintuitive. But if virtual sensitivity+ provides the right modal relationship between the facts and a person's belief, then we can explain why this is the case. In SPECTRUM INVERSION the individual's belief is connected via a reliable mechanism to the facts of the real world. That mechanism involves a number of steps: (i) the content of the virtual world has been designed intentionally to match the real world, (ii) the virtual world inverts periodically, (iii) S's perceptual and belief-forming mechanism compensates for the inversion stage. The result is that when a belief is generated at the end of this chain, it is sensitive to the facts of the real world.

To demonstrate this last point, let us consider what would happen in the nearby worlds where P is false, i.e. where stop signs are not red. Let us suppose that they are blue. Then the virtual world created would invert colours for 5 minutes every hour. S happens to experience the colour of the virtual stop sign during these 5 minutes, which whilst inverted, is actually red. Because their perceptual and cognitive faculties have compensated for this fact, they form the belief that stop signs are blue in the virtual world, and therefore, that stops signs are blue in the real world. In the nearby world where P is false in the real world, they also come to believe that it is false in the virtual world. Therefore, virtual sensitivity+ is satisfied and this fits with our intuitions that in this case the person would have attained knowledge.

6. Conclusion

I have argued that beliefs can amount to knowledge even if the sole basis for believing them comes from virtual reality. This outcome is the one that is most consistent with the empirical evidence surrounding the role of virtual reality and related technologies in training and education. Accommodating this knowledge within Nozick's truth-tracking framework requires modifying the sensitivity principle to fit the unique stages of belief-formation present when a belief from a virtual world is transferred to the real world. Given that bespoke versions of the sensitivity principle have been shown to be necessary for other methods of belief-formation, this is not a unique problem for virtual reality. Ultimately, philosophers need to pay more attention to the epistemic aspects of virtual reality, which, as the examples highlighted above demonstrate, is likely to play a much greater role in the formation of our beliefs in the future.[43]

[43] Acknowledgements: I am grateful to the participants of the 7thNational Meeting in Analytic Philosophy (Lisbon Portugal, 09/2018) for helpful comments and constructive criticism on an earlier draft of this paper. I also wish to thank the Centre for Ethics, Politics and Society at the University of Minho (Braga Portugal, 09/2018) who invited me to present this paper as part of their guest seminar series.