

**volume V ♦ issue 2**

2014

# Logos & Episteme

an international journal  
of epistemology

**Romanian Academy  
Iasi Branch**



**“Gheorghe Zane” Institute  
for Economic and Social  
Research**

INSTITUTUL EUROPEAN

## Editorial board

---

### Director

Teodor Dima

### Editor-in-Chief

Dan Gabriel Simbotin

### Executive Editors

Eugen Huzum  
Cătălina-Daniela Răducu

### Assistant Executive Editor

Vasile Pleșca

### Editors

Daniela Dunca  
Petru Dunca  
Irina Frasin  
Cornelia Margareta Găspărel  
Bogdan Ștefanachi  
Ioan Alexandru Tofan

---

### Contact address:

Institutul de Cercetări  
Economice și Sociale „Gh.Zane”

Iași, str.T.Codrescu, nr.2, cod 700481  
Tel/Fax: 004 0332 408922  
Email: logosandepisteme@yahoo.com

[www.logos-and-episteme.proiectsbcr.ro](http://www.logos-and-episteme.proiectsbcr.ro)

This journal was edited within “Innovation and Development in the Patterning and Representation of Knowledge Through PhD and Post-PhD Fellowships” Project, supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU 159/1.5/S/133675.

The content of these materials does not necessarily represent the official position of the European Union and the Romanian Government.

For detailed information about the other programs co-financed by the European Union, we invite you to visit [www.fonduri-ue.ro](http://www.fonduri-ue.ro)

ISSN 2069-053

## Advisory board

---

Fred Adams  
University of Delaware, USA

Daniel Andler  
Université Paris-Sorbonne, Paris IV, France

Alexandru Boboc  
Academia Română, Romania

Vihren Bouzov  
"St. Cyril and St. Methodius" University of Veliko Turnovo, Bulgaria

Panayot Butchvarov  
University of Iowa, USA

Mircea Dumitru  
Universitatea din București, Romania

Sanford Goldberg  
Northwestern University, Evanston, USA

Alvin I. Goldman  
Rutgers, The State University of New Jersey, USA

Susan Haack  
University of Miami, USA

Stephen Hetherington  
The University of New South Wales, Sydney, Australia

Paul Humphreys  
University of Virginia, USA

Thierry Martin  
Université de Franche-Comté, Besançon, France

Jürgen Mittelstraß  
Universität Konstanz, Germany

Christian Möckel  
Humboldt-Universität zu Berlin, Germany

Maryvonne Perrot  
Université de Bourgogne, Dijon, France

Olga Maria Pombo-Martins  
Universidade de Lisboa, Portugal

Nicholas Rescher  
University of Pittsburgh, USA

Rahman Shahid  
Université Lille 3, France

Alexandru Surdu  
Academia Română, Romania

John Symons  
University of Texas at El Paso, USA

---

Logos &  
Episteme

## TABLE OF CONTENTS

### Articles

Rodrigo BORGES, How to Moore a Gettier: Notes on the Dark Side of Knowledge.....	133
Thomas HALL, In Defense of the Compossibility of Presentism and Time Travel.....	141
Ned MARKOSIAN, Do You Know That You Are Not a Brain in a Vat?.....	161
Rinat M. NUGAYEV, Maxwellian Scientific Revolution: A Case Study in Kantian Epistemology.....	183
Shane RYAN, A Humean Account of Testimonial Justification.....	209

### Debate

Michael HUEMER, Alternative Self-Defeat Arguments: A Reply to Mizrahi.....	223
Notes on the Contributors.....	231
<i>Logos and Episteme</i> . Aim and Scope.....	233
Notes to Contributors.....	235



## ARTICLES



# HOW TO MOORE A GETTIER: NOTES ON THE DARK SIDE OF KNOWLEDGE

Rodrigo BORGES

**ABSTRACT:** The Gettier Problem and Moore's Paradox are related in a way that is unappreciated by philosophers. If one is in a Gettier situation, then one is also in a Moorean situation. The fact that S is in a Gettier situation (the fact that S is "Gettiered"), like the fact that S is in a Moorean situation (the fact that S is "Moored"), cannot (in the logical sense of "cannot") be known by S while S is in that situation. The paper starts the job of mapping what can be said about this feature of Gettier situations. The goal is to stimulate further exploration into this yet uncharted territory.

**KEYWORDS:** Gettier problem, Moore's paradox, luminosity, blindspot

The state of being Moored ( $p$  is true, but  $S$  does not know that  $p$  is true) and the state of being Gettiered ( $S$  has a justified true belief that  $p$ , but  $S$  fails to know that  $p$ ) are importantly connected. If one is in the first state, then one is also in the second state. Contemporary philosophy has failed to acknowledge this important fact. In this paper I begin exploring some of the ways in which being Moored and being Gettiered are related. Most prominently, it is logically impossible for one to know that one is in either of those states, while one is in those states. As with all first explorations into uncharted territory, the task of these notes is to partially map conceptual space and not to arrive at any particular conclusion. Hopefully, what I say here will be enough to stimulate others to look further into these issues.

## 1. Being Moored

It is possible for it to be raining and for me not to know that it is raining. That happens whenever it is raining while I am sleeping, or whenever it rains and I am working in my windowless office. Let such facts be called "Moorean facts," after G.E Moore, and the state the relevant person is in one of being "Moored."<sup>1</sup> Now, it can be easily shown that it is logically impossible for  $S$  to know she is Moored while she is Moored:

1.  $K_s(p \& \neg K_s p)$  [assume for a *reductio ad absurdum*]
2.  $K_s p \& K_s \neg K_s p$  [1, by distribution of knowledge over conjunction]

---

<sup>1</sup> I borrow the label "Moored" from Michael Veber "I Know I Am Not Gettiered," *Analytic Philosophy* 54, 4 (2013): 401–420.

Rodrigo Borges

3.  $K_s p$  [2, by conjunction elimination]
4.  $K_s \neg K_s p$  [2, by conjunction elimination]
5.  $\neg K_s p$  [4, by the factivity of knowledge]
6.  $\neg K_s(p \& \neg K_s p)$  [1, 2-5 by *reductio ad absurdum*]

We should understand each step of this argument as being indexed to a certain time  $t$ , for one can clearly know sometime after  $t$  that one is Moored at  $t$ .<sup>2</sup> Moorean facts present a synchronic, rather than a diachronic puzzle. The argument then shows that  $S$  may not know that she is Moored at the time she is Moored. Since it may well be true that  $p$  and that  $S$  fails to know that  $p$ , being Moored is not an impossible state like “ $S$  is wet and  $S$  is not wet” is at any given time  $t$ . What is more, someone other than  $S$  may come to know, at  $t$ , that  $S$  is Moored at  $t$ . So, Moorean facts not only pose a synchronic puzzle, but they pose a synchronic puzzle to a particular subject, namely, the person who is Moored.<sup>3</sup>

Thus, Moorean facts have the weird characteristic of being out of the cognitive reach of the individual who is a part of that fact, while, at the same time, being within the cognitive reach of others.<sup>4</sup>

Things get worse, however. A different Moorean fact holds whenever  $p$  is the case, but one fails to believe that  $p$  is the case. Even though  $p$  is true and  $S$  does not believe that  $p$  is true,  $S$  cannot know she is in that situation. The argument is similar to the one about knowledge:

1.  $K_s(p \& \neg B_s p)$  [assume for a *reductio ad absurdum*]
2.  $K_s p \& K_s \neg B_s p$  [1, by distribution of knowledge over conjunction]
3.  $K_s p$  [2, by conjunction elimination]
4.  $K_s \neg B_s p$  [2, by conjunction elimination]
5.  $\neg B_s p$  [4, by the factivity of knowledge]
6.  $B_s p$  [3, knowledge entails belief]
7.  $\neg K_s(p \& \neg K_s p)$  [1, 2-5 by *reductio ad absurdum*]

---

<sup>2</sup> G.E. Moore, “Moore’s Paradox,” in *G. E. Moore: Selected Writings*, ed. Thomas Baldwin (New York: Routledge, 1993), 207-212.

<sup>3</sup> This is not meant to exclude the possibility that groups may be Moored. Roy Sorensen argues that groups of people are sometimes Moored. In the case of groups, the “individual” is the relevant group of individuals. See Roy Sorensen, *Blindspots* (Oxford: Oxford University Press, 1998).

<sup>4</sup> From now on I will stop emphasizing that Moorean facts present a synchronic problem. Unless, I state otherwise, this feature of Moorean facts will be presupposed.



Both arguments about our epistemic limitation with respect to Moorean facts seem to be sound and fully general (“p” can refer to any proposition whatsoever). What is more, S can reach the conclusion that S cannot know S is Moored a priori, independently of experience. Someone different from S can also know a priori that S cannot know S is Moored. This shows that “S does not know she is Moored” is a contingent truth we know a priori to be true.

This situation raises obvious problems about thought. The arguments above show that S’s belief that S is Moored can never reach the status of knowledge. If a successful belief is a knowledgeable one, and knowing that one cannot successfully  $\varphi$  gives one sufficient reason not to  $\varphi$ , then the arguments above constitute an a priori reason for one not to believe one is Moored. There is a normative reason for one not to believe that one is Moored. Since believing that p is a way of thinking that p, thinking that one is Moored is always an inappropriate way of thinking about oneself. Thinking that you are Moored has a self-defeating feature to it, similar (but not identical) to believing that you have no beliefs.

Moorean facts also raise problems in speech. G.E Moore argued that it is absurd for S to assert “It is raining, but I do not know that it is” or “It is raining, but I do not believe that it is.”<sup>5</sup> If proper assertion requires that one knows what one asserts, then the fact that we do not know Moorean facts explains why asserting propositions expressing those facts is inappropriate: hearers expect one to know the propositions that speakers assert, and they know the speaker does not know the Moorean fact.<sup>6</sup>

## 2. Being Gettiered

Edmund Gettiered famously showed that one might have a justified true belief that p and yet fail to know that p.<sup>7</sup> In one of the cases Gettier discussed, Smith infers a justified true belief that the man who will get the job has ten coins in his pocket from the false but justified belief that Jones will get the job and Jones has ten coins in his pocket. Smith is Gettiered. Like being Moored, one cannot know one is Gettiered. Here is the proof:

---

<sup>5</sup> Moore, “Moore’s Paradox.”

<sup>6</sup> cf. Peter Unger, *Ignorance: A Case for Scepticism* (Oxford: Oxford University Press, 1975); Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000); Keith DeRose, *The Case for Contextualism: Volume One* (Oxford: Oxford University Press, 2009); Matthew Benton, “Two More for the Knowledge Account of Assertion,” *Analysis* 71, 4 (2011): 684–687; John Turri, “The Express Knowledge Account of Assertion,” *Australasian Journal of Philosophy* 89, 11 (2011): 37–45.

<sup>7</sup> Edmund Gettier, “Is Justified True Belief Knowledge?” *Analysis* 23 (1963): 121–123.

1.  $K_s(J_s p \& p \& B_s p \& \neg K_s p)$  [assume for a *reductio ad absurdum*]
2.  $K_s J_s p \& K_s p \& K_s B_s p \& K_s \neg K_s p$  [1, by distribution of knowledge over conjunction]
3.  $K_s \neg K_s p$  [2, by conjunction elimination]
4.  $\neg K_s p$  [3, by the factivity of knowledge]
5.  $K_s p$  [2, by conjunction elimination]
6.  $\neg K_s (J_s p \& p \& B_s p \& \neg K_s p)$  [7,8-11 by *reductio ad absurdum*]<sup>8</sup>

As with being Moored, being Gettiered poses a synchronic, rather than a diachronic, epistemic problem. S cannot know she is Gettiered while she is Gettiered. After Smith comes to know he will get the job and that he has ten coins in his pocket, he can then come to know he had a justified true belief that the man who would get the job has ten coins in his pocket, but that he did not know that. Like being Moored, and unlike being wet and not wet, the state of being Gettiered is not an impossible one. What is more, “Gettierian facts” pose a synchronic puzzle to the individual who is Gettiered, not to others; there is no obstacle preventing epistemologists from knowing Smith is Gettiered.<sup>9</sup>

As with being Moored, S can reach the conclusion that S cannot know S is Gettiered a priori, independently of experience. Epistemologists also know a priori that Smith cannot know he is Gettiered. This shows that “S does not know she is Gettiered” is a contingent truth we know a priori to be true.

Being Gettiered similarly raises problems about thought. One cannot know one is Gettiered. If a successful belief is a knowledgeable one, and knowing that one cannot successfully  $\varphi$  gives one sufficient reason not to  $\varphi$ , then the arguments

---

<sup>8</sup> This argument assumes that, if one is Gettiered with respect to p, then one does not know that p. In Stephen Hetherington, *How To Know: A Practicalist Conception of Knowledge* (Malden: Wiley-Blackwell, 2011) it is argued that one can know that p even if one is Gettiered with respect to p. For the purpose of this paper, I will assume that Hetherington is wrong about this. I discuss and reject Hetherington’s radical contention in Rodrigo Borges, “Knowledge from Knowledge: an Essay on Inferential Knowledge,” PhD thesis, Rutgers University, 2014.

<sup>9</sup> If it is possible for groups to know things, then, as far as I can see, nothing prevents those individuals from being Gettiered. Perhaps the following is a case in which a group is Gettiered: group A has a justified true belief that the group that won the public bidding has at least ten years of experience because they received a memo from the government asserting that organization B has ten years of experience and organization B won the bidding. However, the person typing the memo made a mistake and organization A won the bidding and A also has more than ten years of experience. Group A has a justified true belief that the group that won the bidding has at least ten years of experience, but group A does not know that, for its justified true belief depends essentially on a falsehood.

above constitutes an a priori reason for one not to believe one is Gettiered. There is, therefore, a normative reason for one not to believe that one is Gettiered. Since believing that  $p$  is a way of thinking that  $p$ , thinking that one is Gettiered is *prima facie* inappropriate.

If knowledge is the norm of proper assertion, then one should not assert one is Gettiered; it is clearly inappropriate for Smith to assert (before he learns he will get the job), “I am justified in believing the man who will get the job has ten coins in his pocket, it is true that he has, but I do not know that he does.”

There might be one subtle difference between being Moored and being Gettiered. While believing one is Moored does not prevent one from being Moored (unless believing one does not know that  $p$  automatically destroys one’s belief that  $p$ ), one might think that believing one is Gettiered prevents one from being Gettiered because believing one is Gettiered with respect to  $p$  destroys one’s belief that one is justified in believing that  $p$ . If one believes that one is Gettiered, one acquires a defeater for the claim that one is justified in believing that  $p$ , thereby falsifying the first conjunct of “ $J_p \& p \& B_p \& \neg K_p$ .” The result is that one acquires a false belief whenever one believes one is Gettiered. This result, if correct, is compatible with my claim that one cannot know that one is Gettiered, while one is Gettiered. Perhaps this provides an explanation of why one cannot know one is Gettiered.<sup>10</sup>

### 3. The Dark Side of Knowledge

That Moorean and Gettierian facts share all those features is explained by the fact that, if  $S$  is Gettiered, then  $S$  is also Moored:

$$J_p \& p \& B_p \& \neg K_p$$

entails

$$p \& \neg K_p$$

The converse, of course, is not the case. That is, one may be Moored but not Gettiered with respect to  $p$ . This fact rarely stressed by philosophers.<sup>11</sup> Since being Gettiered entails being Moored, whatever “badness” we assign to the state of being Moored is also present in the state of being Gettiered. Gettiered agents are in a much worse epistemic situation than philosophers have so far realized, for Smith

<sup>10</sup> Thanks to Georgi Gardiner for discussion here.

<sup>11</sup> Michael Veber is the only exception I know of on this subject (see Veber, “I Know I Am Not Gettiered”). However, he mentions this point in passing, without stopping to discuss it. Here I take Veber’s remark further.

not only fails to know the truth he justifiably believes, but it is logically impossible for him to know he is in that situation, and, if proper assertion requires knowledge, he may not assert that she is Gettiered either. There are a few things we can say about this bad epistemic place in which Gettiered agents find themselves. The following considerations will also apply to the badness of being Moored, since one cannot know one is Moored either.

Timothy Williamson has argued that no non-trivial condition C is “luminous.” A condition C is luminous if it satisfies the following definition:

- (L) For every case  $\alpha$ , if in  $\alpha$  C obtains, then in  $\alpha$  one is in a position to know that C obtains

Williamson argues that no non-trivial condition satisfies (L).<sup>12</sup> It can be shown that, given (L), being Moored and being Gettiered are non-trivial, non-luminous conditions. They are non-trivial conditions because neither of them holds necessarily or necessarily fails to hold.<sup>13</sup> The condition of being taller than oneself satisfies (L) vacuously and is thus trivially luminous, while the condition of being identical to oneself is trivially luminous because it obtains in every case  $\alpha$ .

To see that being Moored or being Gettiered are not luminous conditions, let us assume, with Williamson, that, if one is in a position to know that C obtains, and one has done what one is in a position to do to decide whether C obtains, then one knows that C obtains.<sup>14</sup> The arguments in the two previous sections show that there is no case  $\alpha$  such that S is Gettiered or Moored in  $\alpha$  and S is in a position to know she is Gettiered or Moored in  $\alpha$ , for there is nothing one can do in  $\alpha$  that will give one knowledge that one is Moored or Gettiered. Thus, being Moored or being Gettiered are not luminous conditions, according to Williamson, for neither condition satisfies the definition of a luminous condition, (L).

In fact, we can see that something stronger is true of the conditions of being Moored and being Gettiered, for they are not only non-trivial, non-luminous conditions, but they are, what we may call “non-trivial dark conditions.” A condition C is non-trivially dark if it satisfies the following definition:

- (D) For every case  $\alpha$ , if in  $\alpha$  C obtains, then in  $\alpha$  one is not in a position to know that C obtains

(D) is not the mere denial of (L). The fact that C is not luminous is necessary, but not sufficient for C to be a dark condition. One is sometimes in a position to know that condition C obtains even if C is not luminous (that C is not

---

<sup>12</sup> Williamson, *Knowledge and Its Limits*, 97–102.

<sup>13</sup> Williamson, *Knowledge and Its Limits*, 107–8.

<sup>14</sup> Williamson, *Knowledge and Its Limits*, 95.

luminous entails only that one is not always in a position to know that C obtains). If C is a dark condition, however, then there is no case  $\alpha$  such that one is in a position to know, in  $\alpha$ , that C obtains.

Dark conditions are not trivial conditions. The conditions of being Moored and of being Gettiered obtain in some cases but not in others, but they are always out of our cognitive reach. The arguments in the previous sections establish that being Gettiered and being Moored are dark conditions in the sense of (D). Williamson does not discuss dark conditions as I have defined them, but, given the arguments above, a full story about luminosity ought to include a discussion of conditions that, like being Gettiered and being Moored, obtain in some but not all cases and with respect to which one is never in a position to know they obtain (at the time in which they obtain). In short, a discussion of luminosity is not complete without a discussion of dark conditions.

The argument in the previous section shows that being Gettiered is a cognitive “blindspot,” in the sense introduced by Roy Sorensen.<sup>15</sup> “Blindspot” is a technical term for Sorensen, to be distinguished from “blind spot.” The latter term is used in physiology to refer to the area in one’s retina devoid of rods and cones, and, therefore, insensitive to light. The former term refers to a relational property whose relata are a person S, a fact p, and a time t. Sorensen defines a blindspot thus:<sup>16</sup>

(B) the fact that p is a blindspot for an agent S at a time t if and only if S is not in a position to know that p at t.

Notice that blindspots are dark conditions in the way defined by (D). The converse is also true: if the fact that p is a blindspot for S at t, then the condition, C, p is about is a dark condition. It is only fitting that dark conditions constitute epistemic blindspots. Even though Sorensen recognized that Moorean facts are blindspots, he failed to see that Gettierian facts are also Moorean facts. This paper contributes to the discussion of blindspots by showing that Gettierian facts are blindspots.

#### 4. Conclusion

This ends my brief exploratory notes on the striking similarities between the states of being Moored and of being Gettiered. Being Gettiered entails that one is Moored, and it is logically impossible for one to know that one is in either of those

---

<sup>15</sup> Sorensen, *Blindspots*.

<sup>16</sup> Sorensen, *Blindspots*, 1-13.

Rodrigo Borges

states while one is in them. Those states are dark conditions, synchronically always in our cognitive blind spot.<sup>17</sup>

---

<sup>17</sup> Thanks to Michael Veber for discussion on some of the issues in this paper and for the inspiration to write it. Thanks also to Georgi Gardiner for fruitful discussion on an earlier draft of this paper. Special thanks to Peter Klein for discussing with me different drafts of this paper. The research for this paper was partially funded by the Capes/Fulbright Commission. I am grateful for their support.

# IN DEFENSE OF THE COMPOSSIBILITY OF PRESENTISM AND TIME TRAVEL

Thomas Hall

ABSTRACT: In this paper I defend the compossibility of presentism and time travel from two objections. One objection is that the presentist's model of time leaves nowhere to travel to; the second objection attempts to equate presentist time travel with suicide. After targeting some misplaced scrutiny of the first objection, I show that presentists have the resources to account for the facts that make for time travel on the traditional Lewisian view. In light of this ability, I argue that both of the objections fail.

KEYWORDS: presentism, time travel, nowhere argument, suicide machine argument

## 1. Introduction

Presentism is a thesis about temporal ontology according to which it is necessarily the case that only present entities exist.<sup>1</sup> Many philosophers have said that presentism precludes the possibility of time travel. For example, William Grey has said that the possibility of time travel would have to “presuppose that the past or future were somehow real,”<sup>2</sup> and Steven Hales has said that “there is no such thing as time travel under presentism.”<sup>3</sup> In addition, several philosophers have endorsed the view that time travel is possible only if eternalism (sometimes called ‘four-dimensionalism’) is true – the view that past, present, and future entities all exist on an ontological par. William Godfrey-Smith, for example, has said that “the metaphysical picture which underlies time travel talk is that of the block

---

<sup>1</sup> There is some variation in how presentism is formulated. Some authors add that *it is always the case that* everything is present (e.g. see Thomas M. Crisp, “Presentism,” in *The Oxford Handbook of Metaphysics*, eds. Michael J. Loux and Dean W. Zimmerman (Oxford: Oxford University Press, 2003), 215), and others add that *the only properties and relations that present entities instantiate are those they currently instantiate* (e.g. see Theodore Sider, “Traveling in A- and B- Time,” *The Monist* 88, 3 (2005): 329). Also, some authors do not (explicitly) formulate presentism as a metaphysically necessary thesis. Nothing in this paper will turn on such variation. For defenses of presentism, see Crisp, “Presentism;” Ned Markosian, “A Defense of Presentism,” in *Oxford Studies in Metaphysics, Volume I*, ed. Dean W. Zimmerman (Oxford: Oxford University Press, 2003); and Craig Bourne, *A Future for Presentism* (Oxford: Oxford University Press, 2006).

<sup>2</sup> William Grey, “Troubles with Time Travel,” *Philosophy* 74, 1 (1999): 56.

<sup>3</sup> Steven Hales, “No Time Travel for Presentists,” *Logos & Episteme* 1, 2 (2010): 360.

universe,”<sup>4</sup> and Ken Perszyk and Nicholas Smith agree that “[f]or time travel to be possible, we need a different conception of time: four-dimensionalism.”<sup>5</sup> Although there have been a few dissenters,<sup>6</sup> this appears to be the common view.

The aim of this paper is to join the dissenters and show that the common view isn’t any good. Towards this end, I will evaluate two arguments that purport to show that presentism rules out the possibility of time travel: the Nowhere Argument (§2) and the Suicide Machine Argument (§3). After exposing some misplaced scrutiny of the first argument, I will ultimately respond to the arguments on behalf of the presentist.<sup>7</sup>

## 2. The Nowhere Argument

Simon Keller and Michael Nelson dub the following argument the ‘Nowhere Argument:’

On the presentist model, the past and the future do not exist, so there is nowhere for the time traveller to go. Travelling to Portland is possible, because Portland is right there waiting for you. But travelling to the Land of Oz is impossible, because there is no such place. Travelling to the past or future is more like travelling to the Land of Oz, if presentism is true. You can’t travel to somewhere that doesn’t exist, so, if presentism is true, you can’t travel to other points in time.<sup>8</sup>

---

<sup>4</sup> William Godfrey-Smith, “Travelling in Time,” *Analysis* 40, 2 (1980): 72.

<sup>5</sup> Ken Perszyk and Nicholas Smith, “The Paradoxes of Time Travel,” in *Maui and the White Rabbit: Maori and Pakeha Concepts of Time*, ed. Hamish Campbell (Te Papa: National Museum of New Zealand Press, 2001), 4.

<sup>6</sup> See Simon Keller and Michael Nelson, “Presentists Should Believe in Time Travel,” *Australian Journal of Philosophy* 79, 3 (2001): 333-45; Paul Daniels, “Back to the Present: Defending Presentist Time Travel,” *Disputatio* 4, 33 (2012): 469-84; and Ryan Wasserman, *The Paradoxes of Time Travel* (manuscript).

<sup>7</sup> One noteworthy restriction. I will only be concerned with *closed-future* presentism – that is, presentism plus the view that future-tensed contingent statements are either determinately true or determinately false. Kristie Miller has argued that time travel is incompatible with *open-future* presentism – that is, presentism plus the view that future-tensed contingent statements are neither determinately true nor determinately false. See Kristie Miller, “Time Travel and the Open Future,” *Disputatio* 1, 19 (2005): 223-32; and Kristie Miller, “Backwards Causation, Time, and the Open Future,” *Metaphysica* 9, 2 (2008): 173-91. For a response to Miller, see Daniels, “Back to the Present.”

<sup>8</sup> Keller and Nelson, “Presentists Should,” 334-5.



Initially the argument is quite compelling.<sup>9</sup> However, we should take a closer look:

THE NOWHERE ARGUMENT

(P1) If presentism is true, then necessarily, the past and future do not exist.

(P2) It is impossible to travel somewhere that does not exist.

(C1) If presentism is true, then it is impossible to travel to the past or future.

(P3) If time travel is possible, then it is possible to travel to the past or future.

(C2) If presentism is true, then time travel is impossible.

The first premise goes analytic if by ‘the past and future’ we have in mind concrete temporal locations. Although presentists may believe that all *abstract* times – plausibly thought of as maximal consistent propositions – exist in the present, they will claim that only one of these propositions is true, and moreover, it is the one and only *concrete* time that makes it so.<sup>10</sup> And it is surely the concrete conception of times at work in the Nowhere Argument, for I doubt it makes much sense at all to speak of traveling to a proposition. So only (P2) and (P3) are up for grabs. The rationale for (P2) can be put as follows: *traveling* is a two-place relation, one that holds between a traveler and a destination. Given the general principle that no relation can hold without coexisting relata, it follows that no one can travel to a nonexistent destination. The rationale for (P3) is presumably intuitive: in order to travel in time, one must travel to the past or future. Given these rationales, the argument naturally progresses as follows. In cases of *time* travel, the destination relatum would be a concrete time.<sup>11</sup> But, if presentism is true, there are no nonpresent concrete times. As such, presentism entails that traveling to the past or future is impossible because (i) no past or future concrete times exist according to presentism, and (ii) the *traveling* relation cannot hold with respect to the nonexistent.<sup>12</sup> From here, the second conclusion comes easy:

---

<sup>9</sup> The Nowhere Argument goes back (at least) to Grey, “Troubles,” 56-7, and is endorsed (among others) by Grey, “Troubles;” Perszyk and Smith, “The Paradoxes;” and Peter Eldridge-Smith, “Paradoxes and Hypodoxes of Time Travel,” in *Art and Time*, eds. Jan Lloyd Jones, Paul Campbell, and Peter Wylie (Melbourne: Australian Scholarly Publishing, 2007).

<sup>10</sup> On the abstract-/concrete-time distinction see Markosian, “A Defense,” 32-3.

<sup>11</sup> It is popular to identify concrete times with three-dimensional hyperplanes of spacetime or maximal mereological sums of contemporaneous concreta. For my purposes, however, it won’t be necessary to endorse any particular account.

<sup>12</sup> As I have described it, the Nowhere Argument is an instance of the familiar problem of cross-time relations. There is a considerable amount literature on this topic. For some useful discussion, see Roderick Chisholm, “Referring to Things That No Longer Exist,” *Philosophical*

since time travel surely requires traveling to the past or future, and presentists cannot allow for such travel, presentism entails the impossibility of time travel.

### 2.1. Against Objections to the Second Premise

Unfortunately, many philosophers have focused their scrutiny on (P2) of the Nowhere Argument. In this section, I will consider – and reject – two popular ways of objecting to this premise. The more common way of fleshing out this objection is with a simple reductio: if (P2) is true, then the argument rules out too much: the ordinary passage of time.<sup>13</sup> This is supposedly because the passage of time involves persisting objects traveling to the immediate future, which (according to presentists) is just as nonexistent as the distant past. This is a bad objection. Presentists will not equate ordinary persistence with *traveling* to the immediate future. To see why, we need to understand what presentists say the passage of time consists in.

Presentism is a version of the A-theory of time – the view that tensed statements are irreducible, typically in the sense that tensed sentence tokens cannot be given tenseless truth conditions.<sup>14</sup> In order to express tensed truths, presentists traditionally utilize primitive tense operators (e.g. WAS, WILL) that attach to present-tensed sentences. For example:

- (1) WAS (There are dinosaurs)

According to A-theorists, tensed statements like (1) change their truth value over time. For example, (1) is true now, as it was a thousand years ago. But 1 billion years ago, (1) was false. This is what the passage of time for presentists (*qua* A-theorists) consists in – the *constant and inexorable* change in the truth values of

---

*Perspectives* 4 (1990): 546-56; Theodore Sider, "Presentism and Ontological Commitment," *Journal of Philosophy* 96, 7 (1999): 325-47; Markosian, "A Defense;" Thomas Crisp, "Presentism and 'Cross-Time' Relations," *American Philosophical Quarterly* 42, 1 (2005): 5-17; and Bourne, *A Future*, 95-108.

<sup>13</sup> For endorsements of this objection, see Phil Dowe, "The Case for Time Travel," *Philosophy* 75, 3 (2000): 443; Keller and Nelson, "Presentists Should," 335; Sider, "Traveling," 329; and Daniels, "Back to the Present," 472.

<sup>14</sup> In particular, A-theorists resist the reduction of tense to times. Those that endorse this reduction (i.e. B-theorists) would say that an utterance of 'there were dinosaurs' expresses a truth if and only if there exists some time *t* such that (i) there are dinosaurs at *t*, and (ii) *t* is earlier than the time of utterance. For more on this detensing strategy, see Theodore Sider, *Four Dimensionalism: An Ontology of Persistence and Time* (Oxford: Oxford University Press, 2001), 11-25.

tensed propositions.<sup>15</sup> This makes the phenomenon of ordinary persistence quite different from *traveling* to the immediate future. After all, traveling essentially consists in a change of relations to external objects, but “keeping flow” with the passage of time on this conception does not, for the continuation of the successive process by which the truth values of tensed propositions change is not dependent upon any objectual change in external relations.<sup>16</sup> On presentism, therefore, ordinary persistence does not amount to anything like traveling. (P2) thereby poses no threat to the passage of time, and so this first objection fails.

Other philosophers have rejected (P2) on different grounds; they claim that one can travel to a nonexistent place *as long as it exists upon arrival*.<sup>17</sup> Here is one example involving spatial travel to a nonexistent location, due to John W. Carroll:

Suppose you are a big fan of amusement parks and hear that they are planning to build a new one in Argentina. It doesn't exist yet, but you are so excited that you start now to hitchhike your way there from Raleigh, NC. It seems that you are traveling to the amusement park even though it doesn't exist. What seems important is not that the destination exists when you start to travel, but that it exists when you arrive.<sup>18</sup>

The case is unpersuasive. A more accurate description would go like this: you start to travel to some *existing* place – Argentina – which is such that it *will* be the case that when you arrive there, it contains an amusement park. We can preserve the original intuition by saying that an utterance of ‘I'm traveling to the amusement park in Argentina’ expresses something like a “quasi-truth”<sup>19</sup> when the following closely related facts obtain:

- (2) I'm traveling to Argentina, and
- (3) WILL (An amusement park exists in Argentina and I arrive there)

---

<sup>15</sup> Non-presentist A-theorists may wish to insist that the passage of time consists in the process by which objects successively possess different A-properties like *pastness*, *presentness*, and *futurity*. Presentists cannot endorse this account of passage because on their view it is impossible for anything to possess genuine *pastness* or *futurity*.

<sup>16</sup> This is related to the debate over whether time could pass without change. I take it that presentists are firmly in the camp that says “yes.” For a representative defense of the possibility of temporal passage without change, see Sidney Shoemaker, “Time Without Change,” *Journal of Philosophy* 66, 12 (1969): 363-81.

<sup>17</sup> For endorsements of this objection, see Dowe, “The Case,” 443; Miller, “Time Travel,” 226.

<sup>18</sup> “The Nowhere Argument,” A Time Travel Website, accessed December 30, 2013, ed. John W. Carroll, <http://timetravelphilosophy.net/topics/nowhere/>.

<sup>19</sup> Roughly, a *quasi-truth* is something appropriate to assent to in everyday circumstances, although not literally true. On the notion of quasi-truth, see Sider “Presentism;” and Markosian, “A Defense,” 24-5.

This response makes the case of spatial travel on offer inapplicable as a counterexample to (P2).<sup>20</sup>

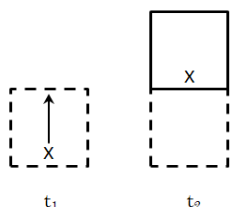
Given that presentists think time is very different from space, they will likely want to resist the likening of time travel to spatial travel. *Contra* spatial travel, time travel simply cannot be a genuinely relational affair for presentists. This is because presentists cannot (or at least should not) believe that it is possible for any genuine cross-temporal relation to hold, for the simple and powerful reason that *no* relation can hold without coexisting relata. As such, I think presentists should welcome the truth of (P2) with open arms. However, now aware of the explanation for this concession, we should be skeptical of the third premise:

- (P3) If time travel is possible, then it is possible to *travel* to the past or future.

In order for the argument to be valid, the sense of ‘travel’ at work here must be the same relational sense that backs (P2). Importantly, this exposes a crucial bit of the rationale for (P3) that was not initially apparent – namely, that *time travel* is genuinely relational. But now we can see that (P3) entails that there must exist a (concrete) nonpresent time in order to time travel, which is tantamount to the position that traveling in time requires that presentism be false! Although *prima facie* this premise appeared conceptually true (probably because ‘time travel’ contains the word ‘travel’), closer inspection has revealed it to be inconspicuously question-begging. As such, if it can be shown that presentists can make metaphysical sense of time travel being non-relational, this will give us a reason to reject (P3). I hope to do just this. My plan is to show that presentists have the

---

<sup>20</sup> It may be possible to repair the spatial-analogy case. For example, imagine a case where some person, x, knows that (say) God is going to create a new closed spatial region. God’s plan is to connect the new region to the finite open space that x inhabits. Given her knowledge, x begins to travel to the nonexistent space at  $t_1$ , arriving when it appears at  $t_2$ :



But is it strictly-speaking true that x is traveling to the nonexistent closed space at  $t_1$ ? I think not. Again, we can cite “in the ballpark” facts that suffice to make it quasi-true, like: (i) x is traveling to the “edge” of the open space at  $t_1$ , and (ii) WILL (there exists a closed space connected to the “edge” of the open space and x arrives there).

resources to account for the facts that make for time travel on the traditional Lewisian view without commitment to any nonpresent entities. I will turn to this task in the next section.

## 2.2. A Presentist-Friendly Endorsement

Keller, Nelson, Paul Daniels, and Ryan Wasserman all agree that presentists can translate time travel talk into appropriately tensed language.<sup>21</sup> In this section, I follow them and show exactly how presentists can account for the facts that make for time travel on the traditional Lewisian view of what time travel consists in. The traditional view is best articulated by David Lewis himself:

What is time travel? Inevitably, it involves a discrepancy between time and time. Any traveler departs and then arrives at his destination; the time elapsed from departure to arrival (positive, or perhaps zero) is the duration of the journey. But if he is a time traveler, the separation in time between departure and arrival does not equal the duration of his journey.<sup>22</sup>

Lewis goes on to make a useful distinction between *external time* and *personal time*. The former is simply time itself, which provides an objective ordering and metric of an object's stages.<sup>23, 24</sup> An object's personal time, by contrast, provides a subjective ordering – the assignment of coordinates to its stages which maintains the regularities and physical processes common to its kind.<sup>25</sup> The motivation for this distinction is easy to see. Before entering her time machine, a pastward time traveler may appropriately utter 'In just a few seconds I will see dinosaurs,' but in this case the traveler is really going millions of years

---

<sup>21</sup> Keller and Nelson, "Presentists Should;" Daniels, "Back to the Present;" Wasserman, *The Paradoxes*.

<sup>22</sup> David Lewis, "The Paradoxes of Time Travel," *American Philosophical Quarterly* 13, 2 (1976): 145.

<sup>23</sup> For the purposes of this paper, I am assuming that time is one-dimensional and continuous. Moreover, I will not consider the possibility of branching timelines or "closed time-like curves." For a useful discussion on how closed time-like curves relate to presentist time travel, see Bradley Monton, "Presentists Can Believe in Closed Timelike Curves," *Analysis* 63, 3 (2003).

<sup>24</sup> I will often use vocabulary associated with perdurantism, the view that material objects persist by virtue of having different temporal parts – or "stages" – at each moment they exist. Endurantists deny the existence of temporal parts of material objects, and will therefore need to understand talk of stages neutrally where needed. One option is to utilize object-time ordered pairs. Another option is to let short-lived events of an object's life take the place of temporal parts. On this latter strategy, see Ned Markosian, "Two Arguments from Sider's *Four Dimensionalism*," *Philosophy and Phenomenological Research* 68, 3 (2004): 675-6.

<sup>25</sup> Lewis only explicitly gives this account for persons, but it easily generalizes. See Lewis, "Paradoxes," 146.

Thomas Hall

into the past, not a few seconds into the future. With the relevant distinction in hand, this utterance becomes less paradoxical: we can say that just a few seconds of the traveler's *personal time* will elapse between her entering the time machine and arriving millions of years in the *external* past. This kind of discrepancy between external time and personal time is what makes for time travel on the traditional Lewisian view:

LEWISIAN TIME TRAVEL (LTT)

Necessarily, something travels in time iff and because there is a discrepancy between its personal time and external time.<sup>26</sup>

Although Lewis's distinction is a good one, his functional account of personal time is inadequate. Wasserman explains:

...consider the case of a single, non-time traveling electron. Suppose that some of its stages are labeled in order (1, 2, 3, etc.) according to external time. And suppose further that every electron remains intrinsically unchanged throughout its entire career. In that case, one can assign coordinates to the stages of our particular electron in many different ways and still preserve the kinds of regularities we ordinarily see in other electrons. For example, one can simply reverse the ordering of all the electron-stages. In that case, there would be discrepancy between the "personal" time of the electron and external time, but that would not make the electron a time traveler.<sup>27</sup>

To solve this problem, Wasserman proposes a revised account according to which "an object's personal time is the assignment of coordinates to its stages that matches the coordinates given by the relevant causal relation"<sup>28</sup> – where 'relevant causal relation' picks out whatever immanent causal relation makes for identity over time. The background assumption for this account is the common belief that identity over time (for material objects) requires the right kind of causal dependence between an object's stages.<sup>29</sup> Importantly, this account appropriately privileges a unique ordering of the electron's stages and thereby avoids the worry above.

---

<sup>26</sup> LTT is endorsed (among others) by Lewis, "Paradoxes;" Dowe, "The Case;" Sider, *Four Dimensionalism*; Sider, "Traveling;" Keller and Nelson, "Presentists Should;" Monton, "Presentists Can;" Alasdair Richmond, "Recent Work on Time Travel," *Philosophical Books* 44, 4 (2003): 297-309; Frank Arntzenius, "Time Travel: Double Your Fun," *Philosophical Compass* 1, 6 (2006): 599-616; Steven Hales, "No Time Travel;" Wasserman, *The Paradoxes*.

<sup>27</sup> Wasserman, *The Paradoxes*, 39.

<sup>28</sup> Wasserman, *The Paradoxes*, 34.

<sup>29</sup> See David Lewis, "Survival and Identity," in *The Identities of Persons*, ed. Amélie Oksenberg Rorty (Berkeley: University of California Press, 1976). Reprinted in his *Philosophical Papers*, Vol. I (Oxford: Oxford University Press, 1983).

## In Defense of the Compossibility of Presentism and Time Travel

By introducing a sentential personal tense operator, as utilized below, we can give a more careful account of personal time that employs Wasserman's revision. First, let's take a look at how eternalists would characterize the account (for simplicity, I've only provided truth conditions for the non-metric future-tensed personal tense operator, and only as it applies to presently existing objects):

### ETERNALIST PERSONAL TIME (EPT)

WILL-BE-FOR-O ( $\Phi$ ) at  $t \leftrightarrow$  (i) O exists and has features  $F_1-F_n$  at  $t$ , (ii) there exists some time  $t^*$  and some object  $x$  such that  $x$  has features  $G_1-G_n$  at  $t^*$  and  $\Phi$  at  $t^*$ , and (iii) O existing and having features  $F_1-F_n$  at  $t$  immanently causes it to be the case (in the relevant way) that  $x$  exists with features  $G_1-G_n$  at  $t^*$ .

For the eternalist, the personal tense operator reduces to temporally-indexed qualitative facts and immanent causal relations involving timelessly existing stages. On this account, to say that *it will be the case for some object O that  $\Phi$*  is (basically) to say that O has a stage causally downstream at some time where  $\Phi$ . To see the account at work, consider:

FIGURE 1

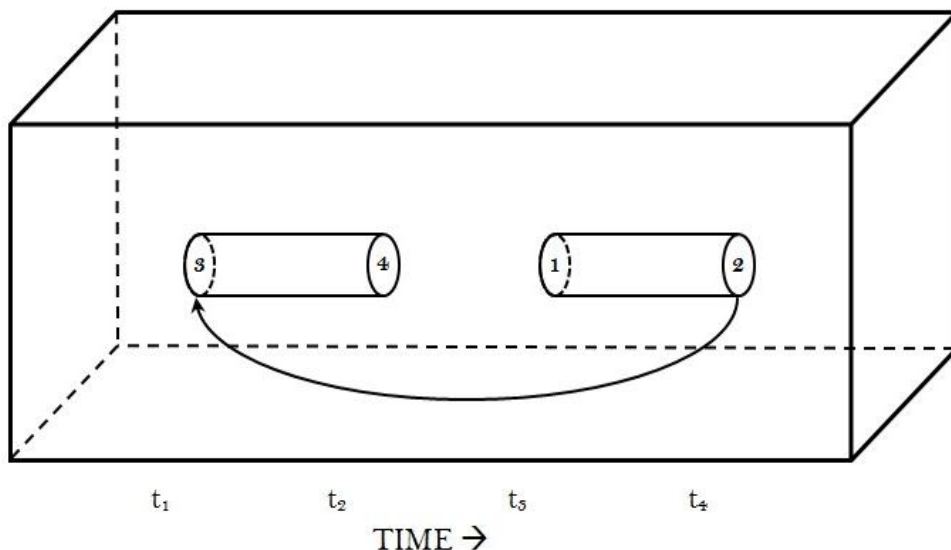


FIGURE 1 depicts a case of discontinuous time travel to the past. The x-axis ( $t_1-t_4$ ) represents external time, whilst the numbered circles (①-④) represent the

ordering of the object's stages according to the relevant immanent causal relation. The object persists normally from  $t_3$  to  $t_4$ , where its departure-stage (i.e. ②) activates a time machine, causing it to be the case (represented by the curved arrow) that its arrival-stage (i.e. ③) appears at  $t_1$ , where it again persists normally until  $t_2$ . There is a discrepancy between the object's relevant causal ordering and its external ordering: ③ is in the *causal future* and the *external past* of ②. By EPT, the object's relevant causal ordering is equivalent to its personal time ordering, thereby allowing us to capture the illustrated LTT-discrepancy: ③ is both a *personal successor* and an *external predecessor* of ②.

Presentists cannot describe the discrepancy in this way. To see why, assume that presentism is true and  $t_4$  is present. In this case, ③ simply doesn't exist, and so it can neither be temporally prior nor a personal successor of anything.<sup>30</sup> But we should not expect presentists to be able to conform to the eternalist's characterization of the discrepancy – they will want to do it in their own terms. The general approach I wish to take towards a presentist-friendly description involves making use of appropriately tensed claims, as first suggested by Keller and Nelson:

But the presentist can have just the same patterns of events happening at just the same times. Or at least, it can be the case on the presentist model that the right sorts of events *will* happen, or *did* happen, or *are* happening, at the right sorts of times.<sup>31</sup>

In agreement here, I take it that if presentism is even to make it to the starting line, it must allow for some way or other of understanding the facts that eternalists utilize to characterize time travel cases – namely, (i) facts about the ordering of an object's stages according to external time, and (ii) facts about how an object's stages relate causally. Fortunately, a presentist-friendly description of the external ordering an object's stages can be found with ease. With respect to FIGURE 1, for example, instead of saying that ③ is an external predecessor of ②, presentists can say the following (where  $t_4$  is the present time):

---

<sup>30</sup> Endurantists will need to understand this claim in their own terms, for on their view the departure and arrival "stages" are really one and the same wholly present object. As I noted in fn.24, there are two promising ways for the endurantist to understand stage-talk. One way is for the endurantist to replace the perdurantist's stages with short-lived *events* of an object's life-event, rather than literal parts of the object. The endurantist could then take the numbered circles to denote the relevant instantaneous events. Alternatively, the endurantist could take the numbered circles to denote ordered pairs of the object and a particular concrete time. Thanks to Hud Hudson for pushing me on this.

<sup>31</sup> Keller and Nelson, "Presentists Should," 338.



- (4)            ② exists and ③ does not, and  
 (5)            WAS (③ exists and ② does not)

An important feature of the presentist's primitive tense operators is that they are *ontologically noncommittal*: the apparent existential commitments of quantifiers and referring terms within their scope are negated.<sup>32</sup> As such, presentists can accept the truth of (5) without being committed to the existence of past arrival-stage ③.

The more difficult task is articulating a presentist-friendly account of personal time that will allow us to capture the relevant causal ordering of the object's stages. Although presentists cannot accept EPT because it quantifies over nonpresent entities, they can easily regiment the account with tense operators to begin to alleviate such commitment. But this won't be enough. In order to complete such an account of personal time, we need a presentist-friendly way of capturing the fact that an object has an appropriate causal successor in the external past or future.

Unfortunately, the metaphysics of causation has been especially troublesome for presentists. We can appreciate the standard worry given a few popular assumptions: (i) the causal relata are physical objects (e.g. events), (ii) causation is not always simultaneous, and (iii) if *c* causes *e*, then both *c* and *e* exist. Now here's the punch: if causation at least sometimes occurs at a temporal distance, and if the relevant physical relata need exist in order for said relation to hold, then presentism is straightforwardly false.<sup>33</sup> One response on behalf of the presentist is to drop the *physical* requirement – perhaps the causal relata are best taken to be abstract entities that exist in the present. Presentists are likely to posit a variety of tensed *facts*, for example, and may happily make use of them as the causal relata.<sup>34</sup> However, there are a few reasons to dislike fact-causation to which

---

<sup>32</sup> More familiar operators share this feature as well. For example, modal primitivists who accept the view that everything is actual traditionally take the modal operator 'POSSIBLY' to work in this way; e.g. the truth of 'POSSIBLY (There are talking donkeys)' brings no commitment to talking donkeys.

<sup>33</sup> An argument along these lines is considered by John Bigelow, "Presentism and Properties," *Philosophical Perspectives* 10 (1996): 35-52; Dean Zimmerman, "Chisholm and the Essences of Events," in *The Philosophy of Roderick M. Chisholm*, ed. Lewis E. Hahn (Chicago: Open Court, 1997); Michael Tooley, *Time, Tense, and Causation* (Oxford: Clarendon Press, 1997); Sider, "Presentism;" Markosian, "A Defense;" Bourne, *A Future*; and Brannon McDaniel, "Presentism and Absence Causation: An Exercise in Mimicry," *Australian Journal of Philosophy* 88, 2 (2009): 323-32.

<sup>34</sup> On fact-causation, see Jonathan Bennett, *Events and Their Names* (Indianapolis: Hackett, 1988), 21-49.

I am largely sympathetic.<sup>35</sup> Nevertheless, for those that do not share such quibbles, everything that follows could be accomplished just as well in a fact-causation framework. Although I can't argue for it here, the option I prefer presentists take is to leave behind the idea that causation is a metaphysically distinguished relation.<sup>36</sup> As such, I think presentists ought to deny that 'c causes e' (where 'c' and 'e' purport to name existing relata) is the correct fundamental locution for causation. Following Sider, presentists will benefit from utilizing a two-place sentential causal operator, providing 'BECAUSE ( $\Phi$ ), WILL<sub>N-UNITS-OF-TIME-HENCE</sub> ( $\Psi$ )' for ordinary (forward-directed) causation, and 'BECAUSE ( $\Phi$ ), WAS<sub>N-UNITS-OF-TIME-AGO</sub> ( $\Psi$ )' for backward causation.<sup>37</sup> What is most important for my purposes is that this approach allows presentists to engage in causal-talk without ontological commitment to nonpresent entities. For example, take:

- (6) Jen's currently activating the time machine causes her appearance in Seattle 30 years ago.

By utilizing the causal operator, (6) can be translated into:

- (7) BECAUSE (Jen activates the time machine), WAS<sub>THIRTY-YEARS-AGO</sub> (Jen appears in Seattle)

Although (6) apparently commits us to the existence of the event *Jen's appearing in Seattle*, (7) allows us to countenance the same causal connection without this commitment.

Certain tangential issues aside, I will adopt this general approach to causation.<sup>38,39</sup> But before we can usefully employ the causal operator in a presentist-friendly account of personal time, we also need to be able to pick out the right kind of causal explanation – namely, the one that makes for identity over time. To explicitly represent the explanation required, I will simply subscript the causal operator with 'ID,' which will do the same stipulatory work as 'in the

---

<sup>35</sup> For one, the fact-causation theorist sacrifices neutrality with respect to the granularity of facts. Moreover, she will have to justify taking the causal relata to be causally-inert entities.

<sup>36</sup> This is where my view differs substantially from that of Keller and Nelson, who agree that "if causation can be a cross-time relation on the four-dimensionalist view, then it can be one on the presentist view too." See Keller and Nelson, "Presentists Should," 341.

<sup>37</sup> See Sider, "Presentism," 338; and Sider, "Traveling," 5.

<sup>38</sup> For some worries involving this approach, see Sider, "Presentism," 339-40.

<sup>39</sup> It is worth noting that adopting the causal operator does not commit one to any substantive thesis about the metaphysical basis for causation. In other words, presentists certainly don't have to take the causal operator as primitive. One plausible reductive view that presentists could opt for, for example, says that the causal operator reduces to the laws of nature plus the instantiation of qualitative properties and single-time relations.

relevant way' in EPT. Presentists can then say 'BECAUSE<sub>ID</sub> (O exists and has such-and-such features), WILL/WAS (there exists a  $y$  with such-and-such features)' to express that some object O has an *appropriate* causal successor in the external future/past. We can now produce a complete account of presentist personal time that substantively mirrors EPT (again, I've only provided truth conditions for the non-metric future-tensed personal tense operator, and only as it applies to presently existing objects):

PRESENTIST PERSONAL TIME (PPT)

WILL-BE-FOR-O ( $\Phi$ )  $\leftrightarrow$  O exists and has features  $F_1-F_n$  and either:

- (i) WILL (There exists some  $y$  such that  $y$  has features  $G_1-G_n$  and  $\Phi$ ) and BECAUSE<sub>ID</sub> (O exists and has features  $F_1-F_n$ ), WILL (There exists some  $y$  such that  $y$  has features  $G_1-G_n$ ), or
- (ii) WAS (There exists some  $y$  such that  $y$  has features  $G_1-G_n$  and  $\Phi$ ) and BECAUSE<sub>ID</sub> (O exists and has features  $F_1-F_n$ ), WAS (There exists some  $y$  such that  $y$  has features  $G_1-G_n$ ).<sup>40</sup>

PPT reveals that by appropriately employing tense and (subscripted) causal operators, we can straightforwardly emulate EPT in a way that is friendly to a presentist's ontology.<sup>41</sup> For a visual representation of the account at work, we can reconstruct (part of) FIGURE 1 into a presentist-friendly model as shown below:

---

<sup>40</sup> Cf. Wasserman, *The Paradoxes*.

<sup>41</sup> It is worth noting that PPT can be used to (accurately) describe the personal futures of ordinary persisting objects. For example, imagine that I plan to order pizza tonight so that it is now true that *it will be for me that I order pizza*. This is true because *it will be the case that* someone appropriately causally related to me orders pizza. More carefully:

WILL-BE-FOR-O (O orders pizza) iff (i) O exists and has features  $F_1-F_n$ , (ii) WILL (There exists some  $y$  such that  $y$  has features  $G_1-G_n$  and O orders pizza) and (iii) BECAUSE<sub>ID</sub> (O exists and has features  $F_1-F_n$ ), WILL (There exists some  $y$  such that  $y$  has features  $G_1-G_n$ ).

As such, PPT seems to successfully undermine a worry articulated by Sider, "Traveling," 333; *my emphasis*:

That I will view a dinosaur in my personal future amounts merely to the fact that I once viewed a dinosaur, and moreover that this is caused by my entry into a time machine. Since this fact *bears little resemblance to the facts that constitute a normal person's genuine future*, I could not enter the time machine with anticipation and excitement at the thought of seeing a dinosaur, for it is not true that I am about to see a dinosaur, nor is the truth much like being about to see a dinosaur.

FIGURE 2

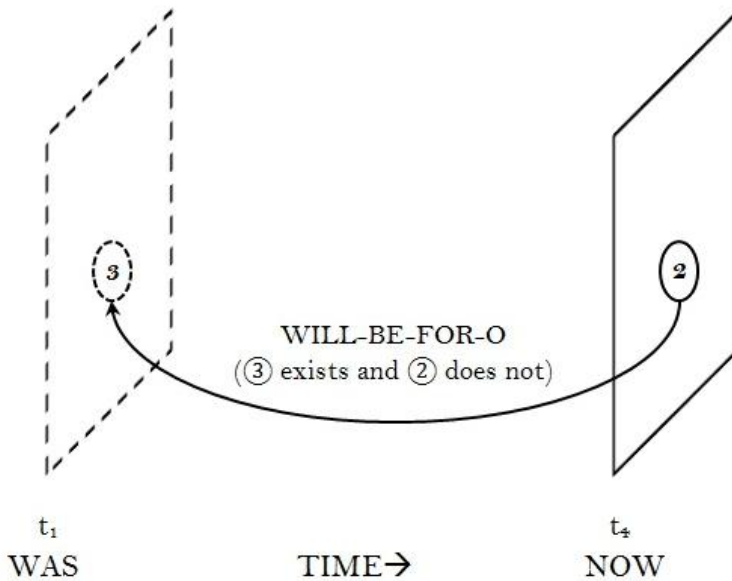


FIGURE 2 depicts “two” instantaneous cross-sections of the eternalist’s block in FIGURE 1.42 The present time is  $t_4$ , and thus according to presentism, ② exists and ③ does not. However, presentists are happy to say that it was the case that ③ exists and ② does not, thereby capturing the relevant external ordering of the object’s stages. Similarly to FIGURE 1, the direction of immanent causation is represented by the curved arrow (although contra FIGURE 1, it is important to remember that it does not signify a genuine causal relation here). Given that ② existing now with various features immanently causally explains (in the relevant way) that it was the case that ③ exists with various features, PPT secures the relevant personal ordering of the object’s stages – that although ② exists and ③ does not, it will be for O that ③ exists and ② does not. Presentists are therefore equally capable of accounting for the illustrated LTT-discrepancy, easily described

<sup>42</sup> Some might object to the possibility of *discontinuous* presentist time travel (as depicted above). One way to do this is to endorse the view that temporal continuity is necessary for identity over time. However, as Keller and Nelson argue, this is a problem equally suited for eternalists and presentists alike. See Keller and Nelson, “Presentists Should,” 339. Another way is to endorse the view that discontinuous *causation* is impossible. But again, it’s not clear how this view could be motivated on purely presentist grounds, and even if it could, the diagrams could be redrawn so as to depict continuous pastward time travel instead.

as a discrepancy between personal tensed facts and ordinary tensed facts: e.g. it both will be for O that and was the case that ③ exists and ② does not.

### 2.3. Against the Third Premise

We are now in a position to forge a strong objection against (P3) of the Nowhere Argument. As I suggested above, I think it should be of no great surprise that presentist-friendly time travel will not involve literally *traveling* to the past or future. As Kristie Miller has said,

in some sense the presentist can never travel to the past... what is the case is that it is now true that some current individual did exist in the past, and that individual's existence in the past is caused by her existence in the present.<sup>43</sup>

Similarly, Wasserman suggests that presentists may say time travel

involves making certain past- or future-tensed statements true by acting on presently existing objects (programming flux capacitors, turning on time machines, etc.).<sup>44</sup>

As I hope to have illuminated above, this kind of approach can be finessed so as to provide a way of countenancing the facts that make for time travel on the standard view without bringing in any commitment to nonpresent entities. This is significant because given the relational sense of 'travel' at work in the Nowhere Argument, (P3) entails that if time travel is possible, then it's possible that a (concrete) non-present time exists. So, this premise can be reasonably rejected on the grounds that (i) LTT is true, and (ii) presentists have the resources to account for the facts that make for time travel according to LTT without commitment to any nonpresent entities. In other words, (P3) is false because presentists can make sense of time travel being a *non-relational* affair. On my view this is accomplished by constructing a presentist-friendly account of personal time through the appropriate employment of tense and causal operators, thereby allowing presentists to characterize LTT-discrepancies as a mismatch between personal tensed facts and ordinary tensed facts. Of course none of this proves that presentist time travel is genuinely metaphysically possible – the point here is simply that the Nowhere Argument does not have the means to rule it out.

---

<sup>43</sup> Miller, "Time Travel," 226.

<sup>44</sup> Wasserman, *The Paradoxes*, 38.

### 3. The Suicide Machine Argument

Steven Hales endorses a different argument for the impossibility of presentism and time travel which he dubs the ‘Suicide Machine Argument.’<sup>45</sup> The basic idea is quite simple: if all of reality is confined to the present, then leaving the present to travel in time ultimately amounts to killing yourself. Hales also says that “the moves that presentists make to get around the Nowhere Argument are not successful to fend off the Suicide Machine Argument.”<sup>46</sup> In this section, I will argue that Hales’ argument fails for much the same reason the Nowhere Argument fails.

First, some thoughts about suicide. Hales says that “[f]or presentists, getting into a time machine is suicide – the occupant goes out of existence.”<sup>47</sup> The principle being relied on here seems to be this:

- (8) For any person,  $x$ ,  $x$  commits suicide by  $\Phi$ ing if  $\Phi$ ing causes  $x$  to go out of existence.

The intuitive force is clear: being taken from the whole of reality is tantamount to annihilation. But before we get on board too quickly, consider the eternalist’s model in which all of an object’s stages exist timelessly within spacetime. On this picture, there is a definite sense (i.e. the wholly unrestricted, atemporal sense) in which it is impossible for an object to go out of existence – but surely, I would think, suicide would not be an impossible feat if eternalism were true. Granted, as long as Hales’ claim is charitably construed merely as a sufficient condition, the fact that it is impossible (on the eternalist’s model) to go out of existence would not entail the impossibility of eternalist suicide. But what alternative principle would the eternalist opt for – that is, what would the eternalist say suicide consists in? Plausibly:

- (9) For any person,  $x$ ,  $x$  commits suicide by  $\Phi$ ing iff  $\Phi$ ing causes  $x$  to have no personal future.

This principle turns out to be good news for the compossibility of presentism and time travel, for the relevant condition is not only sufficient, but

---

<sup>45</sup> Hales, “No Time Travel.” For a response to the argument, see Jimmy Licon, “No Suicide for Presentists: A Response To Hales,” *Logos and Episteme* 2, 3 (2011): 455-64. For a rebuttal, see Steven Hales, “Reply to Licon on Time Travel,” *Logos and Episteme* 2, 4 (2011): 633-36. Also, see replies from Jimmy Licon, “Still No Suicide for Presentists: Why Hales’ Response Fails,” *Logos and Episteme* 3, 1 (2012): 145-51; and Jimmy Licon, “Dissecting the Suicide Machine Argument,” *Logos and Episteme* 4, 3 (2013): 339-52.

<sup>46</sup> Hales, “No Time Travel,” 353.

<sup>47</sup> Hales, “No Time Travel,” 357.

necessary. Analogously to the eternalist's account of personal time, PPT says that the direction of one's personal future is determined by the direction of the relevant immanent causal explanation. As such, as long as this immanent causal explanation can be directed (say) backwards in time, presentism allows time travelers to (i) go out of existence, and yet (ii) have a personal future in the external past.<sup>48</sup> Therefore, by rejecting (8) and instead following eternalists in accepting (9), presentists can sensibly say that time travel is *not* suicide.

Given the worries above, I think the best chance this argument has of succeeding is to formulate it without talk of suicide:

THE SUICIDE MACHINE ARGUMENT

- (P1) If presentism is true, then necessarily, if an object leaves the present, it goes out of existence.
- (P2) Necessarily, if something travels in time, it leaves the present.
- (C1) If presentism is true, then necessarily, if something travels in time, it goes out of existence.
- (P3) If time travel is possible, then it's not the case that necessarily, if something travels in time, it goes out of existence.
- (C2) If presentism is true, then time travel is impossible.

The first premise is uncontroversial: according to presentism *everything* is present, and so if an object leaves the present, it no longer exists. The rationale for (P2) is straightforward as well: there is simply no such thing as a time traveler who never leaves the present.

That leaves (P3). The rationale here is unclear. At one point, Hales says that “[eternalists] insist that any sort of successful travel, spatial or temporal, involves the traveler existing at departure and safely arriving, intact and still in reality, at the arrival.”<sup>49</sup> So perhaps the rationale for (P3) is supposed to be this: in order to successfully travel in time, the traveler must exist upon departure *and* arrival. But notice that this poses no threat to the compossibility of presentism and time travel,

---

<sup>48</sup> Some presentists deny the possibility of backwards causation (e.g. see Bourne, *A Future*, 134), and on such grounds could argue that backwards time travel is impossible. But in order for such an argument to show that there is something about presentism *in and of itself* that rules out the possibility of time travel, the motivation for ruling out backwards causation would need to be presentist-inspired as well. One such motivation is noted by Sider, “Traveling,” fn.5 who suggests that presentists could rule out backwards causation via their acceptance of anti-reductionism about the direction of time, plus the view that causation reduces (in part) to the direction of time.

<sup>49</sup> Hales, “No Time Travel,” 358.

for presentists may assert that *it was the case that* or *will be the case that* a traveler's arrival-stage exists at its respective arrival-time. With respect to FIGURE 2 for example, the depicted traveler's arrival-stage ③ existed at its arrival-time ( $t_1$ ). A more plausible take on Hales' rationale for (P3) requires us to focus on the 'still' in the quote above. The idea here, I presume, is not only that time travelers must exist upon departure and arrival, but that they must exist upon arrival *after* their departure. But 'after' here could mean *personally after* or *externally after*. Let's look at each in turn:

- (R1) Successful time travel requires existing upon arrival personally after departure.
- (R2) Successful time travel requires existing upon arrival externally after departure.

If (R1) is the rationale at work, there is again no threat here. A presentist time traveler's departure-stage only needs to have the relevant immanent causal impact on their arrival-stage in order for PPT to secure the fact that the traveler exists upon arrival personally after their departure. With respect to FIGURE 2 for example, O's departure-stage ② immanently causes it to be the case (in the relevant way) that *it was the case that* O's arrival-stage ③ exists, making it true that *it will be for O that* O exists. And relatedly, this personal tensed fact is true despite the fact that O goes out of existence upon departure. As such, (R1) would not be an effective rationale behind (P3) because the truth of (R1) is compatible with the requirement that time travelers must go out of existence upon departure. To reiterate the point made above, this is because the direction of one's personal future is determined by the direction of immanent causation. As long as a presentist time traveler is able to make the appropriate tensed statements about them true, existing upon arrival *personally after* an existence-ending departure is no problem. And in fact, barring the possibility of branching personal futures, going out of existence upon departure is an unsurprising consequence of securing a personal future beyond the present (if presentism is true). Given that (R1) is both easily embraced by presentists and naturally compatible with the denial of (P3), it fails as a candidate rationale.

What if (R2) is the rationale at work? Before we can evaluate this option, a note of clarification is in order. (R2) should *not* be interpreted as claiming that successful time travel requires the traveler to exist *at some post-departure time*, for this would rule out eternalist pastward time travel just as much as presentist pastward time travel. Rather, (R2) should be interpreted as claiming that successful time travel requires it to be *true at some post-departure time* that the traveler exists *in the wholly unrestricted, atemporal sense of 'exists.'* On this



reading, (R2) does in fact support (P3), for if time travelers must exist externally after their departure, it can't (also) be the case that time travelers must completely go out of existence upon departure. More, the rationale would also constitute a threat to the compossibility of presentism and time travel; presentists cannot allow (say) a pastward time traveler to exist upon arrival externally after their departure, for this would commit them to the existence of the respective *past* arrival-time. Given these features, I take it that (R2) must be the intended rationale behind (P3). But just like the rationale behind the third premise of the Nowhere Argument, (R2) amounts to the question-begging position that time travel requires the existence of a (concrete) nonpresent time!<sup>50</sup> To see this, take a look at FIGURE 2. If it must be true externally after  $t_4$  that O exists upon arrival, then it must also be true externally after  $t_4$  that O's *past* arrival-time,  $t_1$ , exists. (P3) of the Suicide Machine Argument can therefore be rejected on familiar grounds: (i) LTT is true, and (ii) presentists have the resources to account for the facts that make for time travel according to LTT without commitment to any nonpresent entities. As it turns out, the Suicide Machine Argument fails for much the same reason the Nowhere Argument fails.

#### 4. Conclusion

A time traveler in a presentist world cannot *travel* to the past or future, for there are no such places to travel to. More, such a time traveler naturally pays the price of nonexistence in order to secure a personal future at odds with external time. But as long as the requisite tensed statements are made true and such a deviant personal future *is* achieved, the obtaining of the associated LTT-discrepancy will make the traveler just as much a genuine time traveler as one who inhabits the eternalist's block. And as I have argued, neither the Nowhere Argument nor the Suicide Machine Argument have anything to say here.<sup>51</sup>

---

<sup>50</sup> Jimmy Licon similarly accuses the Suicide Machine argument of begging the question. See Licon, "No Suicide."

<sup>51</sup> Special thanks to Ryan Wasserman and Hud Hudson for providing helpful comments on an earlier draft of this paper.



# DO YOU KNOW THAT YOU ARE NOT A BRAIN IN A VAT?

Ned MARKOSIAN

**ABSTRACT:** The topic of this paper is the familiar problem of skepticism about the external world. How can you know that you are not a brain in a vat being fooled by alien scientists? And if you can't know that, how can you know anything about the external world? The paper assumes Evidentialism as a theory about justification, and then argues that you are justified in believing that you are not a brain in a vat, in virtue of the fact that your evidence supports that belief. The paper also considers a number of different objections to this proposal. The upshot is that you do know that you are not a brain in a vat, and that you also know lots of things about the external world.

**KEYWORDS:** skepticism, external world, brain in a vat, evidentialism

## 1. Introduction

The topic of this paper is the familiar problem of skepticism about the external world. How can you know that you are not, for example, a brain in a vat being fooled by alien scientists? And if you can't know that, then how can you know anything about the external world? My aim is to spell out this problem in its strongest, most annoying form, and then to show how it can be solved. The upshot, I will conclude, is that you *do* know that you are not a brain in a vat, and that you do know lots of things about the external world. But let's start with the skeptical problem.

Imagine a human brain in a vat, with tubes and wires hooked up to it. The tubes are for blood flow in and out of the brain, to keep it alive and healthy. The wires are for sensory input, and for monitoring the brain's attempts to move its own body. All of the wires are hooked up to a fancy computer, which ensures that the brain's experiences are completely realistic. Imagine that it seems for all the world to that brain that it is a normal brain in a body, in a world with other humans, and trees, birds, etc.

Now consider this skeptical hypothesis.

**The Brain In a Vat Hypothesis (BIVH):** I am just a brain in a vat being fooled by alien scientists.

It would of course be pretty silly to believe BIVH. After all, you have no special reason to think it is true.<sup>1</sup> But can you know that BIVH is false? Can you know that you are not a brain in a vat? The problem is that all of your experiences are consistent with your being a brain in a vat. Indeed, if you were a brain in a vat, you would be having experiences exactly like the experiences that you are in fact having. So even assuming that you are not a brain in a vat (an assumption that the skeptic is happy to make), it seems like you can't *know* that BIVH is *false*.

Now, unfortunately, pretty much all of your beliefs about the external world are based on the assumption that BIVH is false. (By 'the external world' I just mean the world outside of your own mind.) So, for example, you believe that your best friend is real. But if you are just a brain in a vat, then your best friend is not real. Thus, your belief that your best friend is real is based on the assumption that you are not just a brain in a vat. Similarly with your belief that you have hands, that your bicycle is where you left it, that Barack Obama is President of the US, that grass is green, etc. All of these beliefs are based on the assumption that you are not a brain in a vat. So if you can't know that BIVH is false, then it looks like you can't really know any of your beliefs about the external world.

Here is the argument suggested by these considerations.<sup>2</sup>

**The Brain In a Vat Argument**

- (1) I don't know that BIVH is false.
- (2) If (1), then I don't know anything about the external world.

---

- (3) I don't know anything about the external world.

The problem is that The Brain In a Vat Argument is valid, the premises both seem true, and yet we want to deny the conclusion. Something has to give.

---

<sup>1</sup> What would count as a reason to think BIVH is true? Perhaps experiences like these: You are walking down the street, hearing the ordinary soundtrack of life – birds chirping, the wind in the leaves, etc. Then all of a sudden there is a needle-on-vinyl-like glitch in the soundtrack, immediately followed by a strange, alien voice saying, "How is the experiment proceeding with the human brain in a vat?" This is followed by another strange, alien voice responding, "The experiment is proceeding very well... Hey, don't lean on that button. Get your tentacle off that button!" And this is abruptly followed by a return to the soundtrack of ordinary life.

<sup>2</sup> Hilary Putnam made this kind of argument famous. See Hilary Putnam, *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981), Ch. 1. Putnam has his own strategy for dealing with the argument, which I will not discuss here. (But see Jane McIntyre, "Putnam's Brains," *Analysis* 44 (1984): 59-61, for an excellent summary and criticism of Putnam's strategy.) For yet another anti-skeptical strategy see Jessica Wilson, "The Regress Argument Against Cartesian Skepticism," *Analysis* 72 (2012): 668-673.

## 2. Knowledge and Justification

Before we get to my proposed solution, it will be useful to make explicit some background assumptions that are relevant to our problem. Philosophers generally agree that knowledge involves three things: *belief*, *truth*, and *justification*. Here is an analysis of the concept of knowledge that is often taken as a starting point in epistemology.<sup>3</sup>

**The JTB Analysis of Knowledge:** *S* knows *p* if and only if (i) *S* believes *p*, (ii) *p* is true, and (iii) *S* is justified in believing *p*.

It turns out that there is a well known problem with The JTB Analysis. Edmund Gettier showed in 1963 that belief, truth, and justification are not *sufficient* for knowledge.<sup>4</sup> But The JTB Analysis will do for our purposes, since it is not a controversial matter that *belief*, *truth*, and *justification* are all *necessary* conditions for knowledge, and as we'll see very shortly, that is what is relevant to our skeptical problem. (In other words, we can replace 'if and only if' in The JTB Analysis with 'only if' and we will have something true that is relevant to our skeptical problem.)

Once we have seen that knowledge always involves a justified, true belief, we can see why it appears that I don't know that BIVH is false. It's not that I don't *believe* that the BIVH is false, for I do. And it's not that the BIVH is not *false*, for it is. (Or so, at any rate, we will assume. And remember, the skeptic is happy to grant this assumption. i.e., the fact that BIVH is actually false is not the issue. (Although I will come back to this point later.)) No, the problem is that I have a true belief that does not appear to be *justified*. And the reason my belief that I am *not* a brain in a vat does not appear to be justified is that things seem to me exactly the way they would seem if I *were* just a brain in a vat. I.e., all of my experiences are consistent with BIVH.

Upshot: The reason it seems like I don't know that BIVH is false is that my true belief that it is false appears to be unjustified.

## 3. An Objection to Premise (1) of The Brain in a Vat Argument

I reject premise (1) of the argument. The premise says I don't know that BIVH is false, so to reject the premise is to say that I *do* know that BIVH is false. Basically,

---

<sup>3</sup> Something like The JTB Analysis is often attributed to Plato. See his *Meno* 98. For an alternative approach, see Carrie Ichikawa Jenkins, "Knowledge and Explanation," *Canadian Journal of Philosophy* 36 (2006): 137-164.

<sup>4</sup> See Edmund Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23 (1963): 121-123.

Ned Markosian

then, to reject premise (1) is to say that I know that I am a normal brain in a body. Let's give a name to that proposition.

N = the proposition that I am a normal brain in a body.

As noted above, I believe N. And the skeptic grants that N is true. So here is the situation so far with respect to N.

✓ I believe N.

✓ N is true.

? My belief that N is justified.

Do I know N? Given our assumptions, that depends on whether my belief that N is justified. So to attack premise (1) of The Brain In a Vat Argument, I must argue that my belief that N is justified. (And it is worth emphasizing here that the skeptic agrees with this assessment of the situation. That is, the skeptic agrees with putting those first two check marks there.)

Now, given that what is at issue between the skeptic and me is whether my belief that N is justified, we must consider the question, *What is justification?* I.e., *What conditions must be satisfied in order for it to be true that some person, S, is justified in believing some proposition, p?* Here are some of the main theories that have been proposed by epistemologists in answer to this question.

**Foundationalism:** *S*'s belief that *p* is justified if and only if either (i) *S*'s belief that *p* is indubitable or else (ii) *S* is able to derive *p* from some other, indubitable belief(s).<sup>5</sup>

**Coherentism:** *S*'s belief that *p* is justified if and only if *S* has a coherent system of beliefs and *S*'s belief that *p* fits nicely into that system.<sup>6</sup>

**Reliabilism:** *S*'s belief that *p* is justified if and only if that belief is produced by a reliable belief-forming process.<sup>7</sup>

**Evidentialism:** *S*'s belief that *p* is justified if and only if that belief is supported by *S*'s total evidence.<sup>8</sup>

---

<sup>5</sup> See for example, René Descartes, *Meditations on First Philosophy*. I am using 'indubitable' as a placeholder for whatever characteristic the Foundationalist identifies as the one that makes foundational beliefs justified.

<sup>6</sup> See Erik Olsson, "Coherentist Theories of Epistemic Justification," in *The Stanford Encyclopedia of Philosophy* (2012). <http://plato.stanford.edu/entries/justep-coherence/>.

<sup>7</sup> See Alvin Goldman, "Reliabilism," in *The Stanford Encyclopedia of Philosophy* (2008). <http://plato.stanford.edu/entries/reliabilism/>.

<sup>8</sup> See Earl Conee and Richard Feldman, "Evidentialism," *Philosophical Studies* 48 (1985): 15–34.

## Do You Know That You Are Not a Brain In a Vat?

I vote for Evidentialism, although I won't argue for it here. But I will argue, assuming Evidentialism is true, that my belief that I am not a brain in a vat is justified. If you like, you can take what I say as a big conditional: If Evidentialism is true, then my belief that I am not a brain in a vat is justified. I think that what I will say would also work on the assumption that Coherentism is true, as well as on the assumption that Reliabilism is true, and I will say a little bit about that below. But I will focus mainly on the Evidentialist version of the proposal.

So what does it mean for a belief to be *supported* by your evidence? Like other Evidentialists, I offer no official definition, but here are some examples to help clarify the idea.

- You have read in the newspaper online that the Orioles won the game, and you have also seen reports on TV saying the same thing, and no one has told you otherwise. Your evidence *supports* the belief that the Orioles won the game.
- You feel a tremendous pain in your head. Your evidence *supports* the belief that you are in pain.
- Tim is on his way to ask Maria out again, for the hundredth time, even though she has made it abundantly clear that she has no interest in going out with him. But Tim is optimistic, and believes that this time she will say Yes. Tim's belief that Maria will say Yes is *not supported* by his evidence.
- You have known Tom for 15 years, and he has always been a good, reliable, honest friend. Now he is earnestly telling you that he has no idea what happened to the money you left on the table. So you believe him. But, alas, it just so happens that he is lying to you this time. Still, your evidence that Tom did not take the money *supports* your belief that he did not do it. (This is one of those cases in which your evidence is misleading. Life is hard.)
- You're on the jury in a big murder trial. The first witness is an extremely reliable member of the community, she has nothing to gain from the defendant's being convicted, she says that she saw the defendant commit the murder, and she has a very convincing story. Similarly for a dozen other self-proclaimed eye-witnesses. Meanwhile, the DA presents tons of physical evidence – DNA test results, fingerprints, video footage, etc. – linking the defendant to the crime. Finally, the defendant takes the stand and denies doing it, while telling some ridiculous story. In this case you have *some* evidence that the defendant is innocent (namely, his own testimony), but it is outweighed by the overwhelming evidence against the guy, so that your *total evidence* supports the belief that he did it.

So much for the crucial notion of support. What about evidence? What is that?

I think that each one of us has, at any given time, a great deal of evidence. But I think there are three main kinds. First, and perhaps mostly, there are sensations. These come in the form of visual sensations, auditory sensations, tangible sensations, etc. They are the sights, sounds, smells, tastes, and touches of ordinary (and extraordinary) life. Sometimes they are called appearances. It might also make sense to call them sensory seemings.<sup>9</sup>

Second, there are *a priori* intuitions. It just seems true to me that two plus two equals four, that killing innocent children for no reason is morally wrong, and that a spatially discontinuous object would have to have proper parts. *A priori* intuitions are seemings, too, although of a different kind from sensory seemings.

And third, there are memory impressions. When I introspect and try to examine my own memory impressions, it seems like they come in two main forms. Sometimes they are like vague and fleeting sensations. When I consider the question of where my bicycle is located, a sort of incomplete, flickering image comes into my mind: an image of my bicycle in my garage. But other times memory impressions are more like *a priori* intuitions about certain propositions being true. Right now my memory tells me that I moved to Bellingham in 1998, but it's not the case that I have some vague and fleeting memory impression placing that event in 1998. (I do have fleeting memory impressions of pulling up in a rental truck. But nothing about those impressions places the event in 1998.) Instead, it's more like the relevant proposition is just in the *Remembered Stuff* folder in my mind, and I can tell that by introspection.

Perhaps I have said enough about evidence and support for us to return to our main question: Is my belief that N (the proposition that I am a normal brain in a body) supported by my evidence? I say Yes, and the skeptic says No. Moreover, it is worth emphasizing that this is the very question about which the skeptic and I disagree (on the assumption that Evidentialism is the correct theory of justification, that is).

Here is why I think my belief that N is supported by my evidence. To begin with, I don't have any memory impressions that bear on this point. (But it is

---

<sup>9</sup> For discussions of a number of issues having to do with the nature of evidence from our senses, see G.E.M. Anscombe, "The First Person," in her *Metaphysics and the Philosophy of Mind* (Minneapolis: University of Minnesota Press, 1981); and Susanna Schellenberg, "Experience and Evidence," *Mind* (forthcoming). On the more general topic of self-knowledge, see Brie Gertler, "Self-Knowledge," in *The Stanford Encyclopedia of Philosophy* (2008). <http://plato.stanford.edu/entries/self-knowledge/>.



worth noting that this is largely because I don't have any memory impressions of the kind of glitzy soundtrack involving alien voices mentioned in footnote 1 above.) Nor do I have any relevant *a priori* intuitions. (After all, the question of whether I am a brain in a vat or a brain in a body appears to be a strictly contingent matter, and not something that could be decided from the armchair.) So it comes down to sensations.

Now, let me just check the evidence from my senses here. (Think of this paragraph as the live-blogging portion of the paper.) As I hold my hands out and look at them carefully, my visual sensations forcefully suggest that I have hands and arms. I rub my hands together and pat my arms, and as I do, my tangible sensations suggest the same thing. Clapping my hands lightly results in audible sensations of having hands, and when I hold one hand in front of my nose I get olfactory sensations of having a hand. (Nice soap!) Finally, putting one knuckle gingerly into my mouth produces taste sensations of having a hand. (Soap flavored, but unmistakably a hand.) Meanwhile, I have had no evidence at all (sensory or otherwise) of alien-scientist glitches. So my evidence overwhelmingly supports the proposition that I have hands, and arms, and a body.<sup>10</sup>

Evidentialism says that a belief is justified if and only if it is supported by the agent's evidence. Assuming Evidentialism, then, we have this situation.

N = the proposition that I am a normal brain in a body.

✓ I believe N.

✓ N is true.

✓ My belief that N is justified.

I think it is obvious that my evidence supports the belief that I am a normal brain in a body, since that is what my evidence makes it seem like to me. (And I think the same is true of your evidence and your belief that you are a normal brain in a body.) But here is an additional argument for this claim. Consider two brains:

---

<sup>10</sup> Similar moves have been made, or at least considered, by various other philosophers. See, for example, John Pollock, *Knowledge and Justification* (Princeton: Princeton University Press, 1974), Ch. 2-5; William Alston, *Epistemic Justification* (Ithaca: Cornell University Press, 1989), Essays 1-3; Robert Audi, *The Structure of Justification* (Cambridge: Cambridge University Press, 1993), Ch. 3, 4, 10, and 12; James Pryor, "The Skeptic and the Dogmatist," *Nous* 34 (2000): 517-549; Earl Conee and Richard Feldman, *Evidentialism* (Oxford: Oxford University Press, 2004), 300-301; Michael Huemer, *Skepticism and the Veil of Perception* (Lanham: Rowman & Littlefield, 2001), Ch. V; and Berit Brogaard, "Phenomenal Seemings and Sensible Dogmatism," in *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism*, ed. Chris Tucker (Oxford: Oxford University Press, forthcoming). For criticism of this general idea, see Roger White, "Problems for Dogmatism," *Philosophical Studies* 131 (2006): 525-557.

mine, which is a normal brain in a body, and Sally's, which is a brain in a vat. Now, the whole point of the brain-in-a-vat scenario is that it seems to the brain in the vat that it is a normal brain in a body. So it *looks* to Sally like she is a brain in a body, it *feels* to her like she is a brain in a body, etc. Sadly, Sally's evidence happens to be misleading. *But if it is misleading, then it is leading somewhere.* And where it is leading is toward the belief that she is a normal brain in a body. That's the belief that Sally's evidence supports. Thus Sally is justified in believing that she is a normal brain in a body.<sup>11</sup> And my evidence is just like Sally's in terms of where it is leading. My evidence is leading toward the belief that I am a brain in a body. So Sally and I both have justified beliefs (in her case, that she is a normal brain in a body; in my case, that I am a normal brain in a body).<sup>12</sup> In each case, the belief is justified by the agent's evidence. In my case, the belief is also true. So I have knowledge.<sup>13</sup> In Sally's case, the evidence is misleading and her belief is false. So she does not have knowledge.

Here is the situation with respect to Sally's belief that she is a normal brain in a body.

N2 = the proposition that Sally is a normal brain in a body.

✓ Sally believes N2.

✗ N2 is true.

✓ Sally's belief that N2 is justified.

The upshot is this: Brains in vats don't know that they are brains in bodies (because they have justified but false beliefs in the relevant propositions), but brains in bodies do know that they are brains in bodies (because they have justified, true beliefs in the relevant propositions).

Now, I think similar remarks apply with respect to Coherentism. I have a coherent system of beliefs, and my belief that I am a normal brain in a body fits nicely into that system. So, according to Coherentism, my belief is justified. (Notice that according to Coherentism, Sally also has a justified (but in her case false) belief.)

---

<sup>11</sup> Notice that the only way to resist this argument is to say that Sally's evidence is not misleading. And that's crazy. Poor Sally is the poster child for misleading evidence!

<sup>12</sup> Huemer makes a similar point about a brain in a vat in *Skepticism and the Veil of Perception*, Ch. V, Section 3.

<sup>13</sup> This is a little bit hasty, in light of the Gettier problem. But I am assuming that the skeptic is not resting her case on some tricky, Gettier-style consideration.

The idea seems to work with Reliabilism as well. My belief that I am a normal brain in a body is the result of a process involving my brain and my sensory organs, and this process is quite reliable. So, according to Reliabilism, my belief is again justified. (Notice that according to Reliabilism, however, Sally does not have a justified belief. (For Sally's belief-forming process is, unfortunately, highly unreliable.) So Reliabilism, unlike Evidentialism and Coherentism, entails that a brain in a vat does not have justified beliefs about the external world.)

It's time for a big-picture moment. According to what I have said, I know that I am a normal brain in a body, because my belief that N is supported by my evidence. And for similar reasons I know that BIVH is false. So premise (1) of The Brain In a Vat Argument is false. And so I know lots of things about the external world.<sup>14</sup>

Now let's turn to some likely objections.

#### 4. Objections and Replies

**Objection 1:** "You say that your evidence supports your belief that you are a brain in a body. *But your evidence equally supports the proposition that you are a brain in a vat.* After all, your evidence is perfectly consistent with BIVH!"

---

<sup>14</sup> I want to emphasize that my response to the skeptical challenge posed by The Brain In a Vat Argument is quite different from the famous Moorean response to such challenges. For the Moorean response (as applied to this version of the skeptical challenge) is to accept the conditional 'If I don't know that I am not a brain in a vat, then I don't know that I have hands,' and to *tollens* where the skeptic wants to *ponens*. That is, the Moorean wants me simply to accept as a premise the proposition that I know that I have hands. And the Moorean typically claims that I do not need any justification for the relevant claim because it is more plausible than any premise in any skeptical argument. My approach, on the other hand, is to accept that beliefs like my belief that I have hands are in need of justification, and to claim that such beliefs are in fact justified because they are supported by my sensory evidence. For more on the Moorean approach, see G.E. Moore, "Proof of an External World," reprinted in his *Philosophical Papers* (London: Allen & Unwin, 1959), 126-148; Annalisa Coliva, "The Paradox of Moore's Proof of an External World," *The Philosophical Quarterly* 58 (2008): 234-243; Ernest Sosa, "Moore's Proof," in his *Reflective Knowledge* (Oxford: Oxford University Press, 2011), 3-23; Jessica Brown, "Doubt, Circularity and the Moorean Response to the Sceptic," *Philosophical Perspectives* 19 (2005): 1-14; and Susana Nuccetelli, "Sosa's Moore and the New Dogmatists," *Metaphilosophy* 40 (2009): 180-186. Nor is my response to the skeptic similar to the Russellian "inference to the best explanation" response (see Bertrand Russell, *The Problems of Philosophy* (Oxford: Oxford University Press, 1912/1997), Chapter II). For I am not saying that I can make an abductive inference to the conclusion that I am a brain in a body from some premises; rather, I am claiming that my belief that I am a brain in a body is justified in virtue of being supported by my evidence.

**Reply:** The skeptic is right that my evidence is consistent with BIVH. But the skeptic is wrong about whether my evidence supports the proposition that I am a brain in a vat just as much as it supports the proposition that I am a brain in a body. For my evidence supports the proposition that I am a brain in a body to a very high degree, and it does not support the proposition that I am a brain in a vat at all.

Why does my evidence support the proposition that I am a brain in a body? Because that is how my evidence makes it seem. My evidence makes it seem like I have hands, and arms, etc. (And also there is the Sally argument. Sally's evidence is misleading, which entails that it is leading somewhere. And my evidence is relevantly like Sally's.)

Why does my evidence not support the proposition that I am a brain in a vat? Well, here are some possible ways my evidence could support the proposition that I am a brain in a vat: (i) There could be glitches in the soundtrack of my life, as described in footnote 1 above. (ii) I could have glitchy visual sensations of hands, which would make it look like my hands were flickering in and out of existence. (iii) I could have odd, not-quite-right tactile sensations of having hands. Those are descriptions of possible evidence that would support the proposition that I am a brain in a vat. But none of those possibilities is remotely similar to what my evidence is actually like.

Here is a related point. The skeptic who says that my evidence supports N and BIVH equally well, because it is consistent with both propositions, is making a huge mistake about the relation of support between some evidence and a proposition. It is not (as the skeptic thinks) that some evidence equally supports any hypothesis that is consistent with that evidence. I.e., the following principle is false.

**The Consistency Principle:** For any propositions,  $p$  and  $q$ , and body of evidence,  $e$ , if  $p$  and  $q$  are both consistent with  $e$ , then  $e$  supports  $p$  and  $q$  equally well.

Here's an argument against The Consistency Principle. Suppose you're playing blackjack and have been counting the cards. Suppose you know that there are 10 cards left in the deck, and that 9 of them are face cards. This evidence supports the hypothesis that the next card will be a face card. The evidence is also consistent with the hypothesis that the next card will be a number card, but it does not support that hypothesis *to the same degree*. If you're a betting person, you should bet that the next card will be a face card.

Or suppose you meet some guy who performs an extraordinary act of kindness. Your evidence supports the proposition that this is a nice guy. Your evidence is also consistent with the proposition that this guy is literally the devil,

and has some nefarious reason for convincing you that he is a nice guy; but your evidence does not support this devil hypothesis *as strongly* as it supports the mundane hypothesis that the fellow is a nice guy.

Likewise, my sensory evidence is consistent with BIVH, but it does not support that hypothesis to the same degree that it supports the hypothesis that I have hands and a body. In fact, my evidence doesn't support BIVH at all; for my evidence doesn't make it *seem* like I am a handless, disembodied brain in a vat.

**Objection 2:** "The Consistency Principle is admittedly a bad principle. But there is something in the ballpark that is true, and that entails that your evidence supports BIVH as well as it supports N. For it's not just that your evidence is consistent with BIVH. *It's that this is exactly how things would seem to you if BIVH were true!*"<sup>15</sup>

**Reply:** The skeptic is right in claiming that this is how things would seem to me if I were a brain in a vat (provided the alien scientists were good at what they were doing, and chose to give me sensations like these). But the skeptic is wrong to claim that it follows that my evidence supports N and BIVH equally well. Why? Well, this is the very plausible principle that the skeptic is appealing to here.

**The Counterfactual Principle:** For any propositions,  $p$  and  $q$ , and body of evidence,  $e$ , if  $e$  is qualitatively identical to what the evidence would be like if  $p$  were true, and also qualitatively identical to what the evidence would be like if  $q$  were true, then  $e$  supports  $p$  and  $q$  equally well.

Although it is initially very plausible, The Counterfactual Principle is actually false. Here is a counterexample. Suppose I have seemings as of an empty car parked on the street. This is how things would seem to me if there were an empty car parked on the street. But it is also how things would seem to me if there were a car filled with a dozen cleverly hidden clowns. And yet my evidence does not support these two propositions equally well, since it supports the empty-car hypothesis to a much higher degree than it supports the clown-car hypothesis. After all, my evidence does not make it seem like there are a dozen cleverly hidden clowns in that car – instead, the evidence makes it seem like the car is empty.<sup>16</sup>

---

<sup>15</sup> I am grateful to Katia Vavova for raising this objection.

<sup>16</sup> If you are tempted to say that the evidence *does* support the clown-car hypothesis as much as it supports the empty-car hypothesis, consider that there are an infinite number of such crazy hypotheses (the monkey-car hypothesis, the unicorn-car hypothesis, etc.). The evidence cannot support each of them to the same non-zero degree.

The upshot is that since the objection is based on The Counterfactual Principle, and the principle is false, the objection fails.

**Objection 3:** “It is notable that we have these two hypotheses, BIVH and N, and each of them predicts that things will seem to you the way they do. In fact, the two hypotheses predict your evidence to exactly the same degree. So if the evidence supports one of these hypotheses, then it supports the other one to the same degree.”

**Reply:** This objection is based on the following principle.<sup>17</sup>

**The Prediction Principle:** For any hypotheses,  $h1$  and  $h2$ , and observation,  $o$ , if  $P(o/h1) = P(o/h2)$ , then  $S(h1/o) = S(h2/o)$  (i.e., if the probability of  $o$  given  $h1$  equals the probability of  $o$  given  $h2$ , then  $o$  supports  $h1$  and  $h2$  equally well).

The Prediction Principle (like The Consistency Principle and The Counterfactual Principle) looks, at first glance, like a good principle, but it is not. Here are some counterexamples:

- The empty-car vs. clown-car hypotheses from above. Both predict that I will have experiences as of an empty car. But these empty car experiences I am having do not support the clown-car hypothesis as well as they support the empty-car hypothesis.
- The Obama-is-real vs. the Obama-is-an-elaborate-hoax hypotheses. Both hypotheses predict the experiences I have had as of there being a President Obama, but my experiences do not support the elaborate-hoax hypothesis to the same degree.
- The card counting case above. Let  $H1$  = the hypothesis that 9 of the 10 remaining cards are face cards and the next card will be a face card. Let  $H2$  = the hypothesis that 9 of the 10 remaining cards are face cards and the next card will be a number card. And let  $O$  = my current total evidence. Then  $P(O/H1) = P(O/H2)$ , but  $O$  does not support these two hypotheses equally well (since it supports  $H1$  to degree .9 and  $H2$  to degree .1).

These cases show that it is easy to come up with counterexamples to The Prediction Principle. Since the objection under consideration is based on that principle, and since the principle is false, the objection fails.

---

<sup>17</sup> Why the shift from talk of propositions and evidence to talk of hypotheses and observations? Just because the objection sounds better when put in terms of hypotheses and observations. But I take it that hypotheses are a kind of proposition, and that observations are a form of evidence.

**Objection 4:** “The real problem is that your evidence doesn’t allow you to rule out BIVH. And since that hypothesis is a genuine skeptical alternative to N, your evidence doesn’t support N.”

**Reply:** The idea behind this objection is that you need to be able to rule out competing hypotheses before some hypothesis is supported by your evidence. So the objection is based on the following principle.

**The Competing Hypotheses Principle:** For any hypotheses,  $h1$  and  $h2$ , and body of evidence,  $e$ , if  $h1$  and  $h2$  are competing explanations of  $e$ , and if  $e$  doesn’t allow the relevant agent to rule out  $h1$ , then  $e$  doesn’t support  $h2$ .

I admit that this is an initially very plausible principle. But it’s wrong. We have already seen several counterexamples: (i) the face card vs. number card hypotheses, (ii) the nice-guy vs. the devil hypotheses, (iii) the empty-car vs. clown-car hypotheses, (iv) the Obama-is-real vs. the Obama-is-an-elaborate-hoax hypotheses.

There is also a bigger problem with The Competing Hypotheses Principle. No matter what the evidence, and no matter what the hypothesis, there will always be some competing explanation (perhaps one that is undreamt of by the relevant agent) that is not ruled out by that evidence. So The Competing Hypotheses Principle will give the result that no evidence ever supports any hypothesis. And that can’t be right.

**Objection 5:** “It’s true that The Competing Hypotheses Principle is too strong, since it entails that no evidence ever supports any hypothesis. But there is something in the ballpark that is correct, and that doesn’t have the unwanted consequence that no evidence ever supports any hypothesis. If you are contemplating some hypothesis, and if there is some competing hypothesis that you cannot rule out, *and if that competing hypothesis is somehow made salient to you*, then your belief in the original hypothesis cannot be justified.”<sup>18</sup>

**Reply:** This objection is based on the following principle.<sup>19</sup>

---

<sup>18</sup> I am grateful to David Pitt for pressing this objection.

<sup>19</sup> I take it that the principle discussed in the text is equivalent to this one, which is more in keeping with the original Competing Hypotheses Principle: For any hypotheses,  $h1$  and  $h2$ , body of evidence,  $e$ , and agent,  $S$ , if  $e$  is  $S$ ’s total evidence, if  $h1$  and  $h2$  are competing explanations of  $e$ , if  $e$  doesn’t allow the relevant agent to rule out  $h1$ , and if  $h1$  has become salient for  $S$  as an explanation for  $e$ , then  $e$  doesn’t support  $h2$ .

**The Alternative-Hypothesis-Made-Salient Principle:** You cannot be justified in believing some hypothesis, *h1*, if there is a competing hypothesis, *h2*, such that *h2* is salient for you and you are not in a position to rule out *h2*.

Like the principles considered above, The Alternative-Hypothesis-Made-Salient Principle is initially plausible. But also like the above principles, it is false. Here are some counterexamples:

- You are wondering where your colleague Leila is. All of your evidence points to the hypothesis that she is in her office. Then for no special reason we put some random slips of paper into a hat and pull one out that says “MARS”. The degree of support for the hypothesis that Leila is in her office does not change, and you are still justified in believing that hypothesis.
- A variation on the parked car case from above. You are thinking the evidence supports the empty-car hypothesis. Then a fan of The Alternative-Hypothesis-Made-Salient Principle mischievously says, “Clown-car?” This does not change the degree of support for the empty-car hypothesis, or the fact that you are justified in believing it.
- A variation on the card counting case from above. You are about to bet that the next card will be a face card, when your friend says (for no special reason), “It could be that the next card will be a number card!” This does not change the degree of support for the face card hypothesis, which you are still justified in believing.
- We carefully inspect the room and come to believe that the only larger-than-a-bug, living organisms in the room are several humans. But then I say, just for fun, “What about the hypothesis that there are a bunch of mouse-sized, purple, Martian anthropologists who are super shy and who are sneaking around the room watching us while all the time avoiding detection?” You ask me whether I have any special reason to believe this hypothesis, and I truthfully answer, “No.” This conversation does not change the degree of support for your belief in the humans-only hypothesis, and your belief in that hypothesis continues to be justified.

It’s important to appreciate that in these examples, the only difference between before and after is that some competing hypothesis has become salient for the relevant believer. In particular, it is not the case that any new evidence has come in that undermines the evidence for the original hypothesis, or that functions as a defeater for whatever evidence supported that original hypothesis. And it seems clear that, in these examples, the mere fact that some alternative hypothesis becomes salient for the relevant agent does not change the degree of support by that agent’s evidence for the original hypothesis.



**Objection 6:** “You say that your evidence supports the belief that you are a brain in a body, because your sensations make it seem to you like you have hands. But your evidence equally supports the belief that you are a brain in a vat, since your sensations also make it seem to you like you are a brain in a vat being fed handish sensations. After all, we can imagine some weird person, Hilary, whose sensory evidence is relevantly like yours, but who interprets her evidence differently. Hilary notices the handish sensations she is having and concludes, ‘Hey, it looks like I’m a brain in a vat who is being fed handish sensations!’ Hilary’s belief that she is a brain in a vat being fed handish sensations is supported by her evidence. And since her evidence is relevantly like yours, your evidence also supports the proposition that you are a brain in a vat being fed handish sensations. So your evidence supports BIVH just as strongly as it supports N.”<sup>20</sup>

**Reply:** The objection involves saying that Hilary’s belief that she is a brain in a vat being fed handish sensations is supported by her evidence. But I think it is not. I think that when Hilary interprets her evidence as supporting the belief that she is a brain in a vat, she is doing it wrong. She is simply mistaken about what her evidence makes it seem like to her. Her evidence does not make it seem to her like she is a brain in a vat who is being fed handish sensations; instead, her evidence makes it seem to her like she is a brain in a body, with hands.

It is important to distinguish between the two parts of Hilary’s conjunctive belief: (i) that she is a brain in a vat, and (ii) that she is being fed handish sensations. Her evidence does support the second part. But not the first part.

To see why, consider another person, Dilary, who is having sensations similar to Hilary’s (and mine). Dilary notices the handish sensations she is having and concludes, “Hey, it looks like I am a creature with tentacles who is being fed handish sensations!” Dilary is right that it looks to her like she is being fed handish sensations. But she is wrong in thinking *it looks to her* like she is a creature with tentacles. There is nothing tentacle-ish about Dilary’s sensations (recall that Dilary’s sensations are similar to mine and Hilary’s), just as there is nothing robot-arm-ish or flipper-ish about her sensations.

Or consider Stilary. He believes, for no good reason, that every human is constantly accompanied by an invisible, intangible daemon, who hovers above the human, sometimes protecting the human (on occasions when the human is not harmed) and sometimes negligently failing to protect the human (on occasions when the human is harmed). Stilary experiences sensations of you walking into a room and thinks, “Hey, it looks like I am being fed sensations of a human and a

---

<sup>20</sup> I am grateful to Katia Vavova and Maura Priest for raising this objection.

daemon entering the room!” When it comes to interpreting his evidence and determining which propositions are supported by that evidence, Stilary, like Dilary, is doing it wrong. For his sensations support the proposition that a human has entered the room, but they do not support the proposition that a daemon has entered the room.

Meanwhile, it’s the same with Hilary. Her sensations do not point toward her being a brain in a vat (even if they are consistent with her being a brain in a vat, and even if they are what her sensations would be like if she were a brain in a vat). Her sensations point toward her having hands. Her sensations point toward her having a body. (Recall the argument above about how Sally (who is relevantly like me, but is a brain in a vat) has evidence that is misleading: We can all agree that Sally’s evidence is misleading; and if it is misleading, then it is leading somewhere false, namely, to the proposition that she has hands and a body.)

One upshot of these considerations is that first person reports of how things appear to an agent are fallible. It is possible to be mistaken about what your evidence makes it seem like to you.<sup>21</sup>

Another upshot is that neither Hilary nor Dilary nor Stilary has sensations that support any skeptical hypothesis. Their sensations make it seem to them like they have hands and bodies. And for the same reason, my sensations do not support the skeptical hypothesis that I am a brain in a vat.

**Objection 7:** “But how do you know it’s *true* that you’re a normal brain in a body? You just assumed that.”

**Reply:** I plead guilty. I did just assume that I am a normal brain in a body. But consider this proposition:

T = the proposition that it’s *true* that I am a normal brain in a body.

There is no significant difference between T and N. They are equivalent. Thus, what I have said above ensures that the conditions for knowledge are satisfied with respect to T as well as N.

- ✓ I believe T.
- ✓ T is true.
- ✓ My belief that T is justified.

So that’s how I know it’s *true* that I’m a normal brain in a body.

---

<sup>21</sup> There is of course another usage of ‘what x makes it seem like to S’ – the phrase can be used to mean *the belief that x causes S to have*. But this usage has nothing to do with justification.

**Objection 8:** “You say that you’re a normal brain in a body. But how do you *know* that? After all, for all you know, you might be a brain in a vat.”

**Reply:** Sometimes “How do you know that p?” is a question about your justification for believing p. But I have already answered that question. My justification is all of the evidence I have pointing toward my being a brain in a body. Other times “How do you know that p?” is a question about whether the philosophical conditions for knowledge have been satisfied in the case of your belief that p. But I have already answered that question, too.

**Objection 9:** “It feels like you’re cheating. You’re basically saying that IF you’re not a brain in a vat, THEN you can know lots of things about the external world. But the whole question is supposed to be whether you ARE a brain in a vat!”

**Reply:** No, that is not what the whole question is supposed to be. For I agree with the skeptic about the truth of the following two claims, which should both be completely uncontroversial.

**Claim 1:** Brains in vats don’t know much about the external world.

**Claim 2:** If I am a brain in a vat, then I don’t know much about the external world.

But I disagree with the skeptic about the truth of the following two claims.

**Claim 3:** Brains in bodies don’t know much about the external world.

**Claim 4:** If I am a brain in a body, then I don’t know much about the external world.

Since Claim 1 and Claim 2 are completely uncontroversial, the “whole question” is not about those claims. It is about Claim 3 and Claim 4. And my claim is that, since brains in bodies are justified in believing, for example, that they have hands, they know lots of things about the external world. So I’m not cheating.

Here is an analogy. Suppose I am dealing with a different kind of skeptic, a more mundane one, who says that I do not know where my car is parked (since it could have been stolen or towed since I last saw it). Suppose that we discuss this matter as we walk toward the spot where I think my car is parked. When we arrive, and see that my car is indeed there in the spot where I thought it was, my skeptical friend should not give up and say, “Oh, I guess you did know where your car was parked all along.” Instead, if she is a competent skeptic and a good philosopher, she should say, “Yay! I’m glad your car is where you thought it was,

Ned Markosian

so that we can now drive to the restaurant. But you still didn't know it was here a few blocks ago, when we started this conversation." This shows that it wasn't the *truth* of the relevant belief that was in question all along; it was whether the belief was *justified*.

It's the same with the controversy between the external-world skeptic and me. It is not the truth of the proposition that I am a brain in a body that is in dispute. It is whether that belief is justified. This is why it's not cheating to assume, as I have, that I am a brain in a body.

**Objection 10:** It might appear that I am committed to an inconsistent triad. First, I have said that the *degree of support* that my evidence gives to the proposition that I am a normal brain in a body is very high (greater than .9, say, on a scale from 0 to 1). In short (using some natural notation),

$$(a) \quad S(N/e) > .9,$$

where 'N' is the proposition that I am a normal brain in a body, and 'e' is my total evidence.

Second, it's very natural to say that the *probability* that I am a brain in a body is either unknown or unknowable or inscrutable or undefined. (After all, who knows what the percentage of brains in vats is, throughout the universe, among brains having experiences like mine?) In short (using some familiar notation),

$$(b) \quad P(N/e) \text{ is unknown or unknowable or inscrutable or undefined.}$$

And third, it also looks like I will be committed to what we might call The Equivalence Thesis: The degree of support for some proposition by someone's evidence is equal to the conditional probability of that proposition given that evidence. In short,

$$(ET) \quad S(p/e) = P(p/e).$$

But (a), (b), and (ET) form an inconsistent triad, because .9 is not unknown or unknowable.

**Reply:** I think there might be independent reasons to question (ET), but it is certainly pretty appealing. In any case, I am not going to deny it here. Instead, I think that the best way to deal with this problem is to deny (b). My evidence (namely, my current sensations) not only makes it *seem* like I am a normal brain in a body, but also makes that proposition *very probable*.

## Do You Know That You Are Not a Brain In a Vat?

Here is an argument for the claim that my current total evidence makes it very probable that I am a brain in a body. Suppose I get some very credible new evidence suggesting that the probability that I am a brain in a vat is .5. For example, suppose an oracle appears before me, convinces me that she is omniscient and reliable, and informs me that half of the brains like mine in the universe are brains in vats. This would be a life-changing experience that would seriously undermine my previous sensory experience for N. After the experience with the oracle, the probability that I am a brain in a body, given my total evidence, would be .5. But this would be because the new evidence from the oracle had *dramatically lowered* the probability of N. Which suggests that the probability of N given my evidence before the oracle showed up was *extremely high*. In fact, it looks like the skeptic pushing this objection has to maintain that the new evidence from the oracle *doesn't change the probability of N at all*, and that seems crazy.

Now, you might think the skeptic can get around this argument by saying that the new evidence from the oracle changes the probability of N, given my total evidence, from unknown to .5. But I don't think that response on the part of the skeptic can work, because in other, less dramatic cases – like a brief glitch in the sensory track, or an encounter with a less convincing oracle – it is clear that the new evidence is lowering the probability conferred by my total evidence *incrementally*, rather than flipping it in a binary way from unknown to .5 or whatever.

**Objection 11:** “Given your evidence, the probability that you are just a brain in a vat is equal to the probability that you are a brain in a body. It's 50-50. And if two propositions are equally probable, given some evidence, then that evidence supports them equally well. So your evidence supports BIVH just as much as it supports N.”

**Reply:** This objection involves an appeal to the following principle.

**The Probability Principle:** For any propositions,  $p$  and  $q$ , and body of evidence,  $e$ , if  $P(p/e) = P(q/e)$ , then  $S(p/e) = S(q/e)$ .

The Probability Principle, unlike some of the other principles we have seen, is a good principle. Or at any rate, it follows from (ET), so it is as good as (ET), which I am accepting for the purposes of this paper. The problem with the objection is the claim that  $P(\text{BIVH}/e) = P(N/e)$  (where  $e$  is my total evidence). For as I have just argued (using the oracle example),  $P(N/e)$  is very high, while  $P(\text{BIVH}/e)$  is super low.

Here is a bonus argument for this latter claim. It seems clear that  $S(N/e)$  is very high, and also that  $S(BIVH/e)$  is very low. (I have given two arguments for this conclusion. The first was an argument from the claim that my sensations make it seem like I am a brain in a body, not a brain in a vat. And the second argument for this conclusion was from the claim that Sally (who is an unfortunate brain in a vat) has evidence that is misleading.) So by (ET) we immediately get the result that  $P(N/e)$  is high, while  $P(BIVH/e)$  is low.

Notice that if the skeptic tries to resist this argument by rejecting (ET), then the objection currently under consideration cannot get off the ground, since it needs (ET) to justify The Probability Principle.

## 5. A Final Comment on the Problem of External-World Skepticism

I want to make one final comment on the problem of external-world skepticism. You might have a lingering feeling that I have not really solved the problem. If so, I think this is likely due to a confusion of two different problems.

First, there is an *epistemological problem*: How can I respond to the skeptic who argues that I don't know anything about the external world? To solve this problem I would need a reason to reject one of the premises of The Brain In a Vat Argument. And as we have seen, if I am to reject premise (1) of that argument, this will come down to showing that most of my ordinary beliefs about the external world, including my belief that BIVH is false, are justified.

Second, there is a *practical problem*: Is there some test or procedure that I can perform to determine which situation I am in – a brain in a vat situation or a brain in a body situation? To solve this problem I would need to come up with something analogous to the pinch test for dreaming. That is, I would need to come up with a “pinch test” for the brain in a vat scenario, such that if I perform this test, then it will all of a sudden be revealed to me whether I am a brain in a vat.

I claim to have solved the first problem, the epistemological one. I have argued that ordinary beliefs about the external world, including my belief that BIVH is false, are justified by our sensory evidence. Which means that ordinary knowledge claims about the external world are true – including my belief that I am not a brain in a vat. But I do not claim to have solved the second problem. In fact, I think there is no solution to this practical problem. The answer to the question is *No, there is no “pinch test” for the brain in a vat scenario*. But luckily, this does not change the fact that I do know lots of things about the external

## Do You Know That You Are Not a Brain In a Vat?

world, contrary to what the skeptic and The Brain In a Vat Argument have claimed.<sup>22</sup>

---

<sup>22</sup> Earlier versions of this paper were presented at Western Washington University, The University of Wisconsin – Milwaukee, Cal State – LA, UC – Irvine, Boğaziçi University, Yerevan State University, The University of Illinois, and The University of Alabama. I am grateful to members of all eight audiences for helpful comments. I'm also grateful to Katia Vavova and Dan Korman for helpful comments on an earlier version of this paper, and to Sharon Ryan for helpful discussions of this topic.





# MAXWELLIAN SCIENTIFIC REVOLUTION: A CASE STUDY IN KANTIAN EPISTEMOLOGY

Rinat M. NUGAYEV

**ABSTRACT:** It is exhibited that maxwellian electrodynamics was created as a result of the old pre-maxwellian programmes reconciliation: the electrodynamics of Ampere-Weber, the wave theory of Young-Fresnel and Faraday's programme. The programmes' meeting led to construction of the whole hierarchy of theoretical objects starting from the genuine crossbreeds (the displacement current) and up to usual mongrels. After the displacement current construction the interpenetration of the pre-maxwellian programmes began that marked the beginning of theoretical schemes of optics and electromagnetism real unification. Maxwell's programme did supersede its rivals because it did assimilate some ideas of the Ampere-Weber programme, as well as the presuppositions of the programmes of Young-Fresnel and Faraday. Maxwellian programme's victory over its rivals became possible because the core of Maxwell's unification strategy was formed by Kantian epistemology looked through the prism of William Whewell and such representatives of Scottish Enlightenment as Thomas Reid and William Hamilton. It was Kantian epistemology that enabled Hermann von Helmholtz and his pupil Heinrich Hertz to arrive at such a version of Maxwell's theory that could serve a heuristical basis for the radio waves discovery.

**KEYWORDS:** James Clerk Maxwell, Hermann von Helmholtz, Heinrich Hertz, scientific revolution, Immanuel Kant, epistemology

## Introduction

The aim of the present paper is to answer the question "Why did Maxwell's programme supersede its rivals?" It appears that to answer it one has to take a further step in revealing the inter-theoretical context of Maxwellian electrodynamics genesis and growth. The reconstruction should provide a "theoretically progressive problemshift" relative to other "internal" reconstructions and argue that Maxwellian revolution is a more complex phenomenon than appears from the standpoints of some well-known scientific revolution conceptions.<sup>1</sup> I'll try to demonstrate that *the Maxwellian programme*

---

<sup>1</sup> Thomas Kuhn, "Objectivity, Value Judgement and Theory Choice," in his *The Essential Tension* (Chicago: University of Chicago Press, 1977), 320-339; Imre Lakatos, "The Methodology

*had superseded its rivals because it had constantly communicated with them.* For instance, the Maxwellian programme did assimilate some of the propositions of the Ampere-Weber “hard core,” as well as some propositions of the Faraday and Young-Fresnel programmes. But the opposite statement is not true. Ampere-Weber programme did not assimilate the propositions of the Maxwellian programme. Maxwell’s research programme did supersede that of Ampere-Weber because it was a “*synthetic*”<sup>2</sup> one. It appeared, according to one of Maxwell’s (Kantian) philosophical gurus, one of “successive steps by which we gradually ascend in our speculative views to a higher and higher point of generality.”<sup>3</sup> Contrary to Maxwell’s, the Ampere-Weber programme was a *reductionist*<sup>4</sup> one for it tried to reduce all the theoretical ontologies to one and the same ontology of “action at a distance.”

According to Ludwig Boltzmann, “It is certainly useful to set up Weber’s theory as a warning example for all times that we should always preserve the necessary *mental flexibility*.”<sup>5</sup> Boltzmann constantly emphasized the need for a “*plurality of approaches*,” including both mathematical formalism and picture-based physical theories.

In particular, Maxwell’s programme was not only successful to assimilate the propositions of the Ampere-Weber hard core, combining them with Faraday’s “field” notions, as well as with those of Fresnel-Young optics; it was open for the synthesis with other research traditions as well. For instance, as Heinrich Hertz had put it,

From the outset Maxwell’s theory excelled all others in elegance and in *the abundance of the relations* between the various phenomena which it included.<sup>6</sup>

This “abundance of the relations” was due to that Maxwell did put forward as a synthetic principle the idea, that differed from that of Ampere-Weber by its flexible and contra-ontological, strictly epistemological, Kantian character.

---

of Scientific Research Programmes,” in his *Philosophical Papers, Volume 1* (Cambridge: Cambridge University Press, 1978).

<sup>2</sup> Rinat Nugayev, *Reconstruction of Mature Theory Change: A Theory-Change Model* (Frankfurt am Main: Peter Lang, 1999).

<sup>3</sup> William Whewell, *The Philosophy of Inductive Sciences: Founded Upon Their History*, Second edition, vol. 2 (London: John W. Parker and Son, 1847), 74.

<sup>4</sup> Nugayev, *Reconstruction of Mature Theory Change*.

<sup>5</sup> Quoted from Jed Buchwald, *The Creation of Scientific Effects: Heinrich Hertz and Electric Waves* (Chicago: The University of Chicago Press, 1994), 261.

<sup>6</sup> Heinrich Hertz, *Electric Waves* (London: Macmillan, 1893), 19.

## Maxwellian Scientific Revolution: A Case Study in Kantian Epistemology

By referring everything to the purely geometrical idea of the motion of an imaginary fluid, I hope to attain *generality* and precision, and to avoid the dangers arising from a premature theory professing to explain *the cause* of the phenomena.<sup>7</sup>

For Maxwell, ether was not the ultimate building block of physical reality, from which all the fields and charges should be constructed. “Action at a distance,” “incompressible fluid,” “molecular vortices” were “contrived analogies”<sup>8</sup> for Maxwell, capable only to direct the researcher at the “right” mathematical relations. Maxwellian analogy is contrived and is not intended to illustrate anything in nature. Maxwell gave a new meaning to analogy that comes close to modeling in current usage. For instance, according to John von Neumann, modern sciences do not try to explain, they hardly even try to interpret, they mainly make models. And model is a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a construct is that it is expected to work.

Usually the defining feature of all analogies is supposed to be a *bidirectional* relation between the two domains for which an analogy is established. Neither domain is privileged over the other. Relation holds both ways: one can move from one domain to its analogue and vice versa. But this feature does not hold in Maxwell’s novel methodology of mathematical analogy – it is *unidirectional*, from a fictional system to a physical system, where the purpose of introducing the fictional system is to gain insight into the physical system and ultimately to recast it into the mathematical formalism.

The principle of usual (“physical”) analogy between theories in two different domains that are identical in nature came from William Thomson. But for Maxwell the methodology of analogy was only a tool. Contrary to Thomson, both mathematically identical systems need not exist in nature. In a pair of such systems one of them could be imaginary (“imaginary fluid”), and the other real (“physical”).

From the “presentational” point of view all this hydrodynamic models were doomed to failure efforts to describe what cannot be described in principle – “things in themselves,” the “nature” of electrical and magnetic phenomena. On the contrary, Maxwell aimed his programme at finding empirically meaningful

---

<sup>7</sup> James Maxwell, “On Faraday’s Lines of Force,” *Transactions of the Cambridge Philosophical Society*, X, part 1 (1856). Reprinted in *The Scientific Papers of James Clerk Maxwell*, vol. 1 (New York: Dover, 1952), 155-229.

<sup>8</sup> Giora Hon and Bernard R. Goldstein, “Maxwell’s Contrived Analogy: An Early Version of the Methodology of Modeling,” *Studies in History and Philosophy of Modern Physics* 43 (2012): 236-257.

mathematical relations between the electrodynamics basic objects, i.e. the creation of concordant electromagnetic field equations system.

Hence even Ludwig Boltzmann agreed with Hertz that Maxwell's concepts of charge and current were "irremediably obscure." In his lectures he adopted Hertz's view that electricity was a "thing of thought, serving to picture the integrals of certain equations."<sup>9</sup>

It seems to me that one of the insufficiencies of current Maxwellian revolution studies<sup>10</sup> is an *underestimation of Maxwell's own methodology* created by himself for his ambitious project of mechanics, electrodynamics and optics unification. In every field of creativity (including epistemology) Maxwell always took his own way; and he tried to teach his students in the same way too.

As the author of *Treatise on Electricity and Magnetism* himself put it in one of his letters, "I find I get fonder of metaphysics and less of calculations continually."<sup>11</sup> One can remember Gustav Kirchhoff: "He is a genius, but one has to check his calculations."

It seems to me that one should take Ludwig Boltzmann's comments on Maxwell's works more literally. In his lectures on Maxwell's theory as well as in his comments on Maxwell's electromagnetic papers (that he had translated into German), the founder of statistical mechanics had pointed out that many Maxwell's works but especially his early electrical papers "were *not properly understood*." Perhaps it can be explained by the fact that these works "written according to the *long-term plan*" demonstrate that their author "was *as mastermind in theory of knowledge as he was in the field of theoretical physics*."<sup>12</sup> Maxwell was a great scientist as well as a great innovator of methodology.<sup>13</sup> Maxwell's methodology that sprung out from an intention to find a fruitful

---

<sup>9</sup> Quoted from Buchwald, *The Creation of Scientific Effects*, 258.

<sup>10</sup> See, for instance, Daniel Siegel, *Innovation in Maxwell's Electromagnetic Theory: Molecular Vortices, Displacement Current, and Light* (Cambridge: Cambridge University Press, 1991); Margaret Morrison, *Unifying Scientific Theories: Physical Concepts and Mathematical Structures* (Cambridge: Cambridge University Press, 2000); Olivier Darrigol, *Electrodynamics from Ampere to Einstein* (Oxford: Oxford University Press, 2001).

<sup>11</sup> Lewis Campbell and William Garnett, *The Life of James Clerk Maxwell* (London: Macmillan, 1882), 298.

<sup>12</sup> Ludwig Boltzmann, *Ueber Faraday's Kraftlinien, von James Clerk Maxwell. Herausgegeben von L. Boltzmann.* (Leipzig: W. Engelmann, 1895).

<sup>13</sup> Mary Hesse, "Logic of Discovery in Maxwell's Electromagnetic Theory," in *Foundations of Scientific Method: The Nineteenth Century*, eds. Ronald N Giere, Richard Samuel Westfall (Bloomington: Indiana University Press, 1973), 86-114; Peter Achinstein, "What to Do if You Want to Defend a Theory You Cannot Prove: A Method of 'Physical Speculation,'" *The Journal of Philosophy* 107 (2010): 35-55.

compromise between the extremes of Kantian relativism and Scottish “common sense realism” was a necessary part of his ambitious unification of optics and electromagnetism design.

Previous nineteenth century physics studies have oscillated between *two extremes*. On the one hand, in the more traditional vein, differences between research traditions were considered to be insignificant and communication unproblematic. On the other hand, in the more recent, post-Kuhnian, studies, differences between traditions are often taken to be so radical that communication is impossible among them.

This study originates from an **intermediate picture**. According to it, profound differences between the field and action at a distance research traditions existed at various levels, ranging from ontological commitments and up to epistemological beliefs. Yet these antagonistic traditions were able to communicate in the creative acts of such men of science as Maxwell, Helmholtz and Hertz.

They communicated in the ways that permitted comparisons, adaptations and *cross-fertilizations* of different traditions as well.

For instance, James Clerk Maxwell himself many times, beginning from his first paper and up to the last one, had pointed out that the key ideas of the Ampere- Weber electrodynamics were as useful for electrodynamics development as those of the field theories. Even at the beginning of his electrodynamics studies, on May 1855, a post-graduate student at Cambridge writes a letter to his father, stressing the importance of studying the theories of “heavy German writers.” Further, completing his theory creation on the basis of Lagrangian formalism, in his introduction to *A Dynamical Theory of Electromagnetic Field* (read December 8, 1864) Maxwell characterizes action at a distance theory as “exceedingly ingenious and wonderfully comprehensive.” Yet he had to make the following reservation:

The mechanical difficulties, however, which are involved in the assumption of particles acting at a distance with forces which depend on their velocities are such as to prevent me from considering this theory as an *ultimate* one, though it may have been, and *may yet be useful in leading to the coordination of phenomena*.<sup>14</sup>

And, at last, in his *Treatise on Electricity and Magnetism* (1873), Maxwell renders the creation of his system of equations in the following way:

---

<sup>14</sup> James Maxwell, “A Dynamical Theory of Electromagnetic Field,” *Philosophical Transactions of the Royal Society of London* 155 (1865): 459-512. Reprinted in *The Scientific Papers of James Clerk Maxwell, vol.1* (New York: Dover, 1952), 526-97.

I was aware that there was supposed to be a difference between Faraday's way of conceiving phenomena and that of the mathematicians, so that *neither he nor they* were satisfied with each other's language. I had also the conviction that this discrepancy did not arise *from either party being wrong*.<sup>15</sup>

The intermediate picture of our study comes from the critique of Kuhnian and Lakatosian conceptions' drawbacks: they lack the mechanisms of the paradigms' (or scientific research programmes') *interactions*.<sup>16</sup> To meet the critical comments, a "mature theory-change model" was proposed based on the "communicative rationality" concept.<sup>17</sup> According to the epistemic model, the origins of scientific revolutions lie not in a clash of fundamental theories with facts, but of "old" fundamental theories with each other, leading to contradictions that can only be eliminated in a more general theory. The key role in theory change is played by the proponents of the old paradigms' *dialogue* that leads to intercorrections and interpenetrations of the participants' initial views .

The very realization of reductionist and synthetic research programmes is brought about by the clash of mature theories which they are designed to eliminate. Having compared the heuristic potentials of the reductionist and the synthetic programmes, I favor the latter one since it has the following objective advantages.<sup>18</sup> Firstly, synthetic programmes should provide a greater empirically-progressive shift of problems solved than the reductionist ones. Secondly, only these programmes can rationally explain the use of the so-called *crossbred* theoretical objects which spring from the coincident theories. For instance, if one considers the structures of two mature modern theories – quantum theory and general relativity – he finds that their global theoretical schemes arose from the unification of the crossbred theoretical ones.

Every case of different programmes' meeting leads to a situation when a domain of *hybrid* models occurs formed by simple conjunctions from the models of different research programmes. However, the mongrel models appear to be self-contradictory; and when this is realized, the crossbreeds from the basic objects of all the cross-theories are constructed. A new mature theory is formed due to crossbred domain growth.

---

<sup>15</sup> James Maxwell, *A Treatise on Electricity and Magnetism*. 2 vols, third ed. (New York: Dover, 1954), 499.

<sup>16</sup> Rinat Nugayev, "A Study of Theory Unification," *The British Journal for the Philosophy of Science* 36 (1985): 159–73; Rinat Nugayev, "The History of Quantum Theory as a Decisive Argument Favoring Einstein over Lorentz," *Philosophy of Science* 52 (1985): 44–63.

<sup>17</sup> Rinat Nugayev, "Einstein's Revolution: a Case Study in Communicative Rationality," *Foundations of Science* 4 (1999): 155–204.

<sup>18</sup> Nugayev, *Reconstruction of Mature Theory Change*.

The aforesaid is not to diminish the role of experiments in science. On the contrary, the epistemic model proposed seems to elaborate further the point of view stated in the current literature that both theorists and experimentalists have breaks in their respective traditions, but they are not typically simultaneous.<sup>19</sup> Theory development must have, to some extent, a life of its own. The development of two main cultures within science does not mean that the two do not speak to each other.

The epistemic model was illustrated with reference to physics in the early twentieth century, the three “old” theories in this case being Maxwellian electrodynamics, statistical mechanics and thermodynamics.<sup>20</sup> The world of “old,” pre-Einsteinian physics was conceptually and socially fragmented. It was split on at least 3 research traditions belonging to electrodynamics, thermodynamics and mechanics. Traditions organized around different groups of phenomena generated little support for one another. The practitioners of each theoretical tradition acknowledged the existence of the other but went their own separate ways. With the advent of relativity and quantum theory, the conceptual unification of worldviews was accompanied by a social unification of practice.

### 1. Maxwellian Methodology of Synthetic Mature Theory Construction

Maxwell was not the first to unify optics and electromagnetism. Yet he was dissatisfied with the way his predecessors had done it. Why? The following passage helps to find the answer: the theories of action at a distance were too formal and abstract to grasp the intricate connections between the electromagnetic phenomena.

No electrical theory can now be put forth, unless it shows the *connexion* not only between electricity at rest and current electricity, but between the attractions and inductive effects of electricity in both states... The results of this simplification may take the form of a purely mathematical formula or of a physical hypothesis. In the first case we entirely lose sight of the phenomena to be explained; and though we may trace out the consequences of given laws, we can never obtain more extended views of the *connexions* of the subject.<sup>21</sup>

His predecessors were Hans Christian Oersted (1777-1851), Andre-Marie Ampere (1775-1836), Wilhelm Weber (1804-1890), Michael Faraday (1791-1867)

---

<sup>19</sup> Andrew Pickering, *Constructing Quarks. A Sociological History of Particle Physics* (Chicago: The University of Chicago Press, 1985); Peter Galison, *How Experiments End* (Chicago: The University of Chicago Press, 1987).

<sup>20</sup> Nugayev, “The History of Quantum Theory.”

<sup>21</sup> Maxwell, “On Faraday’s Lines of Force,” 155.

and William Thomson (1824-1907). Yet Maxwell's Weltanschauung was characterized by **an extraordinary high level of philosophical culture**. A brilliant student at Edinburgh and Cambridge and a post-graduate at Cambridge was enchanted by a profound skepticism of David Hume, George Berkeley and Immanuel Kant at the lectures of Sir William Hamilton on mental philosophy at Edinburgh University.

Hamilton's lectures, which were a prominent element in the Scottish university curriculum, "*interested him greatly*." From the Class of Metaphysics his mind "gained many lasting impressions;"<sup>22</sup> the lectures of Sir Hamilton made a strong impression on him, in "stimulating the love of speculation to which his mind was prone."

Sir William Hamilton (1788-1856) was one of the outstanding representatives of Scottish "common sense philosophy," an heir of Thomas Reid and James Stewart. Yet he stressed *Kant's proposition that all knowledge is relative*; so we know nothing about things themselves except by their *relationship* to other things. He had stimulated a spirit of criticism in his pupils by insisting on the great importance of *psychology* as opposed to the older metaphysical method.

Hamilton's "philosophy of the conditioned" surely had a strong Kantian flavor. Like Kant, he held that we can have knowledge only of "the relative manifestations of an existence, which in itself it is our highest wisdom to recognize as beyond the reach of philosophy." But unlike Kant, however, **he had argued for the position of a "natural realism"** in the Reidian tradition.

The Reverend Thomas Reid (1710-1796) directed his *An Inquiry into the Human Mind on the Principles of Common Sense* (1764) against Hume and Berkeley. It is here – he argued – that the "danger of the idealism" lies – in its reduction of reality to "particular perceptions," **essentially unconnected** with each other. The unit of knowledge is not an isolated impression but a judgement; and in such a judgement is contained the reference both to a permanent subject and to a permanent world of thought, and, implied in these, such judgements, for example, as those of existence, substance, cause and effect. Such principles are not derived from sensations, but are "suggested" on occasion of sensation, in such a way as to constitute the necessary conditions of having perceptive experience at all.

The doctrine of relativity of knowledge has seemed to many – including James Stewart Mill – contradictory to his realism. But for Hamilton, the two are held together by a kind of intuitionism that emphasizes certain facts of

---

<sup>22</sup> Campbell and Garnett, *The Life of James Clerk Maxwell*.



consciousness that are both primitive and incomprehensible. They are though constitutive of knowledge, “less forms of cognitions than of beliefs.”<sup>23</sup>

The relativism or phenomenalism which Hamilton adopted from Kant and sought to engraft upon Scottish philosophy is absent from the original Scottish doctrine. Thus, denying Hume’s skepticism, Hamilton did his best to *find a compromise* between Kant’s relativism and Reid’s realism; and it was namely that that Maxwell have pointed out as a basic thesis of his metaphysical programme on moving from Edinburgh to Cambridge:

in the meantime I have my usual superfluity of plans... 4. Metaphysics – **Kant’s Kritik of Pure reason in German, read with a determination to make it agree with Sir W. Hamilton...**<sup>24</sup>

The “Copernican revolution” in epistemology that had been initiated by Kant consisted in that the world of usual every-day experience (or Edmund Husserl’s “*lebenswelt*”) had lost its dominating position in interpreting things that can be perceived by our senses. Kant had exchanged the world of common experience by the world of Galilean experimental and mathematical physics based on the idealizations of the “*lebenswelt*” phenomena. Hence truth became something not spontaneously revealing and disclosing itself but something that can be comprehended only by a special (“scientific”) method.

On the other hand, if truth is comprehended only in experience and we can grasp not “the things by themselves” but just the “phenomena,” it is necessary to reject the opportunity of reaching the absolute truth. Our sensory representation is by no means a representation of things “in themselves,” but only of the way in which they appear to us. Hence the “**analogies of experience**” are especially important in Kant’s epistemology. Kant states that the cognition *according to analogy* does not signify, as the word is usually taken, an imperfect similarity between two things, but rather a perfect similarity between two *relations* in wholly dissimilar things. For instance, such is an analogy between the legal relation of human actions and the mechanical relation of moving forces. One can never do anything to another without giving him a right to do the same to me under the same condition; just as a body cannot act on another body with its motive force without thereby causing the other body to react just as much on it. Right and motive force are here completely dissimilar things, but in their relation there nonetheless complete similarity. By means of such an analogy one can

---

<sup>23</sup> Quoted from Robert Audi, ed., *The Cambridge Dictionary of Philosophy* (Cambridge: Cambridge University Press, 1999), 360.

<sup>24</sup> Campbell and Garnett, *The Life of James Clerk Maxwell*, 77.

therefore provide a concept of a relation to things that are absolutely unknown to him.<sup>25</sup>

In more detail, in *The Critique of Pure Reason* [1787], Kant considers a more interesting example:

we are able, with the guidance of these analogies, to reason in the series of possible perceptions from a thing which we do really perceive to the thing we do not perceive. Thus, *we cognize the existence of a magnetic matter penetrating all bodies from the perception of the attraction of the steel-fillings by the magnet*, although the constitution of our organs renders an immediate perception of this matter impossible for us.<sup>26</sup>

It is rather important that *even the example of the analogy of experience was taken by Kant from the domain of electromagnetism* thus paving the way to Maxwell. The latter had pointed out many times that things we can measure directly, like mechanical force, are merely the outward manifestations of deeper processes, involving entities like electric field strength, which are *beyond* our power of visualization.<sup>27</sup>

A more detailed exposition of Maxwell's research programme that he had followed through all his life is given in his truly philosophical works – in a speech “Are There Real Analogies in Nature?” read at the “Apostles” Cambridge club in 1856 (just after the publication of his most profound paper “On Faraday's Lines of Force,” 1855-1856) – and in his trailblazing paper “Helmholtz” (1877).

The Cambridge lecture is not a crude exposition of Kant's epistemology but a tense discussion of Maxwell with “Kant in himself.” It is not accidental that the very heading of the speech contains a question and not an assertion: “Are There Real Analogies in Nature?”

Maxwell gives no definite and unambiguous answer – in full accordance with Kant's antinomies that occur to Human Reason as attempts to overstep the Limits of Experience. He multiplies arguments pro and contra the proposition that certainly there are real analogies in Nature. Certainly Maxwell's thinking in terms of Kantian antinomies is not accidental. Following Hamilton's traditions, Maxwell tries to find *his own way* between the Scylla of Kantian transcendentalism and the Charybdis of Scottish common sense realism.

---

<sup>25</sup> Immanuel Kant, “Prolegomena to any future metaphysics that will be able to come forward as science,” in *Immanuel Kant. Theoretical Philosophy after 1781*, trans. Gary Hatfield (Cambridge: Cambridge University Press, 2002), 146-47.

<sup>26</sup> Immanuel Kant, *The Critique of Pure Reason* (University Park: The Pennsylvania State University Press, 2010), 170.

<sup>27</sup> Basil Mahon, *The Man Who Changed Everything. The Life of James Clerk Maxwell* (London: John Wiley, 2003).

In modern literature the Scottish view of knowledge is characterized by the following principles:<sup>28</sup>

1. All knowledge is relational.
2. Analogies are among the chief such relational ways of knowing.
3. Analogies are necessary for psychological reasons. For most people, understanding requires the use of analogies for simplifying and organizing knowledge.
4. Strong psychological tendencies in the Scottish Common Sense tradition admit reconciliation with logical and analytical trends of Kant's philosophy.

Hence for Maxwell the philosophical resolution of the antinomies comes from adopting *partial points of view*, as all human knowledge is partial. No absolute truth is attainable. What remains is establishing correspondences or analogies. Whenever one sees a relation between two things he knows well, and thinks that there must be a similar relation between things less known, he reasons from the one to another.

It supposes that although pairs of things may differ widely from each other, the relation in the one pair may be the same as that in the other. Since in a scientific point of view the *relation* is the most important thing to know, a knowledge of the one thing leads us a long way toward a knowledge of the other. If all that one knows is relation, and if all the relations of one pair of things correspond to those of another pair, it will be difficult to distinguish the one pair from the other, although not presenting a single point of resemblance, unless one has some difference of relation to something else whereby to distinguish them. Such mistakes can hardly occur except in mathematical and physical analogies.

Thus, the first lesson taught by Kantian epistemology – (I) “*the principle of relational character of scientific truth*” stating that the *relation* is the most important thing to know. It should be pointed out that even the examples of the analogies are taken by Maxwell from Kant's “Prolegomena.” Hence it is not surprising that the second principle – (II) “*theory laidness of observation*” – is also extracted by Maxwell from Kant:

The dimmed outlines of phenomenal things all merge into one another unless we put on the focusing glass of theory, and screw it up sometimes to one pitch of

---

<sup>28</sup> John Mertz, *A History of European Thought in the Nineteenth Century*, 4 vols. (Edinburgh: William Blackwood and Sons, 1964); Richard Olson, *Scottish Philosophy and British Physics, 1750-1880: A Study in the Foundations of the Victorian Scientific Style* (Princeton: Princeton University Press, 1975).

definition and sometimes to another, so as to see down into different depths through the great millstone of the world.<sup>29</sup>

The importance of the principle (II) for Maxwell's methodology cannot be overestimated. In nature all the phenomena are interconnected and merge into one another; all the differences in theoretical approaches are due to the fact that their authors focus on the different facets and different levels of the phenomena investigated. Hence a theoretician's task is to provide the "appropriate ideas" (Whewell's term) to cover the various domains of experience. But where should he find them? – In experience, from immediate generalizations of the experimental data? – Another piece of Maxwell's creativity – a part of his 1854 letter – makes it possible to take a glance at his thought laboratory:

It is hard work grinding out 'appropriate ideas,' as Whewell calls them. I think they are coming out at last, and by dint of knocking them against all the facts and half-digested theories afloat, I hope to bring them to shape, after which I hope to understand something more about inductive philosophy that I do at present.<sup>30</sup>

Now it is clear where the "appropriate ideas" come from: they are not the slavish copies of things, but are the a priori forms by which a chaos of sensations is "brought to order." At first the "appropriate ideas" are vague and dim; however in the long run they are "grinded out" by knocking them with the "facts" and the other theories. However the theoretician's task is not only to introduce and polish subtle notions "reflecting" the different facets of the phenomena under consideration, but also to unify the notions in synthesis.

The outlines and the stages of such a synthesis are described in Maxwell's paper "Hermann Ludwig Ferdinand Helmholtz." Maxwell points out that the ordinary growth of human knowledge is by accumulation round a number of *distinct* centers. Yet the time must sooner or later arrive when two or more departments of knowledge can no longer remain independent of each other, but *must* be fused into a consistent whole. But though scientists may be profoundly convinced of the necessity of such a fusion, the operation itself usually is a most arduous one. For though *the phenomena of nature are all consistent with each other*, we have to deal not only with these, but with the hypotheses which have been invented to systematize them. It by no means follows that because one set of observers have labored with all sincerity to reduce to order one group of phenomena, the hypotheses which they have formed will be **consistent** with those by which a second set of observers have explained a different set of phenomena.

---

<sup>29</sup> Maxwell; Quoted from Campbell and Garnett, *The Life of James Clerk Maxwell*, 125.

<sup>30</sup> Maxwell; Quoted from Campbell and Garnett, *The Life of James Clerk Maxwell*, 112.

Each science may appear tolerably consistent within itself, but before they can be combined into one, each must be stripped of the “daubing of untempered mortar” by which its parts have been prematurely made to cohere.<sup>31</sup>

This paper is not accidental for Maxwell. In other works Maxwell himself emphasized the value of the next principle (III) – “*cross-fertilization* of the sciences” evoking the image of bees pollinating flowers.<sup>32</sup>

The typical example of “the daubing of untempered mortar elimination” principle (IV) for Maxwell was the progress of mechanics in Newton’s time which consisted in getting rid of the celestial machinery with which generations of astronomers had encumbered the heavens, and thus “sweeping cobwebs off the sky.”<sup>33</sup>

## 2. Stages of Maxwellian Programme

*A Treatise on Electricity and Magnetism* (1873) was mainly an encyclopedia and a textbook; the basic electromagnetic results were obtained in a sequence of three papers: “On Faraday’s Lines of Force” (1855-56), “On Physical Lines of Force” (1861-1862) and “A Dynamical Theory of Electromagnetic Field” (1864).

The first paper (1855-56) is dedicated to elaboration of the “analogies” method founded on Kantian epistemology (see the last part for details). The method rejects the “ontological” approaches looking for the “essences” of electrical and magnetic phenomena and proclaiming that “in reality” electricity and magnetism are “fields” and not “action at a distance” phenomena, or vice versa. Maxwell’s proposal is to consider Faraday’s lines of force as a kind of tubes filled with ideal incompressible fluid.

I propose then, [...] ; and lastly to show how by an extension of these methods, and the introduction of another idea due to Faraday, the laws of the attractions and inductive actions of magnets and currents may be clearly conceived, without making assumptions as to the **physical nature** of electricity, or adding anything to that which has been already proved by experiment.<sup>34</sup>

It is crucial for a Kantian that the incompressible fluid has nothing to do with experimental reality. The constraints on the theory proposed consist in the

---

<sup>31</sup> James Maxwell, “Hermann Ludwig Ferdinand Helmholtz,” *Nature*, XV (1877). Reprinted in *The Scientific Papers of James Clerk Maxwell*, vol. 2 (New York: Dover, 1952), 592.

<sup>32</sup> Peter Harman, *The Natural Philosophy of J.C. Maxwell* (Cambridge: Cambridge University Press, 2001).

<sup>33</sup> James Maxwell, “On Action at a Distance” (1873). Reprinted in *The Scientific Papers of James Clerk Maxwell*, vol. 1 (New York: Dover, 1952), 315-20.

<sup>34</sup> James Maxwell, “On Faraday’s Lines of Force,” 159.

demand that the mathematical constructs should not contradict each other. In all the other matters the physical analogies method admits an unlimited freedom of imagination. Even the conservation laws can be broken down!

Maxwell stresses the generality of the lines of force paradigm, for it can account for any kind of force. For instance, it does not exclude the force of action at a distance which varies inversely as the square of the distance, as force of gravity or as observed electric and magnetic phenomena.

This is a significant remark which is probably intended to undermine possible objections that, in principle, the method excludes the dominant theory based on action at a distance.<sup>35</sup>

And in the other parts of the paper Maxwell renders the ways by which the idea of incompressible fluid motion can be applied to the sciences of statical electricity, permanent magnetism, magnetism of induction, and uniform galvanic currents. The core element of his innovations consisted in the construction of *neutral "language game"*<sup>36</sup> for description and comparison of the consequences from the rival theories.

Maxwell's "neutral language" was not Carnap's and Reichenbach's "observation language" springing out from the "protokolsatze" generalizations. Maxwell is aware of the theory-ladenness of the observation data ("experimental laws already established, which have generally been expressed in the *language of other hypotheses*."<sup>37</sup> He clearly understands that every observation always carries the footprints of the theoretical language that helps to describe it ("The daubing of untempered mortar," as he will call them later in his "Helmholtz" paper).

In order to compare and to unite in a theoretical scheme lacking contradictions all the results of the different experiments carrying the footprints of different theoretical languages, it is necessary to construct an *artificial* theoretical language equally distant from the languages of theories under comparison. This language appeared to be the language of solid state mechanics (with hydrodynamics as its part). Maxwell's ultimate aim was to rewrite all the known empirical and theoretical laws of electricity and magnetism using the

---

<sup>35</sup> Hon and Goldstein, "Maxwell's Contrived Analogy," 243.

<sup>36</sup> Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe (London: Basil Blackwell, 1953).

<sup>37</sup> James Maxwell, "On Physical Lines of Force," *Philosophical Magazine*, Series 4, 21 (1861-1862), 161-175, 281-291. *Philosophical Magazine*, Series 4, 23 (1861-1862), 12-24, 89-95. Reprinted in *The Scientific Papers of James Clerk Maxwell*, vol. 1 (New York: Dover, 1952), 162.

neutral language and then to compare them in order to create a system without contradictions.

The final result of the 1856 paper was a system of equations lacking the “displacement current.” It was not accidental that one of the main drawbacks of the incompressible fluid theory consisted in that the latter, apart from some simple cases, was unable to explain interrelations and interactions of electrical and magnetic fields and electric currents, as well as Faraday’s (1845) interconnection between optical and electromagnetic phenomena.

Maxwellian programme’s ultimate goal was to reveal the connection “between electricity at rest and current electricity” absent in the Ampere-Weber electrodynamics. Was it reached in 1856? – Certainly not. The connection between the current density  $\mathbf{j}$  and the charge density  $\rho$  was lacking in Maxwell’s initial 1856 scheme. It was to appear later, after the “displacement current” introduction and finding out its consequence – the continuity equation  $\text{div } \mathbf{j} + \partial\rho/\partial t = 0$ .

So, in 1861 the publication of Maxwell’s second paper consisting of four parts begins. Its aim was to rederive the results of Weber and Neumann theories on the basis of a new mechanical hypothesis containing the *vortices of incompressible fluid*. The theory started from W. Thomson’s investigations; he showed that the connection between magnetism and electricity has the same mathematical form as that between certain parts of phenomena, of which one has a *linear* and the other a *rotatory* character. It is important that W. Thomson introduced the vortices theory in incompressible fluid while reflecting on Faraday’s experiments on the rotation of the plane of polarized light when transmitted along the lines of magnetic force. So, they were the efforts to theoretical reconstruct the Faraday effect that provided the meeting of optics and theory of magnetism.

In the second Maxwellian theory the magnetic field was represented by a set of vortices in incompressible fluid with the axes of rotation coinciding with the direction of magnetic field at a point. But now a role of neutral language is played not by tube hydrodynamics but by a theory of stresses in the medium where the necessary relations among the forces are described by mathematicians with the help of mathematical entities that now are called *tensors*; the most general type of a tensor describing the most general type of stress consists of a combination of three principal pressures or tensions, in direction at right angles to each other. The tensor apparatus of solid state mechanics provided the creation of new neutral language “dialect;” it enabled to calculate the force upon an element of the medium:  $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4 + \mathbf{F}_5$ . The first term  $\mathbf{F}_1$  refers to the force acting on

magnetic poles; the second term  $F_2$  refers to the action on bodies capable of magnetism by induction; the third  $F_3$  and fourth  $F_4$  terms refer to the force acting on electric currents; the fifth term  $F_5$  refers to the effect of simple pressure that lacks an electromagnetic analogy.

But one of the most intricate problems of the vortices theory that puzzled even Daniel Bernoulli who invented it in XVIII-th century<sup>38</sup> was: how can the rotation be transferred from one vortex to another so that “vortices in a medium exist side by side, revolving in the same direction about parallel axis?” – The only conception that aided Maxwell in conceiving this kind of motion was that of the vortices being separated by a layer of particles called the “*idle wheels*.” Is it possible to connect these particles with electricity?

And in the second part of his 1861 paper “The Theory of Molecular Vortices applied to Electric Currents” Maxwell comes up to the *hardest problem* of his research programme: what is “the physical connexion of these vortices with electric currents, while we are still in doubt as to the *nature of electricity*.” It is this point where Maxwell has to admit the principal *limits* of pure mechanical theories and to **borrow the elements of action at a distance theory!** Or, using Nugayev’s methodological language, one can conclude that Maxwell had to construct the “*crossbred theoretical objects*” from the languages of both cross-theories that combine the properties of quite different theoretical schemes.

According to Maxwell’s theory, an electric current is represented by the transference of the moveable particles interposed between the neighboring vortices. As a result, these particles, in his theory, *play the part of electricity*. Their motion of translation constitute an electric current and their rotation serves to transmit the motion of the vortices from one part of the field to another. The tangential pressures thus called into play constitute electromotive force. Nevertheless,

the conception of a particle having its motion connected with that of a vortex by perfect rolling contact may appear somewhat *awkward*. I do not bring it forward as a mode of connexion existing in nature.<sup>39</sup>

On introducing such abstract objects as “electrical particles” and “electric current representing the motion of such particles” Maxwell had deviated significantly from Faraday’s notions. According to Michael Faraday, the electrical charges should be considered as created by the ends of lines of force; they lack an

---

<sup>38</sup> Edmund Whittaker, *A History of the Theories of Aether and Electricity: From the Age of Descartes to the Close of the Nineteenth Century* (London, New York: Longmans, Green and Co, 1910).

<sup>39</sup> Maxwell, “On Physical Lines of Force,” 345.



independent substantial existence. Correspondingly, in his *genuine* research programme the electric current has to be considered not as the motion of real particles but as an “energy axis.”

This is the nub of the British field programme: the fields are primary, and the particles are only secondary. Later on Maxwell’s eclecticism was followed by H.A. Lorentz’s dualism. Lorentz initiated it in a 1875 paper:

I shall start with instantaneous action at a distance: thus we will be able to found the theory on the most direct interpretation of observed facts.<sup>40</sup>

So it was not a temporary retreat. Even after 1861 Maxwell many times introduced the notions of the Ampere-Weber atomism into his theories.<sup>41</sup>

Yet the results obtained were of course insufficient; the theoretical derivation of Coulomb’s law was lacking. Namely that was done in the third part of 1861-1862 paper “The Theory of Molecular Vortices applied to Statical Electricity.” It is important that the vortices theory contained too many ad hoc assumptions. To eliminate at least some of them “*Maxwell’s miracle*” was invented. It appeared that if one transposes in the course of Fresnel optics and electromagnetism theory meeting the ether properties from optics to electromagnetism, he can eliminate at least one ad hoc supposition.

Indeed, it is necessary to suppose, in order to account for the transmission of rotation from the exterior to the interior parts of each cell, that the substance in the cells possesses *elasticity of figure*, similar in kind, though different in degree, to that of observed in solid bodies. The undulatory theory of light requires one to admit this kind of elasticity in the luminiferous medium, in order to account for transverse vibrations. Hence he need not then be surprised if the magneto-electric medium possesses the same properties. This peculiarity has a vital significance for Maxwell’s neutral language. If we can now explain the condition of a body with respect to the surrounding medium when it is said to be ‘charged’ with electricity, and account for the forces acting between electrified bodies, we shall have established a connexion between *all* the principal phenomena of electrical science. Thus, the extrapolation of the molecular vortices theory on the electrostatic domain became possible due to the elasticity of the vortices that enabled the medium to maintain the elasticity waves. As a result,

the velocity of transverse undulations in our hypothetical medium, calculated from the electromagnetic experiments of M.M. Kohlrausch and W. Weber, agrees so exactly with the velocity of light calculated from the optical

---

<sup>40</sup> Quoted from Darrigol, *Electrodynamics from Ampere to Einstein*, 323.

<sup>41</sup> Olivier Darrigol, *Electrodynamics from Ampere to Einstein*.

experiments of M. Fizeau, that we can scarcely avoid the inference that light consists of the same medium which is the cause of electric and magnetic phenomena.<sup>42</sup>

The displacement current introduction was due to Maxwell's efforts to link the equations relating to electrical current with that of electrostatics, which demanded the Ampere law modification for the sake of a new term introduction; the term had to describe the elasticity of the vortices medium. The displacement current introduction driving force came from Maxwell's efforts to unify all the main empirical laws belonging not only to electricity and magnetism but to optics as well.

As a result Maxwell obtained his famous system of equations along with the continuity equation describing that electrical particles that transform the rotations from one vortex to another does not appear from nothing and cannot disappear to nowhere. But one could not state any final unification of optics and electromagnetism in 1861. It was possible to tell only on the *beginning* of their reconciliation, on the beginning of rather different theoretical ontologies "grinding."

And at last in 1864 Maxwell proposed a modified version of his 1861-1862 paper that tried to avoid any special suppositions on the nature of molecular vortices. In his 1864 paper Maxwell derives his equations from abstract dynamics of Lagrange. The Lagrangian function  $L$  is found as the difference between the kinetic and potential energies of a system. From those he was able to derive the basic wave equation of electromagnetism without any special assumptions about molecular vortices or forces between electrical particles. Although displacement retained a prominent position in "A Dynamical Theory of Electromagnetic Field," its role was rather different from the role it played in 1861-1862 paper. It was no longer associated with changes in positions of rolling particles; rather, Maxwell defined it simply as the motion of electricity, that is, in terms of a quantity of charge crossing a designated area.

However, despite Maxwell's claim to provide deductions from (three) experimental facts, his account still required the **postulation** of a displacement current, something that could neither be verified by nor deduced from experiment.<sup>43</sup>

And at last Maxwell's creativity ends with *A Treatise on Electricity and Magnetism* conceived as an encyclopedia of the electrical and magnetic effects. In his *Treatise* Maxwell goes further in purifying his deductions from the model

---

<sup>42</sup> Maxwell, "On Physical Lines of Force," 22.

<sup>43</sup> Morrison, *Unifying Scientific Theories*; Darrigol, *Electrodynamics from Ampere to Einstein*.

remnants and in strengthening the Lagrangian approach. In the final chapter XX, dedicated to the electromagnetic theory of light, the basic argument in defense of electromagnetic waves is posited out:

To fill all space with a new medium whenever any new phenomena is to be examined is by no means philosophical, but if the study of two *different* branches of science has independently suggested the idea of a medium, and if the properties which must be attributed to the medium in order to account for electromagnetic phenomena are of the same kind as those we attribute to the luminiferous medium in order to account for the phenomena of light, the evidence for the physical existence of the medium will be considerably strengthened.<sup>44</sup>

Yet it is important that in his *Treatise* Maxwell faced with the same problem as in 1864 paper: the problem of Lagrangian mathematical formalism application to the case of electromagnetic field. Maxwell himself used a fitting comparison with a belfry. He aimed to develop a Lagrangian formulation of electromagnetism in which the ether mechanism would be the analogue of the mechanism in the belfry, whilst the positions and velocities of the ropes would have their analogues in measurable charge and current distributions serving to determine the electromagnetic energy.

However on twenty pages of his *Treatise*' chapter Maxwell gave a detailed Lagrangian treatment for interacting closed conduction currents only. And when, two chapters later, he came to build on his Lagrangian formulation to formulate the *general* equations of his electromagnetic theory, he *simply added* the displacement to the conduction current "by hands" to give the total current.

But this move by Maxwell in fact undermined the major attraction of his Lagrangian method.<sup>45</sup> The first direct experimental evidence for the existence of displacement currents emerged only with Hertz's experiments culminating in production of radio waves in 1888. As always, the Lagrangian formulations were retroactive attempts to accommodate results obtained by other means.

But let me return to Maxwell's synthetic programme. Eventually Maxwell found that his elastic vortex medium *would* propagate waves whose velocity, calculated from electromagnetic constants, was that of light. Yet he said nothing about *how* electromagnetic waves might be generated, nor did he attempt to derive the laws governing reflection and refraction. Hence the task of extracting a cogent theory from the *Treatise* and of casting it into a form in which it could

---

<sup>44</sup> Maxwell, *A Treatise on Electricity and Magnetism*, 781.

<sup>45</sup> Alan Chalmers, "Maxwell, Mechanism and the Nature of Electricity," *Physics in Perspective* 3, 4 (2001): 425-38.

command general assent fell to others. Later they were called “the Maxwellians:” George Francis Fitzgerald (1851-1901), Sir Oliver Lodge (1851-1940), and Oliver Heaviside (1850-1925).

Of their advances one should mention the Bath meeting where the Maxwellians made clear that Maxwell’s displacement current was not just a dispensable appendage to the theory, but its keystone: remove it, and the whole theoretical structure would collapse. Without displacement currents, electromagnetic waves could not exist.

But the most important step in consequent optics and electromagnetism unification, i.e. in electrodynamics principles extrapolation on optical phenomena was made in 1879 by Francis Fitzgerald. He first broached the possibility of combining Maxwell’s theory with Mac Cullagh’s. In 1839 James Mac Cullagh had devised a Hamiltonian formulation of wave optics which yielded equations describing the main optical phenomena, including reflection, refraction and double refraction. Fitzgerald, by drawing correspondences between the terms in Mac Cullagh’s theory and electromagnetic terms, was able, in 1879, to translate Mac Cullagh’s theory into an electromagnetic theory of light. It should be noted, however, that Mac Cullagh’s theory suffered from serious mechanical difficulties, pointed out in 1862 by Gabriel Stokes. Stokes showed that Mac Cullagh’s theory implied attributing elastic properties to the ether which were quite unlike those of any known substances.

The merger not only resuscitated Mac Cullagh’s theory but extended Maxwell’s own theory in important new directions, yielding as one of its first fruits a prize that had eluded Maxwell himself: an electromagnetic theory of the reflection and refraction of light.

In his last scientific work – in a review of George Fitzgerald’s paper (1879) – Maxwell described his own treatment of the Faraday effect in 1862 paper as a “**hybrid**” one in which he had combined his electromagnetic theory of light with elements of an elastic solid theory. He had treated light waves as actual motions of the ether and had traced how these would disturb the spinning of the magnetic vortices in such a way as to cause the plane of polarization of the light to rotate. Maxwell had found this detour into a “hybrid theory,” in which electrical and mechanical actions were combined, the least satisfactory part of his own explanation of the Faraday effect.<sup>46</sup>

And Fitzgerald’s 1879 paper brought out, more clearly than before, the fundamental incompatibility between Maxwell’s theory and an elastic ether. Fitzgerald had shown that Maxwell’s theory was mathematically equivalent to

---

<sup>46</sup> Bruce Hunt, *The Maxwellians* (Ithaca: Cornell University Press, 2005), 18.

Mac Cullagh's, while Stokes had shown in 1862 that Mac Cullagh's theory, considered as an elastic solid theory, was untenable. The following conclusion was inescapable: if Maxwell's theory were to survive, it had to be cut loose from reliance on an elastic solid ether and given a new basis. Attempts to produce a 'hybrid' theory, such as Maxwell had pursued in his own account of the Faraday effect, had to be abandoned.<sup>47</sup>

Thus, in his encyclopedia on the phenomena of electricity and magnetism Maxwell sums his results up. His Copernican deeds consisted in combining arguments for electromagnetic and luminiferous ethers' identification and constructing the crossbred theory with displacement current that was capable of electromagnetism and optics unification.

Nicolas Copernicus had pioneered in considering the Earth as an ordinary planet orbiting the Sun; hence he had created a crossbred theoretical object capable of extrapolating the mathematical principles from divine phenomena on the mundane ones. On the other hand, through the same crossbred object the physical principles were extrapolated from mundane objects on the skies.<sup>48</sup>

Similarly, James Maxwell had constructed a crossbred object – the displacement current – and was able to extrapolate the electromagnetic principles on the optical phenomena, and vice versa. Introducing a kind of “*complementarity principle*” in the XXIII chapter called “Theories of Action at a Distance,” Maxwell comes to the following conclusion. We are ignorant of what is really moving between magnets and conductors, but if we decide to describe it we have no other “appropriate” images except “waves” and “particles.”<sup>49</sup>

### 3. Maxwellian Electrodynamics in Germany: Helmholtz and Hertz

Due to Kantian background, Maxwell's programme development should be especially fruitful in Germany. And it was. Maxwell's efforts to find a reasonable compromise between the three research programmes (that of Young-Fresnel, Faraday and Ampere-Weber) were picked up by Maxwell's friend Hermann Helmholtz. He had sought from the middle of 1860-s to reach consensus between the major directions in electromagnetic research of the second half of the nineteenth century, namely, Newton's instantaneous action-at-a-distance concept as used by Weber, and Faraday's contact action concept. In Helmholtz's paradigm charges and currents were considered as the sources of electrical and magnetic

---

<sup>47</sup> Hunt, *The Maxwellians*.

<sup>48</sup> Rinat Nugayev, "The Ptolemy-Copernicus Transition: Intertheoretic Context," *Almagest* 4, 1 (2013): 96-119.

<sup>49</sup> Maxwell, *A Treatise on Electricity and Magnetism*, 488.

fields. It led directly to H.A. Lorentz's dualistic picture of the field equations and the equations of motion in his 1892-1900 papers.

By the time of Helmholtz's first attempt of reconciliation (1870), the research programmes of Weber and Faraday had successfully incorporated all well-established empirical facts. Hence when trying to arrive at results similar to Maxwell's without losing the elements of action at a distance, Helmholtz assumed that the electrostatic forces are constantly present as a field in space and that the change in the polarization or the displacement of the charges signaled the change in the electrostatic field.<sup>50</sup> Under these assumptions, Helmholtz in his 1870 paper successfully derived generalized equations very similar to those of Maxwell and found that in a limited case they yield equations identical to Maxwell's. Yet in addition to the ordinary transverse electromagnetic waves, Helmholtz discovered the existence of longitudinal electric waves which turned to be instantaneous at the Maxwell's limit  $k = 0$ .

To check the consequences from his theory in 1879 Helmholtz proposed a prize competition "to establish experimentally a relation between electromagnetic action and the polarization of dielectrics" and persuaded one of his pupils whose name was Heinrich Hertz to take part.

And in 1886-1888, at Karlsruhe, Hertz attempted to establish the compatibility of the theories of Helmholtz and Maxwell in a new series of experiments. He designed his measurement procedures, taking into account Helmholtz's ingenious separation of the total electric force into the electrostatic and electrodynamic parts to which different velocities of propagation were ascribed. In his own words. According to Coulomb's law, the electrostatic component was thought to be proportional to the inverse square of the distance, whereas the electrodynamic part was only proportional to the inverse of the distance. In the usual theory of the Lienard-Wiechert potential it would correspond to decreasing rates of the bound-field, or longitudinal, component and the radiation field, or transverse component, respectively.

Hertz's experiments were carried out within Helmholtz's research programme. According to Hertz,

Notwithstanding the greatest admiration for Maxwell's mathematical conceptions, I have not always felt quite certain of having grasped the physical significance of his statements. Hence it was not possible for me to be guided in my experiments directly by Maxwell's book. I have rather been guided by

---

<sup>50</sup> Hermann Helmholtz, *Wissenschaftliche Abhandlungen*, vol.1 (Leipzig: J.A. Barth, 1882), 611-628.

## Maxwellian Scientific Revolution: A Case Study in Kantian Epistemology

Helmholtz's work, *as indeed may plainly be seen from the manner in which the experiments are set forth*.<sup>51</sup>

Hertz had planned a series of experiments and his efforts appeared to be fruitful. Yet it should be noted that the title of his 1888 paper "On the Finite Velocity of Propagation of Electromagnetic Action" is perhaps misleading nowadays, because usual Maxwellian electrodynamics does not employ the Helmholtzian "action" terminology, nor does it split the total electric force into electrodynamic and electrostatic parts. But for Hertz's contemporaries who supported the Helmholtz theory, the underlying meaning of the presented results was clear enough: Hertz's experiments could qualitatively conclude about the finite propagation of the electromagnetic part, but could say nothing definite about the electrostatic component

According to one of modern action at a distance devotees,<sup>52</sup> some of Hertz's measurements tended towards the instantaneous nature of the electrostatic modes. Yet he was still not convinced of this instantaneity and preferred to be cautious.

Furthermore it was Hermann Helmholtz who convinced Berlin Academy of Science to set up a special prize for experimental confirmation of Maxwell's theory. And it was Helmholtz's pupil Heinrich Hertz who got the prize in 1888. From two possible explanations of his experiments,<sup>53</sup> Hertz (1889) had chosen the simplest one:

Helmholtz distinguishes between two forms of electric force – the electromagnetic and the electrostatic – to which, until the contrary is proved by experience, two different velocities are attributed. An interpretation of the experiments from this point of view could certainly not be incorrect, but it might perhaps be *unnecessary complicated*.<sup>54</sup>

It seems to me that it was namely the attempt to justify the rationality of choosing the simplest explanation that appeared one of the reasons to force Heinrich Hertz after 1888 to give up his electromagnetic experiments fruitful both from heuristic and technological vistas and to devote the last three years of his life to his extremely ambitious project of classical mechanics rebuilding.

---

<sup>51</sup> Hertz, *Electric Waves*, 20.

<sup>52</sup> Roman Smirnov-Rueda, "Were Hertz's 'Crucial Experiments' on Propagation of Electromagnetic Interaction Conclusive?" in *Instantaneous Action at a Distance in Modern Physics: Pro and Contra*, eds. Andrew E. Chubykalo, Viv Pope, and Roman Smirnov-Rueda (New York: Nova Science Publishers, 2001), 57-69.

<sup>53</sup> Smirnov-Rueda "Were Hertz's 'Crucial Experiments.'"

<sup>54</sup> Hertz, *Electric Waves*, 123.

In his *Principles of Mechanics* he put it clear that it is premature to attempt to base the equations of motion of the ether upon the laws of mechanics until we have obtained a perfect agreement as to what is understood by this name.<sup>55</sup>

It is important that the methodological principles for classical mechanics rebuilding were to be found by Hertz in Kantian epistemology; even before he met Helmholtz, Hertz had attended in Dresden a course on Kantian philosophy.

We form for ourselves images or *symbols* of external objects [...] When from our accumulated previous experience we have once succeeded in deducing images of the desired nature, we can then in a short time develop by means of them, as by means of *models*, the consequences which in the external world only arise in a comparatively long time, or as the result of our own interposition [...] *The images* which we have speak of are *our* conceptions of things. With the *things themselves* they are in conformity in one important respect, namely, in satisfying the above-mentioned requirement. For our purpose it is not necessary that they should be in conformity with the things in any other respect whatever [...]. The images which we may form of things are not determined without ambiguity by the requirement that the consequents of the images must be the images of the consequents. Various images of the same objects are possible, and these images may differ in various respects. We should at once denote as inadmissible all images which implicitly contradict the laws of our thought.<sup>56</sup>

As a result, scrupulous analysis of the simplicity criterion ends by the following conclusion:

A doubt which makes an impression on our mind cannot be removed by calling it metaphysical; every thoughtful mind as such has needs which scientific men are accustomed to denote as metaphysical [...] It is true we cannot a priori demand from nature simplicity, nor can we judge what in her opinion is simple. But with regard to images of our own creation we can lay down requirements. We are justified in deciding that if our images are well adapted to the things, the actual relations of the things must be represented by simple relations between the images.<sup>57</sup>

Hertz's Kantian background manifested itself not only in the epistemological scheme described. According to Jed Z. Buchwald,<sup>58</sup> already in 1884 Hertz had proposed a version of Maxwell's equations that was free of the ether notion completely.

---

<sup>55</sup> Heinrich Hertz, *The Principles of Mechanics Presented in a New Form*, trans. D.E. Jones (London: Macmillan, 1899), XXI.

<sup>56</sup> Hertz, *The Principles of Mechanics*, 1.

<sup>57</sup> Hertz, *The Principles of Mechanics*, 23.

<sup>58</sup> Buchwald, *The Creation of Scientific Effects*, 278.



And, what is more important, quite unlikely Maxwellian field theory, in Hertz's theoretical scheme the source continued to exist as an entity in and of itself. In Hertz's diagram the material object remains unknown, whereas the inferred field is known. This diagrammatic inversion encapsulates the originality of Hertz's physics. It was because Hertz ignored the physical character of the object that produced his radiation – because he boxed it in with a mental quarantine against asking questions against it – he was able to make progress.<sup>59</sup>

Being a pupil of Helmholtz, Hertz learned to watch for novel interactions between laboratory objects without worrying overmuch about the hidden processes that account for the object's effect-producing power.

Thus the nature of electromagnetic waves appeared to Hertz as a kind of "thing in itself" that admits a variety of interpretations. Researcher chooses the version that is the simplest one to work with. The most important thing is the equations depicting the relations between the objects under investigation.

---

<sup>59</sup> Buchwald, *The Creation of Scientific Effects*, 272.



# A HUMEAN ACCOUNT OF TESTIMONIAL JUSTIFICATION

Shane RYAN

**ABSTRACT:** I argue that a Humean account can make sense of the phenomenology associated with testimonial justification; the phenomenology being that in standard cases hearers regularly simply accept a testifier's assertions as true – hearers don't engage in monitoring. The upshot is that a Humean account is in a better position dialectically than is usually supposed. I provide some background to the debate before setting out two challenges facing accounts of testimonial justification. The first challenge is to provide an account that accords with the phenomenology of testimonial reception; the second challenge is to provide an account that can make sense of some testimonial beliefs enjoying greater justification than others. I show the credulist position to be vulnerable to the second challenge and the Humean position to be vulnerable to the first challenge. I argue that a Humean account, by drawing on dual process theory, can overcome the first challenge.

**KEYWORDS:** David Hume, testimony, epistemic justification

## Introduction

I take testimony to be "... the assertion of a declarative sentence by a speaker to a hearer or to an audience."<sup>1</sup> I regard this as an approximation of what testimony is, but one that suffices for my purposes.<sup>2</sup> I further take it that not only is testimonial knowledge possible but that we are dependent on testimony for much of our knowledge. Following Adler, I describe this as the Far-Reaching Dependence Thesis.<sup>3</sup>

---

<sup>1</sup> Jonathan Adler, "Epistemological Problems of Testimony," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Winter 2010 Edition), URL = <<http://plato.stanford.edu/archives/win2010/entries/testimony-episprob/>>.

<sup>2</sup> Nevertheless it should be noted that testimony needn't be in the form of spoken word and "audience" here shouldn't be thought of as restricted to hearers. Testimony can, for example, be given in written form and sign language, and, might, be given by bodily movements with communicative potential such as pointing, winking, etc., in the appropriate contexts as Lackey argues. Jennifer Lackey, *Learning from Words: Testimony as a Source of Knowledge* (Oxford: Oxford University Press, 2008), *Oxford Scholarship Online*, Accessed March 17<sup>th</sup> 2011, <<http://dx.doi.org/10.1093/acprof:oso/9780199219162.001.0001>>, 25-26.

<sup>3</sup> Adler, "Epistemological Problems of Testimony."

At first glance, the mere idea that we can gain knowledge from testimony might seem implausible, let alone the claim that we are dependent on testimony for much of our knowledge. A Cartesian way of thinking about knowledge lends itself to this thought. To elaborate, it might be thought that if even our perceptual beliefs are vulnerable to sceptical worries, then our beliefs dependent on the assertions of others are all the more vulnerable to sceptical worries. Knowledge, for the Cartesian, is reason based conviction which there can be no alternative reason for doubting; as doubts are removed, certainty increases.<sup>4</sup> On this way of thinking of knowledge, it's very difficult to see how one can gain true beliefs with knowledge conducive epistemic justification from standard cases of testimony. If one attains knowledge, one's knowledge will be equipped to withstand any sceptical challenge; if one knows, one will be certain.<sup>5</sup>

The approach commonly taken in contemporary epistemology to the examination of the nature of knowledge is notably distinct from the methodology of doubt employed by Descartes for the same purpose. First, fallibilist accounts of knowledge are widely accepted within the field of epistemology, fallibilist in that we can know *p* even if our belief that *p* could, in a certain sense, have been wrong but isn't. More specifically, the justification of our belief may be such that it is possible that we could be wrong.<sup>6</sup> Second, and relatedly, it's standard practice when examining the structure of knowledge in contemporary epistemology to assume that we do know more or less what we ordinarily take ourselves to know. Third, it's also standard practice to extrapolate from such cases of knowledge as well as hypothetical cases in which we would take ourselves to know, to build an account of the nature of knowledge.<sup>7</sup>

---

<sup>4</sup> Lex Newman, "Descartes' Epistemology," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Fall 2010 Edition), URL = <<http://plato.stanford.edu/archives/fall2010/entries/descartes-epistemology/>>.

<sup>5</sup> For an early treatment of how testimonial belief fits in with certainty and what we would now call entitlement, see the writings of the much overlooked John Henry Newman, *An Essay in Aid of a Grammar of Assent* (London: Elibron Classics, 2007). (Anthony Kenny, *Philosophy in the Modern World: A New History of Western Philosophy, vol. 4* (Oxford: Oxford University Press, 2007), 30, in his history of Western philosophy, writes that Newman's is "... arguably the best treatment of the topics of belief and certainty between Hume and Wittgenstein.")

<sup>6</sup> Duncan Pritchard, *What is This Thing Called Knowledge?* (London and New York: Routledge, 2006).

<sup>7</sup> Relevant to this discussion is Roderick Chisholm, *The Problem of the Criterion* (Milwaukee: Marquette University Press, 1973), in which it is argued to be problematic to build an account of knowledge both from what we take the extension of knowledge to be without already drawing on what we take the intension of knowledge to be and similarly to start from what we take the

This is not to say that there is a denial of the sceptical problem, rather responding to the sceptic is treated as an enterprise distinct from that of providing an account of the nature of knowledge. C. A. J. Coady for example distinguishes what he calls positive epistemology from negative epistemology.<sup>8</sup> The former investigates the structure of knowledge or the body of beliefs that can reasonably be thought of as knowledge, and leaves sceptical worries aside; while the latter is concerned with the theoretical problems raised by scepticism.<sup>9</sup> Greco, in like vein, distinguishes what he calls “the project of explanation,” which seeks to explain “what knowledge is and how knowledge is possible,” and “the project of vindication,” which is the project “of showing that we have knowledge.”<sup>10</sup>

Claiming that what follows falls within the realm of positive epistemology or the project of explanation doesn't quite yet allow us to set sceptical worries with regard to testimonial knowledge to one side. After all, it might be thought that much of what we take ourselves to know doesn't include, or at least not to any significant degree, testimonial beliefs in propositions. But given a little consideration this will surely seem mistaken.

My knowledge that Mount Everest is the highest mountain in the world, that smoking can cause cancer, that there exist other planets in our solar system, that the French Revolution occurred, are all based on the testimony of others.<sup>11</sup> Not only is much of my knowledge of the broader world and its past based on testimony, but much of my knowledge of important facts about my own life and the lives of those around me is also based on testimony; my knowledge, for example, of my date of birth, the occupations of numerous friends and family members, and that, say, two of my friends holidayed in Italy last year, are each based on testimony. I take it that most of us are in a similar epistemic position and that therefore the much of what we take ourselves to know includes knowledge that comes from testimony. These examples suggest not only that much of what

---

intension of knowledge to be without already drawing on what we take the extension of knowledge to be.

<sup>8</sup> C.A.J. Coady, *Testimony: A Philosophical Study* (Oxford: Oxford University Press, 1992), 3.

<sup>9</sup> Confusingly, “positive epistemology” and “negative epistemology” are used in a different sense in some of the literature. For example, an alternative use of “positive epistemology” is one that describes an epistemology that assumes the negation of sceptical premises and considers why the original premises should be denied. William G. Lycan, “Moore Against The New Sceptics,” *Philosophical Studies* 103 (2001): 44.

<sup>10</sup> John Greco, *Achieving Knowledge* (Cambridge: Cambridge University Press, 2010), 5.

<sup>11</sup> History as a subject is particularly dependent on testimony. Testimony seems to be the basic raw material or data of history as a discipline.

Shane Ryan

we take ourselves to know includes some testimonial knowledge but that the Far-Reaching Dependence Thesis is true.

So far I have acknowledged that there might be resistance to treating testimonial knowledge as possible. I have explained that I will follow the contemporary epistemological approach of setting sceptical worries to one side and assume that we do know much of what we take ourselves to know. I have also countered the worry that, even if the contemporary epistemological approach is accepted, it might be objected that testimonial knowledge does not account for a significant portion of what we ordinarily take ourselves to know.

### **1. Testimonial Justification: Two Problems**

I take an account of testimonial justification to face two challenges. The challenges, set out in question form, are as follows:

1. How can we account for testimonial justification in a way that accords with the phenomenology of testimonial reception?
2. What makes some testimonial beliefs more justified than other testimonial beliefs?

Relevant to addressing the first challenge is what Adler calls the Uniformity Claim; in standard cases hearers regularly simply accept a testifier's assertions as true.<sup>12</sup> I take this to be a claim about the phenomenology of testimonial reception as well as being an empirical claim; in other words I take the Uniformity Claim as also describing how testimonial reception is commonly experienced. A further claim that I take to be relevant to addressing the first question is that the testimony believed in such cases is ordinarily justified. The motivation for the latter claim is that ultimately we want an account of how testimonial justification accords with the phenomenology of testimonial reception in such a way that we can explain how the Far-Reaching Dependence Thesis is the case. If both claims are indeed true then this would suggest that testimony carries its own justification; or, to put it another way, the fact that something is testified to ordinarily makes believing what has been testified justified, rather than testimony being justified at the expense of some reasoning process, say inference, on the part of the believer. Just as something visually seeming to be the case ordinarily makes believing what seems to be the case justified, so too with testimony; testimony that p ordinarily makes believing p justified.

---

<sup>12</sup> Adler, "Epistemological Problems of Testimony."

It does, however, seem intuitive to think that one recipient of testimony may be justified in her testimonial belief, while a fellow recipient may be less justified in believing what is said. To see this, consider the following case:

Jana tells her mother, who is a wise old judge, and her young child a fantastic tale of how, in the space of a fortnight, she lost most of her fortune, say when the stocks she had invested in crashed, and then regained a roughly equivalent fortune, say when other stocks she had invested in soared. Although her tale is fantastic, both the judge and the child believe her tale, the content of which is known by Jana and so is in fact true.

Even if they both believe the testimony, we should want an account of testimonial justification to allow for the possibility that the judge is more justified in her belief than the child is in his belief. Such a consideration suggests that the degree of justification enjoyed by a testimonial belief may vary from person to person.<sup>13</sup>

## 2. The Credulist Response

The works of Thomas Reid and David Hume have inspired rival contemporary accounts of testimonial justification. The Reidian or credulist view is that the Almighty intended us to be social beings and so made us with a “propensity to speak the truth” and with a disposition to believe what we are told by others.<sup>14</sup> Given such a propensity and disposition, testimony is likely to be believed and testimony believed is likely to be true, making testimonial belief justified.

The contemporary credulist argument doesn’t make reference to God, but claims that adherence to the norm of assertion makes testimonial claims justified. Accounts claiming to articulate the norm of assertion set out the conditions in which it is regarded as appropriate to assert something. One account of the norm of assertion, is that the speaker can properly assert *p*, only if the speaker knows *p*.<sup>15</sup> If this is right, then it looks like we’re on our way to answering the question as to what makes testimonial belief justified in a way that fits with the Uniformity Claim.

---

<sup>13</sup> Similarly, we can imagine that the judge’s testimonial belief may be more justified in one situation and less so in one in which she has less experience and is less wise.

<sup>14</sup> Thomas Reid, *Inquiry and Essays*, eds. Ronald E. Beanblossom and Keith Lehrer (Indiannapolis: Hackett, 1983), 94-95.

<sup>15</sup> From Adler, “Epistemological Problems of Testimony,” who cites Peter Unger, *Ignorance* (Oxford: Oxford University Press 1975) and Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000).

If there is widespread adherence to the knowledge norm of assertion, then if someone asserts *p*, then an audience is justified in believing *p*. Now we look to be in a good position to explain the Far-Reaching Dependence Thesis while continuing to hold the Uniformity Claim. Simply believing what is said won't prevent the recipient of testimony from having a justified belief.

Now, however, we don't seem to be in a good position to answer the second question. Returning to the case of the judge and the child, if testimonial belief is justified because of the norm of assertion then it's not obvious how the judge might be more justified than the child. Relatedly, claiming that knowledge may be gained simply by believing what's testified seems to set the bar too low for knowledge. Elizabeth Fricker puts the worry about accounts of testimonial justification based on such a claim starkly when she writes that they are "... an epistemic charter for the gullible and indiscriminating."<sup>16</sup> Fricker recommends that

... a hearer should always engage in some assessment of the speaker for trustworthiness. To believe what is asserted without doing so is to believe blindly, uncritically. This is gullibility.<sup>17, 18</sup>

The credulist may warn that requiring such an assessment of testimonial justification discords with the Uniformity Claim, and risks ultimately leaving us unable to explain the Far-Reaching Dependence Thesis. The credulist may claim that without foregoing an explanation of the Far-Reaching Dependence Thesis, the credulist can adjust her account in an attempt to address the concerns raised. The credulist may do so by claiming that there should be "counterfactual sensitivity" to possible defeaters.<sup>19</sup> If I ask for directions to the Sears Tower, now officially known as the Willis Tower, and would have believed the testifier, had he, gleam in eye, unsuccessfully attempted to muffle sniggers while giving directions, then I would be insensitive to what such behaviour normally indicates and not justified in my testimonial belief.

The natural question to ask is whether this move, which claims that counterfactual sensitivity rather than assessment is required, does not too go too far and discords with the uniformity principle. Miranda Fricker criticises this move on the basis that we are left with no explanation of how an agent would go

---

<sup>16</sup> Elizabeth Fricker, "Against Gullibility," in *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, eds. By Bimal K. Matilal and Arindam Chakrabarti (Dordrecht: Kluwer, 1994), 126.

<sup>17</sup> Fricker, "Against Gullibility," 145.

<sup>18</sup> Both quotes are cited in Adler, "Epistemological Problems of Testimony."

<sup>19</sup> Adler, "Epistemological Problems of Testimony."



from simply accepting testimony to, when relevant, critically attending to justification relevant aspects of the testimony.<sup>20</sup> In order for there to be counterfactual sensitivity the worry is that there needs to be a kind of continuous assessment or monitoring; the thought being that only this can explain the possibility of noting a defeater and responding accordingly. But the credulist needn't be committed to such continuous assessment or monitoring in order for there to be counterfactual sensitivity. For example, a person may notice that another person's eyebrow has been shaved without requiring of the first person that he always be engaged in monitoring the eyebrows of the other person.

The picture of testimonial justification that emerges from this move is of testimonial justification being more complex than the original credulist idea of testimony carrying its own justification. Rather the picture we get is that testimony enjoys default or prima facie justification but that believing testimony alone isn't sufficient for justification, at least for knowledge conducive justification the agent's belief must have the outlined counterfactual sensitivity.

Even with these significant adjustments to the credulist account of testimonial justification with which we started out, it's not obvious that we are in a position to adequately address the second challenge of accounting for differing degrees of justification of testimonial belief, although we can account for one person having testimonial justification from a piece of testimony and another not having testimonial justification from the same piece of testimony. The addition of a simple counterfactual sensitivity condition does not seem sufficiently dynamic to capture the way in which we'd imagine there to be a wide range of factors of differing weights adding to and detracting from the degree of appropriate justification a belief in a piece of testimony enjoys. Putting the point differently, we're not yet better placed to say how the wise old judge might be justified in her testimonial belief to the  $n^{\text{th}}$  degree while another recipient of testimony may be less justified in his belief in the same piece of testimony.

### 3. The Humean Response

Hume also holds that much of our knowledge comes about via testimony.<sup>21, 22</sup> In "On Miracles," he writes that belief in testimony is not due to an a priori

---

<sup>20</sup> Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007), 65-66.

<sup>21</sup> David Hume, *An Enquiry Concerning Human Understanding*, ed. Tom L. Beauchamp, (Oxford and New York: Oxford University Press, 1999), 170.

<sup>22</sup> Paul Faulkner, "David Hume's Reductionist Epistemology of Testimony," *Pacific Philosophical Quarterly* 79 (1998): 302-313, Duncan Pritchard, "The Epistemology of Testimony,"

connection “between testimony and reality” but is based on “the usual conformity of facts to the reports of witnesses.”<sup>23</sup> There is no necessary connection between what is testified to and how the world is, but as it happens what is testified to usually does reflect how the world is. The level of justification that testimony enjoys is determined inductively; the more regularly facts and witness reports conform, the greater the justificatory weight of the testimony.

Hume doesn't want to just say that induction shows there to usually be conformity between testimony and reality and that therefore we can just take testimony to be justified; rather he discusses considerations that have a bearing on the justificatory force of testimony. A benefit of this more fine-grained approach is that if we are justified in believing competing claims, two or more claims that can't both or all be true, then we may be able to determine which of the claims is of greater justificatory force; something that we would be unable to do if we supposed testimony to be justified only on the basis of it being testimony. Similarly, if we have testimonial justification for believing *p* and we have non-testimonial justification for believing not *p*, then a more fine-grained account of testimonial justification may allow us to determine whether *p* or not *p* enjoys greater justification.

The considerations that Hume regards as significant when determining the justificatory force of testimony include; the presence of contrary testimony; the character of the testifiers; the number of testifiers; the manner in which the testimony is delivered; and the interests the testifiers have in the testimony being affirmed.<sup>24</sup> Another factor is the extraordinariness of the testimony for the recipient of the testimony. The more extraordinary the testimony to the experience of the recipient, the less the testimony should be credited; if, however, the testimony being false would be more extraordinary than the testimony being true, then the testimony should be believed to a degree justified by the remainder, the weight of evidence for *p* after the weight of evidence against *p* has been

---

*Philosophical Issues* 14 (2004): 327, footnote 6, and Axel Gelfert, “Hume on Testimony Revisited,” *Logical Analysis and History of Philosophy* 13 (2010): 60-75 each question how Hume is standardly interpreted in this debate. How he has been standardly interpreted and challenges to that interpretation are not the concern of my paper. Mindful of these worries, however, in representing his views I stick to a close reading of his work and describe the view I find articulated there “Humean” rather than “reductionist” or “inferentialist.”

<sup>23</sup> Hume, *An Enquiry concerning Human Understanding*, 171.

<sup>24</sup> Hume, *An Enquiry concerning Human Understanding*, 171.

subtracted.<sup>25</sup> Hume's list is not exhaustive; he supposes there to be many such particulars.<sup>26</sup>

That such particulars contribute to our evidence and so impact on the justificatory force of testimony is supported by induction. Believing ultimately that, excepting in cases of perceptual knowledge and memorial knowledge, we are dependent on experience for knowledge, and therefore also presumably for knowledge conducive justification, he holds that belief should be attuned to the experienced frequency of conjunctions of events in accordance with his evidential calculus.<sup>27</sup> The wise man's degree of belief will reflect this variability of evidential weight; he "proportions his belief to the evidence."<sup>28</sup>

It appears that Hume is well-placed to explain how the judge can enjoy greater testimonial justification than the child; having more experience to draw on, she can potentially be more justified than the child. However, it seems natural to think that his account of testimonial knowledge discords with the Uniformity Claim. Requiring an agent to check with her experience in order to enjoy testimonial justification, at least consciously, isn't consistent with the claim that in standard cases of testimonial reception we simply accept testimony, and that we can have justified testimonial beliefs in those same cases.<sup>29, 30</sup>

Considering what Hume writes about testimony, it's not obvious why Hume should be taken to be committed to thinking that such checking is necessary. As long as agents do proportion their belief to the evidence, it's hard to see why actually checking with their experiences or not should matter for Hume. Of course the belief couldn't just be formed in a way that makes it so that it is only luckily proportioned to the evidence. An issue that arises is that even if it is correct that it doesn't matter to Hume's assessment of the force of an agent's doxastic justification whether they've actually checked or not, what does matter is

---

<sup>25</sup> Hume, *An Enquiry*, 172.

<sup>26</sup> Hume, *An Enquiry*, 171.

<sup>27</sup> Hume, *An Enquiry*, 110, 170.

<sup>28</sup> Hume, *An Enquiry*, 170. In his own work concerning human understanding, John Locke, *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Oxford University Press, 1975), 697, makes a similar claim. He writes that the mark of a lover of truth is that he does not entertain "any Proposition with greater assurance than the Proofs it is built upon will warrant."

<sup>29</sup> Hume, *An Enquiry*, 171, might challenge the uniformity thesis. He writes that frequently testimony is met with hesitation. However, I'll leave exploration of that possible challenge to one side as I believe it is not immediately crucial to the articulation of a promising Humean view of testimonial justification.

<sup>30</sup> See Coady, *Testimony*, for an example of this criticism of the Humean position.

that it would be unclear how it might happen that the agent's experience might inform their belief in the way that Hume thinks it should in the absence of actually checking. This is a problem given that we want to explain the Far-Reaching Dependence Thesis in a way that fits with the Uniformity claim.

Psychological research suggests that it may be possible to explain how an agent's experience might inform their belief in a way that would be consistent with the uniformity claim. Dual process theorists argue for the claim that "there are two different modes of processing," sometimes described as system 1 and system 2, sometimes as sub-personal and personal.<sup>31, 32</sup> System 1 processes are generally agreed to be "rapid, parallel and automatic in nature," while system 2 is taken to be "slow and sequential" and evolutionarily recent in comparison with system 1.<sup>33</sup> In a more recent work Evans adds that there seems to be widespread agreement that system 1 is unconscious and system 2 is effortful.<sup>34</sup> Given what we know about the two systems, though admittedly in the absence of any definitive experimental results, it seems plausible to expect that system 1 picks up on cues of differing strengths that support and cues that undermine belief in testimony. It further seems plausible that experience of testimonial reception can contribute to an agent being sensitive and appropriately responsive to cues and so enjoying greater justification when they believe a piece of testimony. It would follow from the relevant work being done by system 1 that an agent's experience of standard cases of testimonial reception would conform to the Uniformity Claim. If this is right then we have an explanation of testimonial justification that accords with the Uniformity Claim and goes some way towards explaining what makes some testimonial beliefs more justified than others. To see that this seems supported by the system 1/system 2 distinction, consider how Frankish describes the workings of the two systems thus:

[M]ost of our behaviour is generated without the involvement of personal reasoning. Think about the actions involved in such everyday activities as driving a car, holding a conversation, or playing sports. These are intelligent actions, which are responsive to our beliefs and desires (think of how beliefs about the rules of the game shape the actions of a football player), and a great

---

<sup>31</sup> Jonathan St. B. T. Evans, "Dual-Processing Accounts of Reasoning, Judgement, and Social Cognition," *Annual Review of Psychology* 59 (2008): 256.

<sup>32</sup> Different authors have labelled the modes differently. Other labels for dual processes include experiential and rational, heuristic and systematic, intuitive and analytic, holistic and analytic. Evans, "Dual-Processing Accounts," 257.

<sup>33</sup> Jonathan St. B. T. Evans, "In Two Minds: Dual-Process Accounts of Reasoning," *Trends in Cognitive Sciences* 7, 10 (2003): 454.

<sup>34</sup> Evans, "Dual-Processing Accounts," 270.

deal of complex mental processing must be involved in generating them. Yet, typically, they are performed spontaneously with no prior conscious thought or mental effort.<sup>35</sup>

Nevertheless, even if there is a way for experience to act as an input into testimonial justification in a way that is consistent with the Uniformity Claim and the Far-Reaching Dependence Thesis, such an input isn't quite adequate. An agent won't be more justified in her testimonial belief simply in virtue of that agent being more experienced. Our judge is old but she is also wise. Rather, what is required is that the testimonial recipient be virtuous in her beliefs, though experience plausibly does contribute to the overall virtuousness of the testimonial recipient.<sup>36</sup>

#### 4. Conclusion

In this paper I have shown how a Humean view of testimonial knowledge may be articulated that incorporates attractive features of the credulist view with attractive features of a non-credulist view. In particular I've shown how this Humean view can overcome two related problems for accounts of testimonial justification. While challenges for the view remain, for example explaining how testimonial justification is possible in young children, this is a challenge also facing other leading accounts of testimonial justification.<sup>37</sup>

---

<sup>35</sup> Keith Frankish, "Systems and Levels: Dual-system Theories and the Personal-Subpersonal Distinction," in *In Two Minds: Dual Processes and Beyond*, eds. Jonathan Evans and Keith Frankish (Oxford: Oxford University Press, 2009), 89-107. Interestingly the approach based on dual process theory as outlined offers a way of seeing the counterfactual sensitivity credulist position and the Humean position as being much closer than the pair are usually represented as being.

<sup>36</sup> It's standard in virtue ethics to claim that experience is necessary for the practical wisdom that being morally virtuous requires. For example, see Rosalind Hursthouse, "Virtue Ethics," *The Stanford Encyclopedia of Philosophy* ed. Edward N. Zalta (Summer 2012 Edition), URL = <<http://plato.stanford.edu/archives/sum2012/entries/ethics-virtue/>>.

<sup>37</sup> Thanks for comments and suggestions go to Christos Kyriakou and audiences at the *David Hume and Modern Philosophy* conference in Moscow and at a work in progress meeting at the University of Edinburgh. Special thanks go to Duncan Pritchard for his very helpful comments.



# DEBATE





# ALTERNATIVE SELF-DEFEAT ARGUMENTS: A REPLY TO MIZRAHI

Michael HUEMER

ABSTRACT: I address Moti Mizrahi's objections to my use of the Self-Defeat Argument for Phenomenal Conservatism (PC). Mizrahi contends that other epistemological theories can be supported by parallel self-defeat arguments. I argue that the self-defeat arguments for other theories either (a) are compatible with PC and thus present no problem, or (b) have a false premise, unlike the self-defeat argument for PC.

KEYWORDS: dogmatism, phenomenal conservatism, seemings, self-defeat argument

According to Phenomenal Conservatism (PC),

If it seems to S that P, then, in the absence of defeaters, S thereby has at least some degree of justification for believing that P.

I have previously defended PC using the "Self-Defeat Argument," which includes roughly the following reasoning:

1. All our beliefs (that are reasonable candidates for being justified) are based upon appearances.
2. A belief is (doxastically) justified only if what it is based upon constitutes an adequate source of (propositional) justification.
3. So, if appearances are not a source of (propositional) justification, then all our beliefs are (doxastically) unjustified (including the belief that appearances are not a source of justification).<sup>1</sup>

In an interesting recent paper, Moti Mizrahi has objected to my use of the Self-Defeat Argument for Phenomenal Conservatism.<sup>2</sup> Mizrahi maintains that if my argument works, then analogous self-defeat arguments may be deployed on behalf of other epistemological theories, including, in particular, evidentialism. As Mizrahi would have it, the evidentialist could argue:

---

<sup>1</sup> For a full statement and defense of the argument, see my "Compassionate Phenomenal Conservatism," *Philosophy & Phenomenological Research* 74 (2007): 30-55; and "Phenomenal Conservatism and Self-Defeat: A Reply to DePoe," *Philosophical Studies* 156 (2011): 1-13. For brevity, I shall hereafter omit the parenthetical phrases from the argument.

<sup>2</sup> Moti Mizrahi, "Phenomenal Conservatism, Justification, and Self-Defeat," *Logos & Episteme* 5 (2014): 103-110.

- 1\*. All our beliefs are based upon evidence.
2. A belief is justified only if what it is based upon constitutes an adequate source of justification.
- 3\*. So, if evidence is not a source of justification, then all our beliefs are unjustified.

So far, so good. But now I do not see where the problem for my view is supposed to arise. Why should I be troubled if this evidentialist-friendly argument succeeds?

Here is one possible reason. The success of that argument would trouble me if I rejected its conclusion, and especially if the conclusion were incompatible with, or at least in tension with, PC. But in fact I accept (3\*), and I see no tension between (3\*) and PC. Surely evidence is a source of justification! I would merely go on to note that our evidence consists of appearances. (In fact, depending on what “evidence” means, PC might just be a form of evidentialism.)

Here is a second thought. As Mizrahi observes, the success of the above evidentialist argument shows that PC is not *unique* or *special* in being supportable by a self-defeat argument. But I see no reason why this should be problematic. One who endorses a conclusion based upon a certain argument need not deny that any other conclusions can be supported by the same style of argument. Of course, one should, at least *prima facie*, be troubled if the same style of argument supports *contrary* conclusions, or supports *implausible* conclusions. But evidentialism is not contrary to PC, nor is it implausible.

But did I not claim, in my original presentation of the Self-Defeat Argument, that PC *was* unique in being supportable by this type of argument? Not exactly. What I claimed was that the Self-Defeat Argument tells against any *opposing* theory, that is, any theory that is actually incompatible with PC. I have no objection *per se* to theories that are *compatible* with PC, and I do not deny that some theories *that are compatible with PC* might be supported by a plausible self-defeat argument.

Perhaps evidentialism was simply a poor choice of example, because it happens to be compatible with PC. Mizrahi appears to hold that a (approximately equally persuasive?) self-defeat argument could be given for *any* theory of basic propositional justification (or at least for many such theories?). Certainly it should trouble me if an equally plausible self-defeat argument could be constructed for some competing theories of basic propositional justification (that is, theories incompatible with PC). But I deny that this is the case. The conclusion of the relevant self-defeat argument would be of the form “If X is not a source of justification, then all our beliefs are unjustified” (with the further suggestion of

course being that we should accept X as a source of justification). PC holds, in essence, that appearances are a source of justification (check the formulation of PC above). But no proposition of the form “If X is not a source of justification, then all our beliefs are unjustified,” nor of the form “X is a source of justification,” can be incompatible with the claim “Appearances are a source of justification.” (“a is F” is not incompatible with “b is F.”) To have a theory that is incompatible with PC, one must not merely hold that X *is* a source of justification (for some X), but actually hold that *appearances* are *not* a source of justification.

Perhaps the problem is not that self-defeat arguments can be used to support conclusions incompatible with PC, but merely that they could be used to support *implausible* conclusions, or conclusions that I would reject. For example, suppose someone held the view that tea leaf readings are a basic source of propositional justification. A self-defeat argument might be constructed for this view. It would go as follows:

- 1'. All our beliefs are based upon tea leaf readings.
2. A belief is justified only if what it is based upon constitutes an adequate source of justification.
- 3'. So, if tea leaf readings are not a source of justification, then all our beliefs are unjustified.

But this argument is just utterly implausible, mainly because the first premise is obviously false. This is a counter-example to the claim that any theory of justification can be supported by an *equally plausible* self-defeat argument (e.g., one whose first premise is as plausible as my premise (1)).

What does Mizrahi say by way of explaining why other theories of basic propositional justification could be supported by self-defeat arguments? He writes:

I submit that any basic source of propositional justification [...] can be plugged into the following argument scheme instead of X:

- All our beliefs (in relevant cases) are based upon X.
- A belief is (doxastically) justified only if what it is based upon constitutes an adequate source of (propositional) justification.
- ∴ If X is not a source of justification, then all our beliefs are unjustified [...].

This is so because any theory of basic propositional justification that identifies X as a basic source of justification would have to appeal to X in order to justify itself *on pain of self-defeat*. Since the ‘in the relevant cases’ is supposed to rule out beliefs that are clearly not justified, the remaining beliefs must be justified in virtue of being based upon X.<sup>3</sup>

---

<sup>3</sup> Mizrahi, “Phenomenal Conservatism,” 108-9.

I am not sure exactly how to read this passage. It appears that Mizrahi is considering only *complete* theories of basic propositional justification, that is, theories that purport to identify the *sole* ultimate source of propositional justification, rather than theories that merely purport to identify *a* source of propositional justification. (Aside: PC, as formulated at the outset, is not complete in this sense; it only purports to identify *a* source of justification. Nevertheless, I would in fact be strongly inclined to accept appearances as the sole ultimate source of justification.) As Mizrahi notes, if one holds that X is the only ultimate source of justification, then one is committed to holding that X is the ultimate source of justification for the claim that X is the only ultimate source of justification. My best guess is that this is somehow supposed to show that one who holds that X is the sole ultimate source of justification is then entitled to assert that all our beliefs (in relevant cases) are based upon X. But I cannot see how this is the case. Suppose someone holds that tea leaf readings are the only ultimate source of justification (call this the Tea Leaf Theory). Their merely holding this bizarre view would not in the least render plausible their assertion that all our beliefs in relevant cases are based upon tea leaf readings.

One of the tasks for epistemological theory is to *account for* the justification of certain sorts of beliefs that we antecedently take to be justified – e.g., my perceptual belief that there is a squirrel in the tree outside, my belief that I feel hungry, your belief that  $2+1=3$ . This is not to say that we must start with complete agreement on which beliefs are justified, or even any *definite* agreement on whether any particular belief is justified; it is only to suggest that there are certain initially reasonable candidates for justified beliefs (like the belief about the squirrel), and other beliefs that are not even initially reasonable candidates (like the belief, based on pure wishful thinking, that I am the world's greatest basketball player). It just is obviously false that all the reasonable candidate beliefs are based upon tea leaf readings. It doesn't matter if someone holds that tea leaf readings are the only source of justification – that won't somehow make it plausible to hold that my belief about the squirrel, your belief that  $2+1=3$ , and so on, are all based upon tea leaf readings. We know that they are not.

Here is another thought. Suppose S holds that X is the only ultimate source of propositional justification. Assume also that S accepts the basing requirement for doxastic justification (premise 2 in the self-defeat argument). S would then be committed to the view that, if there are any justified beliefs, they are all based upon source X. S would therefore be committed to accepting both premises in the following argument:

- 1". If there are any doxastically justified beliefs, they are based upon X.

2. A belief is justified only if what it is based upon constitutes an adequate source of justification.
- 3". Therefore, if X is not a source of justification, then all our beliefs are unjustified.

Since the argument is obviously valid, S must hold that the argument is in fact *sound*. We can thus conclude: any complete theory of basic propositional justification can be supported by a self-defeat argument *that the theory's proponents are committed to viewing as sound*.

Perhaps this is in the neighborhood of what Mizrahi was getting at in the passage I quoted above. However, I still see no problem for my view. Suppose the proponent of theory T is committed to the view that there is a sound self-defeat argument for T. That of course does not imply that *I* am committed to the view that there is a sound self-defeat argument for T. Nor does it imply that it is even the slightest bit *plausible* to think that there is a sound self-defeat argument for T. Thus, for example, proponents of the Tea Leaf Theory of justification would be committed to holding that there is a sound self-defeat argument for the Tea Leaf Theory. But this does not trouble me at all, since *I* remain free to hold that the self-defeat argument for the Tea Leaf Theory is ridiculous.

Might opponents of PC react similarly to my argument? Well, not reasonably. Consider some reasonable candidate for a justified belief, say your belief that  $2+1=3$ . If you reflect on this belief, I think you are just going to find it plausible that it is based upon the appearance that  $2+1=3$  (its seeming to you that  $2+1=3$ ), in a way that you will *not at all* find it plausible that the belief is based upon a tea leaf reading that told you that  $2+1=3$ . That is why the self-defeat argument for PC is reasonable, whereas the self-defeat argument for Tea Leaf Theory is unreasonable. Of course, if, when you reflect, you *don't* find it plausible that you believe that  $2+1=3$  because it seems to you that  $2+1=3$ , then my argument won't work on you.

"Okay," you might say, "big deal. So the self-defeat argument can't be used to support the Tea Leaf Theory. But what about theories opposing PC that epistemologists have actually advanced? For example, could a self-defeat argument be used to support the Acquaintance Theory of justification?" The Acquaintance Theory holds that one has non-inferential justification for believing that P if and only if one is acquainted with the fact that P.<sup>4</sup> Imagine the Acquaintance Theorist arguing, by analogy to my own argument:

---

<sup>4</sup> For defense of the acquaintance theory, see Bertrand Russell, *The Problems of Philosophy* (New York: Oxford University Press, 1997). The characterization in the text is somewhat simplified; for a more complex acquaintance theory, see Richard Fumerton, *Metaphysical and*

Michael Huemer

- 1#. All our beliefs are based upon acquaintance.
2. A belief is justified only if what it is based upon constitutes an adequate source of justification.
- 3#. So, if acquaintance is not a source of justification, then all our beliefs are unjustified.

Is this argument as plausible as the analogous Self-Defeat argument for PC?

I think not. Premise (1#) is simply false. For example, if one has a perfectly realistic hallucination of a squirrel, with no reasons for doubting the experience's veridicality, and one believes on the basis of this experience that a squirrel is present, one thereby has a justified belief (certainly, at least, a belief that is a reasonable candidate for being justified). But this belief, at least on its face, is not based upon acquaintance. Similarly, if a philosopher believes the Comprehension Axiom of naive set theory (prior to the discovery of Russell's Paradox or any other objections to the Axiom) on the basis of the philosopher's intuition that the Comprehension Axiom is true, the philosopher has a justified belief which is not based on acquaintance.

Acquaintance theorists might dispute my claims here. Obviously, my belief that a squirrel is present is not directly based upon acquaintance with a squirrel. But it might be *indirectly* based upon acquaintance with, say, a *sensory experience* of a squirrel. According to the usual view, one's acquaintance with the sensory experience directly justifies one's belief that one has that very experience. The belief that one has a sensory experience of a squirrel, combined with the belief that sensory experiences are reliable indicators of facts in the external world, then inferentially justifies the belief that a squirrel is present.<sup>5</sup> Similarly, the philosopher's belief in the Comprehension Axiom might be indirectly based upon acquaintance with an *intuition*, which, when combined with the belief that intuitions are reliable indicators of facts about abstract objects, gives one justification for believing the Comprehension Axiom.<sup>6</sup>

My first line of reply to this sort of view, as I have suggested elsewhere, would be to question that one can really justify the needed major premises (i.e., that sensory experiences and intuitions are reliable), if one does not start out by accepting appearances at least *prima facie*.<sup>7</sup> My second line of reply would be that

---

*Epistemological Problems of Perception* (Lincoln: University of Nebraska Press, 1985), 73-9.

<sup>5</sup> See Russell, *Problems of Philosophy*, Ch. 2.

<sup>6</sup> For a similar approach to justifying introspective beliefs, see Ali Hasan, "Phenomenal Conservatism, Classical Foundationalism, and Internalist Justification," *Philosophical Studies* 162 (2013): 130-32.

<sup>7</sup> See my *Skepticism and the Veil of Perception* (Lanham: Rowman & Littlefield, 2001) and

the inferential justification that acquaintance theorists would rely upon simply is not the actual psychological basis for our perceptual or intuitive beliefs – even if the acquaintance theorist is able to construct convincing arguments for the reliability of sense perception and intuition, few if any people can seriously be claimed to be relying on such arguments when we form perceptual or intuitive beliefs. I therefore think that the acquaintance theorist will not be able to plausibly maintain that our actual beliefs (that are reasonable candidates for being justified) are based upon the reasoning that supposedly provides propositional justification for them. Given (i) the psychological facts about how we actually form beliefs, (ii) the acquaintance theorist’s account of what provides propositional justification for beliefs about the external world and about abstract objects, and (iii) the basing requirement for doxastic justification, the acquaintance theorist will have to say that almost all of our beliefs are doxastically unjustified. This is a highly implausible result.

It does not matter if it is possible to construct an *implausible and unsound* version of the self-defeat argument in defense of other theories of justification. That casts no doubt on my use of the self-defeat argument for PC. What sets PC apart from its rivals (i.e., theories that are incompatible with PC) is that the Self-Defeat Argument for PC has a first premise that is plausible and true, whereas the self-defeat arguments for rival theories have first premises that are implausible and false. Obviously, more needs to be said in defense of that claim. But I have said more – in my earlier defense of the Self-Defeat Argument, I have argued that the plausible candidates for justified beliefs are in fact based upon its seeming to one that P, rather than, for example, on one’s being acquainted with the fact that P, on P’s being delivered by a reliable belief-forming method, etc.<sup>8</sup> The Self-Defeat Argument thus supports the epistemic import of appearances, rather than of one of these other possible sources.

---

*Ethical Intuitionism* (New York: Palgrave Macmillan, 2005), 107-22.

<sup>8</sup> Huemer, “Compassionate Phenomenal Conservatism,” 39-48.





## NOTES ON THE CONTRIBUTORS

**Rodrigo Borges** is a PhD candidate in philosophy at Rutgers University. His main research interest is in epistemology, with a focus on knowledge-first epistemology, but he also has interests in philosophy of language and philosophy of science. He co-edits with Peter D. Klein and Claudio de Almeida the volume “Explaining Knowledge: New Essays on the Gettier Problem” (Oxford University Press, forthcoming). His paper “Inferential Knowledge and the Gettier Conjecture” will appear in that volume. Contact: epistemem@gmail.com.

**Thomas Hall** is a recent graduate of Western Washington University, and is beginning his first year in the philosophy doctoral program at the University of Southern California in the Fall of 2014. His main areas of interest are in metaphysics, metametaphysics, and metaethics. Contact: thomashall13@msn.com.

**Michael Huemer** is professor of philosophy at the University of Colorado at Boulder. He is the author of more than 50 academic articles in ethics, epistemology, political philosophy, and metaphysics, as well as three fascinating books that all readers should buy immediately: *Skepticism and the Veil of Perception* (Rowman & Littlefield, 2001), *Ethical Intuitionism* (Palgrave Macmillan, 2005), and *The Problem of Political Authority* (Palgrave Macmillan, 2013). Contact: ikd3lob02@sneakemail.com.

**Ned Markosian** is a Professor of Philosophy at Western Washington University. He is co-author – with John W. Carroll – of *An Introduction to Metaphysics* (Cambridge University Press, 2010). He writes mainly about metaphysics, but also has interests in epistemology and ethics. Some recent publications include “A Spatial Approach to Mereology” (in Shieva Kleinschmidt, ed., *Mereology and Location*, Oxford University Press, 2014), “Two Puzzles About Mercy” (*Philosophical Quarterly*, 63, 2013), and “Agent Causation as the Solution to All the Compatibilist’s Problems” (*Philosophical Studies*, 157, 2012). Contact: ned.markosian@wwu.edu.

**Rinat M. Nugayev** is a Professor of Philosophy at Russian University of Cooperation. He is a philosopher with contributions in epistemology, philosophy, sociology and history of science. He has authored papers in these areas for *The*

Logos & Episteme

*British Journal for the Philosophy of Science, Philosophy of Science, International Studies in the Philosophy of Science, Journal for General Philosophy of Science, Foundations of Science, Epistemologia, Centaurus, Physis, Historia Scientiarum, Theoria and Historia Scientiarum, Scientia, Almagest.* His books include: *Reconstruction of Mature Theory Change: A Theory Change Model* (Peter Lang, 1999) and the books published in Russian: *Einstein's Revolution: Interdisciplinary Context* (2010), *The Copernican Revolution: Unification of Mundane Physics and Mathematics of the Skies* (2012), and *Maxwellian Scientific Revolution: Reconciliation of research programmes of Ampere-Weber, Faraday and Young-Fresnel* (2013). Contact: [nugayevrinat@gmail.com](mailto:nugayevrinat@gmail.com).

**Shane Ryan** is a postdoctoral research fellow at Soochow University in Taipei City. He was awarded his PhD from the University of Edinburgh in 2013. His main areas of interest are virtue epistemology, social epistemology, and social philosophy. He has several papers under review and forthcoming, including "Standard Gettier Cases: A Problem for Greco?", which is to appear in *Grazer Philosophische Studien*. Contact: [shaneryan27@hotmail.com](mailto:shaneryan27@hotmail.com).

## ***LOGOS & EPISTEME: AIMS & SCOPE***

*Logos & Episteme* is a quarterly open-access international journal of epistemology that appears in printed and on-line version at the end of March, June, September, and December. Its fundamental mission is to support and publish various current reflections and researches that aim at investigating, analyzing, interpreting or philosophically explaining the human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal will publish articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal intends to promote all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.



# NOTES TO CONTRIBUTORS

## 1. Accepted Papers

The journal accepts for publication articles, discussion notes and book reviews submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The editors of *Logos & Episteme* reserve the right to refuse all future papers belonging to the authors who fail to observe this rule.

## 2. Submission Address

Please submit your manuscripts electronically at: [logosandepisteme@yahoo.com](mailto:logosandepisteme@yahoo.com). Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

## 3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

## 4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (\*.rtf) or Microsoft Word document (\*.docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in *The Chicago Manual of Style*, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

## **5. Official Languages**

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

## **6. Abstract**

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

## **7. Author's CV**

A short CV including the author's affiliation and professional address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

## **8. Review Process**

The reason for these requests is that all articles, with the exception of articles from the invited contributors, will be subject to a strict blind-review process. Therefore the authors should avoid in their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it. The review process is intended to take no more than six months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors. Processing of papers in languages other than English may take longer. The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers. The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

## **9. Acceptance of the Papers**

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are

received and they will appear in the journal in the alphabetical order of their authors.

### **10. Responsibilities**

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations. The papers containing racist and sexist opinions assumed by their authors will be rejected. The presence in texts of sexist or racist terms is accepted only if they occur in quotations or as examples.

### **11. Checking Proofs**

Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

### **12. Reviews**

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice și Sociale „Gh. Zane” Academia Română, Filiala Iași, Str. Teodor Codrescu, Nr. 2, 700481, Iași, România. The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

### **13. Property & Royalties**

Articles accepted for publication will become the property of Logos & Episteme and may not be reprinted or translated without the previous notification to the editors. No manuscripts will be returned to their authors. The journal does not pay royalties. Authors of accepted manuscripts will receive free of charge one copy of the issue containing their papers.

### **14. Permissions**

Authors have the right to use their papers in whole and in part for non-commercial purposes. They do not need to ask permission to re-publish their

## Logos & Episteme

papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.