

Volume 8 Issue 1 2021

Symposion

Theoretical and Applied Inquiries
in
Philosophy and Social Sciences

Romanian Academy, Iasi Branch
„Gheorghe Zane” Institute for Economic and Social Research
ISSN: 1584-174X EISSN: 2392-6260

Advisory Board

Ștefan Afloroaei, “Alexandru Ioan Cuza” University of Iași
Marin Aiftincă, Romanian Academy
Scott F. Aikin, Vanderbilt University
Jason Aleksander, Saint Xavier University
Helen Beebee, University of Manchester, United Kingdom
Richard Bellamy, European University Institute
Ermanno Bencivenga, University of California, Irvine
Alex Blum, Bar-Ilan University, Israel
Harry Brighouse, University of Wisconsin-Madison
Thom Brooks, Durham Law School
Gideon Calder, University of South Wales
Paula Casal Ribas, ICREA-Universitat Pompeu Fabra
Daniel Conway, Texas A&M University
Drucilla Comell, Rutgers University
Lucian Dîrdală, „Mihail Kogălniceanu” University of Iași
Eva Erman, Stockholm University
John Farina, George Mason University
Hans Feger, Freie Universität Berlin
Alessandro Ferrara, University of Rome “Tor Vergata”
Nancy Fraser, The New School for Social Research
Miranda Fricker, City University of New York, USA
Moira Gatens, University of Sydney, Australia
Steve Fuller, University of Warwick
Anca Gheaus, University of Sheffield
Paul Gombert, University of California, Davis
Steven Hales, Bloomsburg University
Sally Haslanger, Massachusetts Institute of Technology, USA
Nils Holtug, University of Copenhagen
Axel Honneth, Goethe-Universität Frankfurt/ Columbia University, New York

Franz Huber, University of Toronto
Adrian-Paul Iliescu, University of Bucharest
Alison Jaggar, University of Colorado Boulder, USA
Eva Feder Kittay, Stony Brook University
Thomas Kroedel, Humboldt University of Berlin
Carolyn Korsmeyer, University at Buffalo, State University of New York, USA
Janet Kourany, University of Notre Dame, USA
Franck Lihoreau, University of Coimbra
Clayton Littlejohn, King’s College London
Niklas Möller, Royal Institute of Technology, Sweden
Jonathan Neufeld, College of Charleston
Serena Olsaretti, ICREA-Universitat Pompeu Fabra
Jānis T. Ozoliņš, Australian Catholic University
Philip N. Pettit, Princeton University
Thomas Pogge, Yale University
Eduardo Rivera-López, Universidad Torcuato Di Tella
John Roemer, Yale University
Jennifer M. Saul, University of Sheffield, United Kingdom
Samuel Scheffler, New York University
Lavinia Stan, Saint Xavier University
Alexandru Surdu, Romanian Academy
Vasile Tonoiu, Romanian Academy
Hamid Vahid, Institute for Fundamental Sciences Tehran
Gheorghe Vlăduțescu, Romanian Academy
Jonathan Wolff, University College London
Federico Zuolo, Freie Universität Berlin, Otto Suhr Institut for Political Science

Founding Editor

Teodor Dima (1939-2019)

Editorial Board

Editors-in-Chief

Eugen Huzum, Cătălina-Daniela Răducu

Executive Editors

Ionuț-Alexandru Bârliba, Vasile Pleșca

Consulting Editor for English

Translations

Cristina Emanuela Dascălu

Assistant Editors

Alina Botezat, Irina Frasin, Alina Haller,

Aurora Hrițuleac, Liviu Măgurianu,

Alexandru Sava, Ioan Alexandru Tofan,

Codrin Dinu Vasiliu

Web & Graphics

Virgil-Constantin Fătu, Simona-Roxana
Ulman

Contact address: Institutul de Cercetări Economice și Sociale „Gh. Zane”, Iași, str. T. Codrescu, nr.2, cod 700481, Romania. Phone/Fax: 004 0332 408922. Email: symposion.journal@yahoo.com. www.symposion.acadiasi.ro.

TABLE OF CONTENTS

RESEARCH ARTICLES

Arnold Cusmariu, <i>The Cogito Paradox</i>	7
Michael F. Duggan, <i>Looking for Black Swans: Critical Elimination and History</i>	45
Samson Liberman, <i>Attention Deficit: Alienation in Platform Capitalism</i>	79
Daniel Rønnedal, <i>Perfect Happiness</i>	89
Rajesh Sampath, <i>The Question as to Why We Have to Live Out the Agony of Our Epoch and its Fundamental Un-Answerability: A Reading of the Preface to the 1967 Edition of Klossowski's' Original 1947 Sade My Neighbor</i>	117
Information about Authors.....	139
About the Journal.....	141
Author Guidelines.....	143

RESEARCH ARTICLES

THE *COGITO* PARADOX

Arnold Cusmariu

Abstract: The *Cogito* formulation in *Discourse on Method* attributes properties to one conceptual category that belong to another. Correcting the error ends up defeating Descartes' response to skepticism. His own creation, the Evil Genius, is to blame.

Keywords: Descartes, *Discourse on Method*, *Cogito*, truth, certainty, validity, clear and distinct perception, skepticism, the Evil Genius.

Overview

In *Discourse on Method*, Descartes

- (a) refers to a *Cogito* in argument form as "*cette vérité*," "this truth,"
- (b) describes it as "*si ferme et si assurée*," "so firm and sure,"
- (c) claims it is immune to skeptical attacks on its certainty, and
- (d) concludes he could accept it "without scruple as the first principle ('*le premier principe*') of the philosophy that I was seeking."

But arguments are not true or false, nor certain or uncertain; their components are. To clear up matters, the machinery of modern logic must be deployed. *Zut alors!* Disaster follows, perpetrated by Descartes' own creation, the Evil Genius: (c) is false, toppling (d) and blunting Descartes' response to skepticism.

Discourse Cogito Translations

Sources

To mitigate translation bias, I consulted eight English translations of the *Discourse*: Haldane 1970 [1911], Veitch 1912, Cottingham 1985, Cress 1998, Clarke 1999, Maclean 2006, Kennington 2007 and Bennett 2017. For the French original, I used the Amazon Kindle edition of Descartes' works, which is based on the 1874 Levrault edition.

Comments on Translations

- Haldane, Veitch, Cress, Clarke and Kennington translate "*je pense, donc je suis*" as the familiar "I think, therefore I am"; while Cottingham, Maclean and Bennett translate as "I am thinking, therefore I exist." The occurrent sense, "I am thinking," seems to me preferable to the dispositional "I think," for several reasons.

- First, the occurrent sense is implied by the use of the temporal conjunction "*pendent*," "while," in "*Mais aussitôt après je pris garde que, pendent que je voulois ainsi penser que tout étoit faux ...*"
- Second, later in the same sentence Descartes writes "*il faloit nécessairement que moi qui le pensois fusse quelque chose*," whose correct translation requires the occurrent sense, namely, "it was necessarily the case that I, who was thinking ..."
- Thus, when Descartes states his *Cogito* a few words later in the same sentence, using the familiar "*je pense, donc je suis*" formulation, only a translation that uses the occurrent sense, "I am thinking," is consistent with the passage as a whole.
- Descartes characterizes the *Cogito* as "*si ferme et si assurée*," which translations render in a variety of ways, e.g., "so certain and so assured" (Haldane); "so firm and sure" (Cottingham); "so firm and certain" (Clarke); and "so secure and certain" (Maclean). Descartes seems to me to use two terms, "*ferme*" and "*assurée*," for emphasis, not because epistemic value is somehow additive, or because the certainty of the *Cogito* is enhanced if both terms are used as opposed to either by itself.
- Six of the eight translations render the end of "*il faloit nécessairement que moi qui le pensois fusse quelque chose*" as "was (had to be) something." The "*quelque chose*" language means "something or other" because Descartes recognizes that the nature of the self is a separate issue. This may be why Haldane and Veitch use the awkward, though neutral locution "should be somewhat."

Cogito Passage Translations

Haldane 1970 [1911], I, 101: And since all the same thoughts and conceptions which we have while awake may also come to us in sleep, without any of them being at that time true, I resolved to assume that everything that ever entered into my mind was no more true than the illusions of my dreams. But immediately afterwards I noticed that whilst I thus wished to think all things false, it was absolutely essential that the 'I' who thought this should be somewhat, and remarking that this truth '*I think, therefore I am*' was so certain and so assured that the most extravagant suppositions brought forward by the skeptics were incapable of shaking it, I came to the conclusion that I could receive it without scruple as the first principle of the Philosophy for which I was seeking.

Veitch 1912, 42-43: Finally, when I considered that the very same thoughts (presentations) which we experience when awake may also be experienced when we are asleep, while there is at that time not one of them true, I supposed that all the objects (presentations) that had entered into my mind when awake, had in them no more truth than the illusions of my dreams. But immediately upon this I observed that, whilst I thus wished to think that all was false, it was absolutely necessary that I, who thus thought, should be somewhat; and as I observed that this truth, '*I think, therefore I am*,' was so certain and of such evidence that no ground of doubt, however extravagant, could be alleged by the skeptics capable of shaking it, I concluded that I might, without scruple, accept it as the first principle of the philosophy of which I was in search.

Cottingham 1985, 127: Lastly, considering that the very thoughts we have while awake may also occur while we sleep without any of them being at that time true, I resolved to pretend that all the things that had ever entered my mind were no more true than the illusions of my dreams. But immediately I noticed that while I was trying thus to think everything false, it was necessary that I, who was thinking, was something. And observing that this truth '*I am thinking, therefore I exist*' was so firm and sure that all the most extravagant suppositions of the skeptics were incapable of shaking it, I decided that I could accept it without scruple as the first principle of the philosophy I was seeking.

Cress 1998, 18: And finally, considering the fact that all the same thoughts we have when we are awake can also come to us when we are asleep, without any of them being true, I resolved to pretend that all the things that had ever entered my mind were no more true than the illusions of my dreams. But immediately afterward I noticed that, while I wanted thus to think everything was false, it necessarily had to be the case that I, who was thinking this, was something. And noticing that this truth—*I think, therefore I am*—was so firm and so assured that all the most extravagant suppositions of the skeptics were incapable of shaking it, I judged that I could accept it without scruple as the first principle of the philosophy I was seeking.

Clarke 1999, 24-25: Finally, since I thought we could have all the same thoughts, while asleep, as we have while we are awake, although none of them is true at that time, I decided to pretend that nothing that ever entered my mind was any more true than the illusions of my dreams. But I noticed, immediately afterwards, that while I thus wished to think that everything was false, it was necessarily the case that I, who was thinking this, was something. When I noticed that this truth, '*I think, therefore I am*' was so firm and certain that all the most extravagant assumptions of the skeptics were unable to shake it, I judged that I could accept it without scruple as the first principle of the philosophy for which I was searching.

Maclean 2006, 28: Finally, considering that all the same thoughts which we have while awake can come to us while asleep without any one of them then being true, I resolved to pretend that everything that had ever entered my head was no more true than the illusions of my dreams. But immediately afterward I noted that, while I was trying to think of all things being false, it was necessarily the case that I, who was thinking them, had to be something; and observing this truth: *I am thinking, therefore I exist*, was so secure and certain that it could not be shaken by any of the most extravagant suppositions of the skeptics, I judged that I could accept it without scruple, as the first principle of the philosophy I was seeking.

Kennington 2007, 32-33: And finally, considering that all the same thoughts that we have while away can come to us also while we are sleeping, without there being any that are then true, I resolved to feign that all the things that had ever entered my mind were not more true than the illusions of my dreams. But immediately after, I noticed that while I thus chose to think that everything was false, it was necessarily true that I, who was thinking this, was something. And observing that this truth *I think, therefore I am* was so firm and so assured that all the most extravagant suppositions of the skeptics were incapable of shaking

it, I judged that I could accept it without scruple as the first principle of the philosophy that I was seeking.

Bennett 2017, 14-15: Lastly, I decided to pretend that everything that had ever entered my mind was no more true than the illusions of my dreams, because all the mental states we are in while awake can also occur while we sleep and dream, without having any truth in them. But no sooner had I embarked on this project than I noticed that while I was trying in this way to think everything to be false it *had* to be the case that I, who was thinking this, was something. And observing that this truth **I am thinking, therefore I exist** was so firm and sure that not even the most extravagant suppositions of the skeptics could shake it, I decided that I could accept it without scruple as the first principle of the philosophy I was seeking.

A *Cogito* Valid in the Sentential Calculus (SC)¹

Preliminary

Two claims made in the Overview section require explanation:

- It is a mistake to attribute semantic and epistemic properties to arguments rather than argument components.
- Descartes made this mistake in formulating the *Discourse Cogito*.

As to the first error, syllogistic logic, which Descartes studied, does not give meaning, for example, to “**AAA-1** is true” and “**AAA-2** is false.”² Moreover, descriptions such as “**AAA-1** is justified” and “**AAA-2** is unjustified” merely note that **AAA-1** is one of sixteen syllogisms valid in syllogistic logic, while **AAA-2** is not. The sentential and quantificational calculi likewise do not define arguments as true or false, justified or unjustified. A step in an argument can be said to be justified in the sense that the correct rule of inference was applied correctly. The sixteen valid syllogisms are effectively rules of inference, though they are inadequate for validating even the simplest proofs in Euclid.

As to the second error, several passages in the *Discourse* in addition to the *Cogito* passage attribute semantic and epistemic properties to arguments rather than argument components. That said, it should be noted that there are also *Discourse* passages that attribute semantic and epistemic properties correctly to opinions, thoughts or propositions. How the literature explains (if at all) why Descartes got it right in some places and not others, I do not know. I am not about to accuse him of carelessness.

¹ Objections will no doubt occur as readers work their way through the details below. To avoid disrupting the flow, I have relegated the matter to “Objections and Replies” at the end. I have also avoided debating the literature, which would have required book-length treatment.

² These are standard abbreviations in syllogistic logic. See below.

Discourse Examples of Correct Attributions

- (*Œuvres Complètes*, Kindle Edition, 20): "... *considérant combien il peut y avoir de diverses opinions* (**opinions**) *touchent une même matière, qui soient soutenues par des gens doctes, sans qu'il y en puisse avoir jamais plus d'une seule qui soit vraie* (**true**) ..." A semantic property is attributed to opinions.
- (*Œuvres Complètes*, Kindle Edition, 36): "... *decouvrir la fausseté* (**falsity**) *ou l'incertitude* (**uncertainty**) *de propositions* (**propositions**) *que j'examinais, non par des faibles conjectures* (**feeble conjectures**), *mais par des raisonnements* (**reasonings**) *clairs et assurés* (**clear and certain**) ..." The passage attributes semantic and epistemic properties to propositions and conjectures but, paradoxically, ends by attributing epistemic properties to reasonings.
- (*Œuvres Complètes*, Kindle Edition, 39): "... *des opinions* (**opinions**) *qu'en sait être fort incertaines* (**uncertain**), *tout de même que si elles étaient indubitables* (**indubitable**) ..." Epistemic properties are attributed to opinions.
- (*Œuvres Complètes*, Kindle Edition, 39): "... *et enfin, considérant que toutes les memes pensees* (**thoughts**) *que nous avod étant eveilles nous peuvent aussi venire quand nous dormons, sans qu'il y en ait aucune qui soit vraie* (**not true**) ..." A semantic property is attributed to thoughts: those had while asleep are said to be not true at that time.
- (*Œuvres Complètes*, Kindle Edition, 39): "*Après cela je considèrai en général ce qui est requis à une proposition* (**proposition**) *pour être vraie et certaine* (**true and certain**) ..." Semantic and epistemic properties are attributed to propositions.
- (*Œuvres Complètes*, Kindle Edition, 43): "*En sorte que si nous en avons assez souvent [idées]* (**ideas**) *qui contiennent de la fausseté* (**falsehood**)..." A semantic property is attributed to ideas. Though the term "*idées*" had to be added, it is clear that the reference is to ideas. Earlier in the passage, there is a reference to "*idées ou notions*," "ideas or notions."
- (*Œuvres Complètes*, Kindle Edition, 44): "... *ne doivent aucunement nous faire douter* (**doubt**) *de la vérité* (**truth**) *des pensées* (**thoughts**) *que nous avons éntant éveillés.*" Semantic and epistemic properties are attributed to thoughts.
- (*Œuvres Complètes*, Kindle Edition, 44): "... *elle nous dicte aussi que nos pensées* (**thoughts**) *no pouvant être toutes vraies* (**true**) ..." A semantic property is attributed to thoughts.

Discourse Examples of Incorrect Attributions

- (*Œuvres Complètes*, Kindle Edition, 19): "*Je me plaisais surtout aux mathématique, à cause de la certitude et de l'évidence de leurs raisons.*" Haldane, Veitch, Cottingham, Cress, Clarke Maclean, Kennington and Bennett all agree that "*à cause de la certitude et de l'évidence de leurs raisons*" attributes the epistemic terms "certainty" and "evidence" to reasoning rather than reasons, i.e., arguments rather than argument components, and translate accordingly. Some, however, translate "*raisons*" using the terms "demonstrations," "proofs" or "arguments," while others translate it literally

as “reasonings.” Maclean uses “incontrovertible” instead of the literal “certain” based on interpretive comments in his Introduction (Maclean 2006, I.)

- (*Œuvres Complètes*, Kindle Edition, 29): “... il n’y a eu que les seuls *mathematicisens* qui on pu trouver quelques **démonstrations**, c’est-a-dire quelques **raisons certaines et évidentes** ...” Cottingham, Cress, Clarke, Maclean, Kennington and Bennett agree that “*démonstrations, c’est-a-dire quelques raisons certaines et évidentes* ...” attributes the epistemic terms “certain” and “evident” to reasonings rather than reasons, i.e., arguments rather than argument components, and translate accordingly. Haldane and Veitch try to “rescue” Descartes by attributing epistemic terms to reasons rather than reasonings. Veitch: “... demonstrations, that is, any certain and evident reasons.” Haldane: “... demonstrations, that is to say, producing reasons which are evident and certain.”
- (*Œuvres Complètes*, Kindle Edition, 29): “... bien que je n’en espérasse aucune autre utilité, sinon qu’elles accoutumeraient mon esprit à se repaître de **vérités**, et ne se contenter point de **fausses raisons**.” Haldane, Cress, Maclean, and Kennington translate “*fausses raisons*” literally as “false reasoning(s).” The other four translators try to “rescue” Descartes in various ways from attributing a semantic property to arguments. Veitch: “such reasonings as were unsound.” Cottingham and Bennett: “bad reasoning.” Clarke: “faulty reasoning.”
- (*Œuvres Complètes*, Kindle Edition, 39): “... je rejetait comme **fausse** toutes les **raisons** que j’avois prise auparavant pour les **démonstrations** ...” Cress, Maclean, Kennington, Veitch and Clarke translate “*je rejetait comme fausse toutes les raisons*” literally as “I rejected as false all the reasoning(s) (arguments),” having translated “*démonstrations*” as “demonstrations.” Cottingham and Bennett write: “I rejected as unsound all the arguments ...” giving “*fausse*” the technical meaning “unsound” so it can be applied to arguments with false premises, conveniently absolving Descartes of the mistake of attributing a semantic property to arguments. Whether Descartes would have agreed that the syllogistic logic he knew recognizes the valid-sound distinction is an open question. Haldane writes: “I rejected as false all the reasons ...,” giving “*fausse*” its literal meaning but applying it to argument components, also absolving Descartes of the mistake of attributing a semantic property to arguments.
- Finally, the famous *Cogito* passage (*Œuvres Complètes*, Kindle Edition, 39) attributes truth and certainty to a *Cogito* stated in argument form, “... il fallout nécessairement que moi qui le pensois fusse quelque chose; et remarquant que **cette vérité**, je pense donc je suis, étoit **si ferme et si assurée** ...” The eight translators agree that a semantic property is intended by “*cette vérité*,” translated literally as “this truth.” The translators also agree that an epistemic property is intended by “*si ferme et si assurée*,” which they translate as follows: Haldane: “so certain and so assured.” Veitch: “so certain and of such evidence.” Cottingham: “so firm and sure.” Cress: “so firm and so assured.” Clarke: “so firm and certain.” Maclean: “so secure and certain.” Kennington: “so firm and so assured.” Bennett: “so firm and sure.”

Rescuing the *Discourse Cogito*

I will proceed as follows:

- Taking “therefore” in the *Discourse Cogito* at face value as a conclusion indicator, I will first build an argument that is valid in the sentential calculus (SC).
- Second, I will explain why this argument is not a syllogism.
- Third, I will state the valid *Cogito* argument as a sentence.
- Fourth, I will show that the resulting sentence is true—in fact, necessarily true.

As a result, Descartes avoids the following:

- misattributing truth to arguments;
- misattributing epistemic properties to arguments;
- having to say that the *Discourse Cogito* is a syllogism;
- having to say that the *Discourse Cogito* is an inference (in SC, for the time being);

Now that Descartes can refer to the *Discourse Cogito* as “*cette vérité*” and “*si ferme et si assurée*” without attributing semantic and epistemic properties to arguments, the question is whether he is right that “*cette vérité*” and “*si ferme et si assurée*” can be correctly be attributed to a *Discourse Cogito* in sentential form. Let us consider them in turn.

A *Cogito* Argument Valid in SC

We start by treating the two components of the *Cogito*, “I am thinking” and “I exist,” as premise and conclusion and add a material conditional premise that connects them.

(1) I am thinking.

(2) If I am thinking, then I exist.

Therefore, (3) I exist.

Next, we symbolize “I am thinking” as P ; “I exist” as Q ; the material conditional “if, then” using the arrow symbol \rightarrow ; and the triangular dot symbol \therefore for “therefore.”

(1*) P

(2*) $P \rightarrow Q$

\therefore (3*) Q

The argument sequence from (1) and (2) to (3) is a substitution instance of the argument sequence form from (1*) and (2*) to (3*), which is the SC rule of

inference *Modus Ponens* (**MP**). Thus, the *Discourse Cogito* expressed in argument form is valid in SC.

The *Cogito* Argument Is Not a Syllogism

“Syllogism” is shorthand for “categorical syllogism in standard form,” CSSF:

(I) A CSSF is an ordered sequence of three categorical propositions in standard form, two premises and a conclusion.

(II) Categorical propositions are of four types, each in subject-predicate form:

A: All _ are _ ; **E:** No _ are _ ; **I:** Some _ are _ ; **O:** Some _ are not _ .

(III) Terms flanking the copula are of three types: A subject term, a predicate term and a middle term.

(IV) Subject and predicate terms occur once in the premises and once in the conclusion.

(V) The middle term occurs only in the premises in four configurations called “moods.”

(VI) There are a total of six term occurrences.

Inspection shows that the *Discourse Cogito* argument from (1) and (2) to (3)

(1) I am thinking.

(2) If I am thinking, then I exist.

Therefore, (3) I exist.

is not a CSSF. For example:

- The components of this argument are not one of **A**, **E**, **I**, or **O**.
- Premise (2) is a material conditional and as such is not in subject-predicate form.
- Eliminating two occurrences of “I” to meet condition (VI) is not obvious.
- Reducing the *Cogito* argument to CSSF form is technically complex.³

³ Anthony Kenny writes (1968: 51): “The [*Cogito*] argument could be interpreted in a simple syllogism, provided we are willing to follow Descartes in regarding ‘exists’ as a predicate. ‘Whatever is thinking exists; but I am thinking; therefore, I exist.’” The matter is not as “simple” as Kenny thinks. Flawed accounts of a CSSF can also be found in Williams 1978, 89, Curley 1978, 27, 79), Wilson 1978, 55 and Markie 1986, 175. For Descartes’ views on the syllogism, see Ariew 2011, Ch. 10. For Descartes’ views on deduction see Clarke 1992, 258-285 and Gaukroger 1995, 115-118.

The *Discourse Cogito* Argument as a Sentence

To allow Descartes to refer to the *Discourse Cogito* as “*cette vérité*” and “*si ferme et si assurée*” without attributing semantic and epistemic properties to arguments, we state the argument from (1) and (2) to (3) in the form of a sentence:

(1) I am thinking.

(2) If I am thinking, then I exist.

Therefore, (3) I exist.

This is easy to do because (Copi 1967, 30) to every argument there corresponds a material conditional whose antecedent is the conjunction of its premises and whose consequent is the argument’s conclusion.⁴ This means that we can write the argument sequence from (1) and (2) to (3) in the form of a material conditional,

(4) If I am thinking and if I am thinking, then, I exist, then, I exist,

using the conjunction “and” to make it clear that (1) and (2) are being asserted jointly. *Voilà!* Descartes can now refer to the *Discourse Cogito* sentence as “*cette vérité*” and “*ferme et assurée*” without misattributing semantic and epistemic properties. But, is he right to do so in both cases? Let us consider them in turn.

Does “*Cette Verité*” Apply to the *Discourse Cogito* Sentence?

We can test whether (4) is true using a standard truth table, writing the sequence form from (1*) and (2*) to (3*) as a sentence, so that (4) is a substitution instance of (5),

(5) $(P \& (P \rightarrow Q)) \rightarrow Q$

1	2	3	4	5
P	Q	$P \rightarrow Q$	$P \& (P \rightarrow Q)$	$(P \& (P \rightarrow Q)) \rightarrow Q$
T	T	T	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	T

Table 1. Truth table test

⁴ This statement needs to be qualified somewhat because syllogistic logic does not define the material conditional, or any other truth functions for that matter. So, for example, we cannot write **AAA-1** in syllogistic logic as “If all *M* are *P* and all *S* are *M*, then all *S* are *P*.” This means that Descartes could not have solved the misattribution problem in syllogistic logic, the only kind available in his day. An aside: I have not researched the matter and cannot say whether Descartes was aware of the serious limitations of syllogistic logic for mathematical purposes; and if he was, whether he made an effort to develop an alternative. The syllogism was (famously) considered the “final word” in logic even as late as Kant.

Column 5 shows that (5) is true for any truth-functional assignments in columns 1 and 2. Because (4) is a substitution instance of (5), the *Discourse Cogito* sentence is true and can be referred to as “*cette vérité*.” Moreover, (5) is a tautology. On the assumption that tautologies are necessary truth, (5) is necessarily true, as are substitution instances of it.

Is the Discourse Cogito Sentence “Ferme et Assuré”?

Skepticism is an attack on the epistemic value of a proposition, not its truth value. That is, no matter what the epistemic value of proposition *P* might be as a result of “extravagant” skeptical attacks, it does not follow that *P* is false or that *P* is not necessarily true. Similarly, doubts about *P*, however persuasive, at most can lower our confidence that *P* is true. They cannot change the truth value of *P* from true to false or from necessarily true to contingently true. To avoid confusion, this part of the *Discourse Cogito* statement

“... et remarquant que cette vérité, je pense, donc je suis, étoit si ferme e si assurée, que toutes les plus extravagantes suppositions de sceptiques n’étoient pas capable de l’ébranler ...”

needs to be interpreted in a way that implies

“... et remarquant que les plus extravagantes suppositions de sceptiques n’étoient pas capable d’ébranler la certitude de cette vérité, je pense, donc je suis ...”

Descartes’ claim must be interpreted to mean that no matter how “extravagant,” skeptical “suppositions” cannot change (“*ebranler*,” “shake”) the epistemic value of the *Cogito*. At issue, then, is the certainty of the *Cogito* argument paraphrased as a sentence, (4)

(4) If I am thinking and if I am thinking, then, I exist, then, I exist.

That is, at issue is whether (6) is true:

(6) I am certain that if I am thinking and if I am thinking, then, I exist, then, I exist.⁵

A Doxastic Burden Principle

The weapons for defeating (6) come from Descartes’ own arsenal. One is a doxastic burden principle (DBP) requiring ability to accept what is true and reject what is not true of a proposition in order for it to have an epistemic property.

⁵ Descartes applies “*si ferme et si assurée*” to the entire *Cogito*. Thus, the text does not support placing the epistemic operator this way: “If I am thinking and if I am thinking, then I exist, then I am certain that I exist.” The *Cogito* does not read “I am thinking, therefore I am certain that I exist.” Whether it should have and if so what difference it would make are not issues that can be pursued here.

(DBP) *P* has an epistemic property for person *S* only if *S* is able to accept what is true of *P* and reject what is not true of *P*.⁶

DBP can be derived from Descartes' epistemic principle of clear and distinct perceptions:

Principles I.45, Latin original (AT, 22): *Etenim ad perceptionem, cui certum & indubitatum iudicium possit inniti, non modò requiritur ut sit clara, sed etiam ut sit distincta.*

Principles I.45, French translation (*Œuvres Complètes*, Kindle Edition, 735): *Car la connaissance sur laquelle on peut établir un jugement indubitable doit être non seulement clair, mais aussi distincte.*

Haldane 1970 [1911], 237: For the knowledge upon which a certain and incontrovertible judgment can be formed, should not alone be clear but also distinct.⁷

Veitch 1912, 33: For the knowledge upon which we can establish a certain and indubitable judgment must be not only clear, but also, distinct.

Cottingham 1985, 207: A perception which can serve as the basis for a certain and indubitable judgment needs to be not only clear but also distinct.

This language suggests the following:

(CDP) *P* is certain for person *S* only if *S* clearly and distinctly perceives *P*.

We can derive DPB from CDP in a few easy steps:

(a) *S* clearly and distinctly perceives *P* only if *S* accepts *P*.

(b) *S* accepts *P* only if *S* can accept *P*.

(c) *S* can accept *P* only if *S* is able to accept what is true of *P* and reject what is not true of *P*.

Therefore, (d) *P* is certain for *S* only if *S* is able to accept what is true of *P* and reject what is not true of *P*.

We can reasonably generalize (d) to epistemic properties as such:

(DBP) *P* has an epistemic property⁸ for person *S* only if *S* is able to accept what is true of *P* and reject what is not true of *P*.

⁶ *P* can be a proposition, sentence, statement and even a thought, opinion or judgment in the occurrent sense. The bearers of epistemic properties for Descartes is too complex an issue to discuss in this article.

⁷ Haldane states (202) that her translation "... is made from the Latin version collated with the French," which is why I'm including the French translation of the Latin original.

⁸ My arguments in SC and QC rely on certainty in the epistemic sense in both CDP and DBP. For Descartes, metaphysical certainty must also be considered. For a comment on metaphysical certainty, see Objection 5 below.

Arnold Cusmariu

Applying DBP

Inspection shows that I-VII below are true of (4)

(4) If I am thinking and if I am thinking, then, I exist, then, I exist,

while their negations are not. The list can easily be expanded.

- I. The two occurrences of "if" are synonymous.⁹
- II. The first occurrence of "if" pairs up with the second rather than the first occurrence of "then."
- III. The two occurrences of "am" are synonymous.
- IV. The two occurrences of "thinking" are synonymous.
- V. The two occurrences of "then" are synonymous.
- VI. The two occurrences of "exist" are synonymous.
- VII. The four occurrences of "I" are co-referential.

Applying DBP yields that (4) has an epistemic property for *S* only if *S* is able to accept I-VII, which are true of (4), and reject the negations of I-VII, which are not true of (4).

Descartes' Evil Genius

The other weapon for defeating the certainty of the *Discourse Cogito* is Descartes' Evil Genius (EG), characterized at the end of *Meditation I*:

Heffernan 1990, 96: *Supponam igitur non optimum Deum, fontem veritatis, sed genium aliquem malignum, eundemque summe potentem & callidum, omnem suam industriam in eo posuisse, ut me falleret.*

Cress 1979, 16: Thus I will suppose not a supremely good God, the source of truth, but rather an evil genius, as clever and deceitful as he is powerful, who has directed his entire effort to misleading me.

A powerful and deceitful EG presumably can bring about the following states of affairs:

- (i) a person accepting what is not true;
- (ii) a person rejecting what is true;
- (iii) a person being unable to accept what is true;
- (iv) a person being unable to reject what is not true.

⁹ This article is not the place to address Quine's well-known concerns about synonymy (Quine 1950.) See Cusmariu 1978B, Cusmariu 1982 and Cusmariu 1983.

For purposes of my argument, it is sufficient that the EG be able to bring about states of affairs (iii) and (iv), which need not involve causing doubt (of any kind) in a person.

Three Scenarios

The list of I-VII represents opportunities that the EG can use to cause *S* to run afoul of the doxastic burden principle DBP. An EG who has the power to impact a person's mental life in such a way as to bring about states of affairs (iii) and (iv), presumably has the power to impact a person's mental life in such a way as to cause a person to experience memory lapses while uttering or thinking (4) to himself,

(4) If I am thinking and if I am thinking, then, I exist, then, I exist.

Here are four memory-lapse scenarios under states of affairs (iii) and (iv):

Scenario 1: *S* utters or thinks to himself "if" at the beginning of (4), then continues with "I am thinking and" but just as he is about to utter or think to himself "if" as the next word after "and," the EG immediately causes *S* to experience a memory lapse so that when *S* utters or thinks to himself the "if" after "and," *S* is unable to remember that this "if" is synonymous with the "if" he already uttered or thought to himself, and as a result *S* is unable (iii) to accept what is true of (4), that the two occurrences of "if" are synonymous; and (iv) to reject what is not true of (4), that the two occurrences of "if" are not synonymous.

Scenario 2: *S* utters or thinks to himself the first occurrences of "am," "thinking," "then," and "exist" in (4) but then the EG immediately causes *S* to experience memory lapses so that when *S* comes to the point of uttering or thinking to himself the second occurrences of these terms, *S* is unable to remember that they are synonymous with their first occurrences he already uttered or thought to himself, and as a result *S* is unable (iii) to accept what is true of (4), that the two occurrences of each of these terms are synonymous; and (iv) to reject what is not true of (4), that the two occurrences of these terms are not synonymous.

Scenario 3: *S* utters or thinks to himself "If I" at the beginning of (4), realizes that "I" refers to himself, then *S* continues with "am thinking and if" but just as he is about to utter or think to himself "I am," the EG immediately causes *S* to experience a memory lapse so that when *S* utters or thinks to himself "I am," *S* is unable to remember that this second "I" is co-referential with the one he already uttered or thought to himself. The EG repeats this process for the remaining two occurrences of "I," and as a result *S* is unable (iii) to accept what is true of (4), that the four occurrences of "I" refer to himself; and (iv) to reject what is not true of (4), that the four occurrences of "I" do not refer to himself.¹⁰

¹⁰ There are two basic approaches to the semantics of indexicals: *utterance-based* (Reichenbach 1947, Burks 1949) and *expression-based* (Kaplan 1989)—see Georgi in the *Internet Encyclopedia of Philosophy*. My use of "utterance" language does not mean I am relying on a specific analysis of either kind. I am merely following Descartes' lead in running my argument in terms of utterances. Thus, in *Meditation II* Descartes writes (Heffernan Bilingual Edition

Arnold Cusmariu

The *Cogito* Sentence is Not “*Ferme et Assurée*”

The doxastic burden principle DBP together with inability to accept what is true and reject what is not true of (4)

(4) If I am thinking and if I am thinking, then I exist, then I exist,

imply that (6)

(6) I am certain that if I am thinking and if I am thinking, then I exist, then I exist,

is false. The *Cogito* is valid in SC, the argument can be paraphrased as a true sentence so that “*cette vérité*” can be asserted, but the certainty Descartes claimed is shot down by flak from his own creation, the Evil Genius, with a key assist from a principle derived from Descartes’ epistemic principle of clear and distinct perceptions.

A *Cogito* Valid in the Quantificational Calculus (QC)¹¹

Preliminary

It may well be objected that a premise needed for a valid derivation in SC, “If I am thinking, then I exist,” could lead to a sound argument only if this premise is true in the general case. To answer this objection, a universal material conditional linking occurrent thought to the thinker’s existence must be added to the argument.

(D) Anything that is thinking, exists.

The deductive resources of SC, however, are not adequate to yield a valid inference from a premise in universal form to a conclusion in singular form. To bring about such an inference, a quantificational rule is necessary, specifically, Universal Instantiation (UI). We start with an explanation of UI – what it is, how it works, and what is involved in its correct application.

1990: 100) “... Ego sum, ego existo, *quoties a me profertur* [as it is uttered by me], *vel mente concipitur* [or conceived by the mind] *necessario esse verum*.” The distinction between utterance-based and expression-based semantics matters to the third scenario but not the first two. In any case, the EG should have no difficulty defeating an expression-based indexical theory.

¹¹ The term “quantificational calculus” (QC) seems to me preferable to “quantification theory” (it’s not a theory); “predicate logic” and “predicate calculus” (it’s not just about predicates); or “functional calculus” (functions already have a meaning in mathematics.) See LeBlanc 1966: 83. A recent historical overview of QC is Ferreiros 2001. For the sake of uncluttered text, quotation marks have been kept to a minimum and no use has been made of Quine corner quotes (Quine 1940).

A *Cogito* Valid in QC

The conclusion of the *Cogito* is supposed to follow by **UI** and **MP**:

(1) $(x)(\text{If } x \text{ is thinking, then } x \text{ exists})$	Translation of (D) in QC
\therefore (2) If I am thinking, then I exist.	1, UI
(3) I am thinking.	Assumption
\therefore (4) I exist.	2, 3 MP

UI in QC

The deductive resources of QC include rules of inference from SC such as **MP** and four rules governing universal and existential quantification. Here are **UI** details:

Formalism: $(x)\Phi x / \therefore \Phi a$

Interpretation: From a closed universally quantified wff of the form $(x)\Phi x$, infer a singular closed wff of the form Φa .

Application Sequence: It will be useful to present the application of **UI** to a specific case as a sequence of steps.¹²

Step 1: Determine that the wff to be used as the premise in a potential **UI** argument is a closed universally quantified wff of the form $(x)\Phi x$.

Step 2: Delete the (objectual) universal quantifier binding occurrences of the individual variable x in the closed universally quantified premise of the form $(x)\Phi x$ to obtain an open wff of the form Φx .

Step 3: Write a closed wff of the form Φa from Φx of Step 2 such that:

- (i) a in Φa is an individual constant in the signature of QC;
- (ii) occurrences of a in Φa are of the same individual constant;
- (iii) occurrences of a in Φa replace occurrences of the same individual variable x in Φx ;
- (iv) occurrences of a are in Φa at exactly those places where occurrences of x are in Φx ;
- (v) reference of occurrences of a in Φa has been fixed so they designate the same object in the universe of discourse (domain of quantification) of QC.¹³

Step 4: Enter Φa on a line below $(x)\Phi x$.

¹² Application steps are my own formulation.

¹³ Individual variables and individual constants are non-logical symbols in the signature of QC, as opposed to logical symbols such as material implication and negation. This article is not the place for a precise characterization of the distinction between logical and non-logical symbols, regarding which Tarski wrote (1983, 419) "... no objective grounds are known to me which permit us to draw a sharp boundary between the two groups of terms." See also Peacocke 1976.

Comments on UI Application Steps

- The terms “individual variable” and “individual constant” are standard in logic textbooks, e.g., Hilbert and Ackermann 1950 (1928), 102¹⁴; Kleene 1950, 436; Carnap 1956, 4; Church 1956, 168; LeBlanc 1966, 83, 84, 92; Schoenfield 1967, 10, 12; Copi 1967, 278-9; Thomason 1970, 209; Hunter 1971, 137; Takeuti 1975, 192; Curry 1977, 316; Mendelson 1979, 46; Kalish, Montague and Mar 1980, 440-1; Simco and James 1983, 122-3; Enderton 2001, 70; Hurley 2008, 407-409, 416¹⁵; and Smullyan 2014, 137.¹⁶
- Step 3i captures a basic syntactic fact about the signature of QC: An individual constant must come from the signature of QC to be used in a quantificational rule.
- The reader can easily verify that validity requires Steps 3ii-3iv.
- This is also true of Step 3v. Without this step, the possibility is left open that $(x)(Fx \rightarrow Gx)$ is true while a substitution instance $Fa \rightarrow Ga$ is false because the reference of the two occurrences of the individual constant a has not been fixed to denote the same object, rendering the inference of $Fa \rightarrow Ga$ from $(x)(Fx \rightarrow Gx)$ invalid. This could happen even though Steps 3ii-3iv are satisfied.
- Step 3v shows that **UI** cannot be formulated in purely syntactic terms. What this means and whether the same is true of the other three quantificational rules of QC are technical questions beyond the scope of this article.

What Is an Individual Constant?

Here are points relevant to the *Discourse Cogito* as a derivation sanctioned by **UI**.

- Symbolized using lower-case letters from the beginning of the alphabet $a, b, c \dots$ individual constants enable us to express in symbols that an object in the universe of discourse of a formal theory satisfies an open wff of the form Φx , which can be expressed by writing a closed wff of the form Φa .

¹⁴ In Editor's Notes, Luce writes (Hilbert and Ackerman 1950, 168): "It is not always understood that a constant, like a variable, is a *symbol*, a linguistic expression, but with the important distinction that a constant has a fixed designation, which remains unaltered throughout the discussion in which the constant appears; whereas a variable designates ambiguously, so to speak, assuming any one of a range of values." The main text where the terms "constant" and "variable" occur does not contain these explanations.

¹⁵ Hurley also uses "instantial letter," which he explains in the Glossary (677) as "the letter (variable or constant) introduced by universal instantiation or existential instantiation." The main text does not include this explanation. Presumably "instantial letter" covers both individual variables and constants, so that a better term would have been "instantial symbol."

¹⁶ While the term "variable" is as old as algebra and mathematics generally distinguishes between "real variables" ("free variables" in logic) and "apparent variables" ("bound variables" in logic), the term "individual variable" is from logic. Mathematics does not use the term "constant," speaking instead of "values" of a variable. Construing individual constants as "values" of individual variables, as symbols denoting objects in the universe of discourse of QC is hardly a definition of "individual constant."

- Individual constants enable valid derivations of, for example, wffs about objects taken singly, Φa , from wffs about objects taken collectively, $(x)\Phi x$. All axioms and some theorems are in universally quantified form.
- Individual constants are needed to express property instantiation, e.g., instantiation of *humanity* and *mortality* in an inference from a sentence of the form $(x)(Fx \rightarrow Gx)$ such as “Whatever is human, is mortal” to a sentence of the form $Fa \rightarrow Ga$ such as “If Socrates is human, then Socrates is mortal.”
- By way of definition of what it means for occurrences of an individual constant to be occurrences of the same individual constant—**UI** application Step 3ii—we will stipulate that this is true if and only if individual constant occurrences replace or are replaced by the same individual variables. This is not intended as a general definition, however, because occurrences of an individual constant can be occurrences of the same individual constant for reasons unrelated to substitutivity of individual variables or any other connection to rules of quantification.
- We will stipulate that individual variable occurrences in Ψ are occurrences of the same individual variable—**UI** application Step 3iii—if and only if the same quantifier binds all individual variable occurrences if Ψ is closed or would bind them all if Ψ is open.
- Individual constants in QC are “referentially anonymous” in the sense that they only designate some object or other in the universe of discourse of QC, to mirror the fact that the existential quantifier likewise asserts the existence only of some object or other in the universe of discourse of QC.
- Once referentially anonymous individual constants are replaced with referentially transparent constants such as numerals and proper names, application Step 3v of **UI** applies. The reference of occurrences of such individual constants needs to be fixed so they consistently designate the same object in the universe of discourse of QC. How this can be done will be discussed below.

First Preliminary Application

A **UI** derivation in arithmetic will help identify issues relevant to the *Cogito* argument.

Properties of addition, subtraction and equality allow us to write:

$$(A) (x)(y)(z)(x + y = z \rightarrow x = z - y)$$

UI application instructions should enable us to prove that (A) logically implies (C):

$$(C) (7 + 5 = 12) \rightarrow (7 = 12 - 5)$$

Step 1: Determine that the wff to be used as the premise in a potential **UI** argument is a closed universally quantified wff of the form $(x)\Phi x$.

QC does not recognize (A) as a wff because the symbols for addition and subtraction are not in its signature. For the time being, however, let us stipulate that these symbols have been added to the signature of QC so that a universally

quantified sentence of arithmetic such as (A) can be considered a wff of the form $(x)\Phi x$ in QC.

Step 1 may be understood broadly to include (A) as wff of the form $(x)\Phi x$ because the following assumptions are reasonable:

- The material conditional arrow; quantifiers; individual variables and their occurrences; left and right hand brackets have the same meaning in (A) that they have in QC.
- Individual variable occurrences are bound by quantifiers in (A) exactly as they are in QC.

Step 2: Delete the (objectual) universal quantifier binding occurrences of the individual variable x in the close universally quantified premise of the form $(x)\Phi x$ to obtain an open wff of the form Φx .

Having stipulated that (A) is a wff of the form $(x)\Phi x$, Step 2 can be applied to (A), resulting in (B):

(A) $(x)(y)(z)((x + y = z) \rightarrow (x = z - y))$

\therefore (B) $(x + y = z \rightarrow x = z - y)$ 1, **UI** Step 2

Next, we decide whether Step 3 allows the argument to move from (B) to (C):

(B) $(x + y = z) \rightarrow (x = z - y)$

\therefore (C) $(7 + 5 = 12) \rightarrow (7 = 12 - 5)$

Step 3: Write a closed wff of the form Φa from Φx of Step 2 such that:

- a in Φa is any individual constant in the signature of QC;
- occurrences of a in Φa are of the same individual constant;
- occurrences of a in Φa replaces occurrences of the same individual variable x in Φx ;
- occurrences of a are in Φa at exactly those places where occurrences of x are in Φx ;
- reference of occurrences of a in Φa has been fixed so they designate the same individual in the universe of discourse (domain of quantification) of QC.

However, Step 3 can be applied to (B) to yield (C) provided:

- It can be explained how numerals can be considered individual constants in the signature of QC. For the time being, let us stipulate that the signature of QC has been expanded so as to include numerals as individual symbols. Let us stipulate further that integers have been added to the universe of discourse of QC.
- It can be reasonably stipulated that the two occurrences of the numerals 7, 5 and 12 in (C) are occurrences of the same numerals. This can be done on grounds that they replace occurrences of the same respective individual variables.

(iii) It can also be reasonably stipulated that occurrences of the numerals 7, 5 and 12 in (C) replaced occurrences of the same individual variables x, y and z , respectively, in (B). This can be done on grounds that each variable is bound by the same universal quantifier.

(iv) It can be seen by inspection that (C) contains occurrences of 7, 5 and 12 at exactly those places where (B) contains occurrences of x, y and z , respectively.

(v) Finally, it can reasonably be stipulated that the reference of occurrences of 7, 5 and 12 in (C) has been fixed so that they are co-referential. How exactly this is done need not be spelled out for present purposes.

All five conditions of Step 3 having been met, we can write

$$(x)(y)(z)((x + y = z) \rightarrow (x = z - y)) \ / \therefore (7 + 5 = 12) \rightarrow (7 = 12 - 5)^{17}$$

Second Preliminary Application

Frege and his followers built modern logic to provide a rigorous foundation for mathematical reasoning, which is why the above application of **UI** was straightforward. However, when Russell used his expansion of the uniqueness quantifier in $\exists! \Phi x$ to formulate his theory of descriptions (Russell 1905) and then apply it to “puzzles”¹⁸ (Russell 1956, 47), a precedent was set that analytic philosophy has followed ever since. The methods of formal logic have been used to help elucidate philosophical problems despite significant differences between formal languages and the vernacular, the language in which philosophical problems are stated. Ramsey agreed, describing what started life as notational convenience as a “paradigm of philosophy” (Ramsey 1965, 263). In his *Tractatus* (1963 [1921]), Wittgenstein applied the lessons of the theory of descriptions to philosophy as a whole. In a way, the *Tractatus* is the theory of descriptions “on steroids.”

As a test case closer to the *Discourse Cogito*, let us look at an argument that is intuitively valid and see what is involved in applying **UI** to confirm validity.

(A) $(x)(x \text{ is human} \rightarrow x \text{ is mortal})$

\therefore (B) Socrates is human \rightarrow Socrates is mortal

According to the standard understanding of **UI**, if (A) is true, it follows that every substitution instance of (A) that has the requisite form will be true. So, is (B) a substitution instance of (A), but does it have the requisite form? We follow application steps as specified above.

¹⁷ This sentence can be converted into a material conditional but there is no need to do that.

¹⁸ Omitted from the list of “puzzles” Russell claimed his theory of descriptions (TD) solves is his own paradox. TD allowed Russell to formulate his “no class” solution to the paradox; but then he realized there is also a version of the paradox for what he called “propositional functions, and it was back to square one; the theory of types followed.

Step 1: Determine that the wff to be used as the premise in a potential **UI** argument is a closed universally quantified wff of the form $(x)\Phi x$.

QC does not recognize (A) as a wff. (A) can be regarded as a substitution instance of a closed universally quantified sentence of the form $(x)\Phi x$ that is recognized as a wff in QC,

$$(C) (x)(Fx \rightarrow Gx),$$

provided

(i) the universal quantifier in (A) functions the way it functions in (C), as do individual variables;

(ii) the material implication arrow captured the ordinary meaning of “if ... then ...” that makes vernacular versions of (A) true;

(iii) the predicates “is human” and “is mortal” in (A) can be considered substitution instances of “*F*” and “*G*,” respectively, in (C).

For the sake of argument, let us stipulate that all these conditions have been met. We will leave it to a Substitution Theorem¹⁹ to spell out the precise relationship between (C) and (A) and move on to application Step 2 of **UI**.

Step 2: Delete the (objectual) universal quantifier binding occurrences of the individual variable x in the close universally quantified premise of the form $(x)\Phi x$ to obtain an open wff of the form Φx .

Having agreed that (A) can be regarded as a substitution instance of a wff recognized as such in QC, Step 2 can be applied to (A), resulting in (A1):

$$(A) (x)(x \text{ is human} \rightarrow x \text{ is mortal})$$

$$\therefore (A1) x \text{ is human} \rightarrow x \text{ is mortal} \quad \text{Step 2 of UI}$$

Let us move on to Step 3 of **UI** to decide the validity of this inference:

$$(A1) x \text{ is human} \rightarrow x \text{ is mortal}$$

$$\therefore (B) \text{Socrates is human} \rightarrow \text{Socrates is mortal}$$

Step 3: Write a closed wff of the form Φa from Φx of Step 2 such that:

(i) a in Φa is any individual constant in the signature of QC;

(ii) occurrences of a in Φa are of the same individual constant;

(iii) occurrences of a in Φa replaces occurrences of the same individual variable x in Φx ;

¹⁹ For the Substitution Theorem, see Kleene 1967: 14. For rules of substitution, see LeBlanc 1966: 137. The idea of a rule of substitution comes from Frege. See Zalta 2018A and 2018B. Gödel has stated (Schilpp 1944, 126): “In *Principia*, eliminations are always carried out by substitutions in the theorems corresponding to definitions, so that it is chiefly the rule of substitution which would have to be proved.”

(iv) occurrences of a are in Φa at exactly those places where occurrences of x are in Φx ;

(v) reference of occurrences of a in Φa has been fixed so they designate the same object in the universe of discourse (domain of quantification) of QC.

However, applying Step 3 to (A1) will result in (B) provided:

(I) It can be explained how proper names can be considered individual constants in the signature of QC.

(II) We can reasonably stipulate that the two occurrences of "Socrates" in (B) are occurrences of the same proper name. See below.

(III) We can reasonable stipulate that the two occurrences of "Socrates" in (B) replaced occurrences of the same individual variable x in (A1). See below.

(IV) It can be shown that (B) contains occurrences of "Socrates" at exactly those places where (A1) contains occurrences of x . Inspection shows this is the case.

(V) The reference of the two "Socrates" occurrences of (B) has been fixed so that they are co-referential.

Ordinary Proper Names as Individual Constants

Ordinary proper names (OPNs) can be included as individual constants in a QC signature based on the designative function they already have in natural language. There are three questions that must be answered.

Question 1: How do OPNs acquire a designative function?

Answer: Reference fixing for OPNs uses one or more definite descriptions to assign designation. A definite description is an open sentence of the form "the x such that Fx ," which entails uniqueness according to Russell's analysis (Russell 1905), formalized in *Principia Mathematica* (Russell and Whitehead 1911, 31-2). Thus, under the interpretation of "Socrates" as "the name of the protagonist of Plato's dialogues," the sentence "Socrates is human" is true. On the other hand, under the interpretation of "Socrates" as "the name of the shaggy mutt down the street," the sentence "Socrates is human" is false. Under the interpretation of "Socrates" as "the name formed by letters S-o-c-r-a-t-e-s of the English alphabet," the sentence "Socrates is human" is either false, nonsense or lacks a truth value, depending on one's theory.

An OPN designates the object satisfying the definite description. For example, if an interpretation assigns the definite description "the author of *Waverly*" to the OPN "Scott" in a universe of discourse that includes persons, then, the OPN "Scott" is stipulated to designate the object satisfying the definite description "the author of *Waverly*." Thus, it is true under this interpretation that Scott is identical with the author of *Waverly* and false that Scott is identical with the author of *Hamlet*.

Question 2: Must the designation of each OPN occurrence be fixed separately?

Answer: No. Consider as an example “If Socrates is human, then Socrates is mortal.” In order for this sentence to be true, it is sufficient that the same interpretation be assigned to the two occurrences of “Socrates,” which is what happens in ordinary contexts of use.

Question 3: How can the reference of OPN occurrences be fixed so they are co-referential.

Answer: The reference of OPN occurrences can be fixed so they are co-referential if and only if the same interpretation, e.g., definite description, determines the identity of the object to which they refer. Note that from the fact that two OPN occurrences are of the same OPN, it does not follow that they are co-referential. In one and the same true sentence, “Socrates” might denote the Greek philosopher and the shaggy dog down the street its owners happened to name after the Greek philosopher.

Having satisfied all conditions of Step 3, we can now write

$$(x)(x \text{ is human} \rightarrow x \text{ is mortal}) / \therefore \text{Socrates is human} \rightarrow \text{Socrates is mortal}^{20}$$

Applying UI to the *Cogito*

Here is our *Cogito* argument again:

(D) Anything that is thinking, exists.

(1) $(x)(x \text{ is thinking} \rightarrow x \text{ exists})$ Translation of (D)

\therefore (2) I am thinking \rightarrow I exist. 1, **UI**

(3) I am thinking. Assumption

\therefore (4) I exist. 2, 3 **MP**

Let us follow **UI** application instructions one step at a time.

Step 1: Determine that the wff to be used as the premise in a potential **UI** argument is a closed universally quantified wff of the form $(x)\Phi x$.

QC does not recognize (1) as a wff. The issue is whether (1) can be included in our **UI** argument as a substitution instance of a closed universally quantified sentence of the form $(x)\Phi x$ that is recognized as a wff in QC, namely,

(C) $(x)(Fx \rightarrow Gx)$.

Application Step 1 of **UI** can be applied to (1) understood as a substitution instance of (C) provided familiar conditions are satisfied:

(i) the universal quantifier in (1) functions the way it functions in (C), as do individual variables;

(ii) the material implication arrow captures the ordinary meaning of “if ... then ...” that makes (1) true and (1) as a correct translation of (D);

²⁰ There is also no need to convert this sentence into a material conditional.

(iii) the predicates “is thinking” and “exists” in (1) can be considered substitution instances of “*F*” and “*G*,” respectively, in (C).²¹

We will leave it to a Substitution Theorem again to spell out the precise relationship between (C) and (1) and move on to the Step 2 of **UI**.

Step 2: Delete the (objectual) universal quantifier binding occurrences of the individual variable *x* in the close universally quantified premise of the form $(x)\Phi x$ to obtain an open wff of the form Φx .

Let us stipulate that (1) is a substitution instance of a wff of the form $(x)\Phi x$ recognized as such in QC, namely, (C), so Step 2 can be applied to (1), resulting in (1a):

(1) $(x)(x \text{ is thinking} \rightarrow x \text{ exists})$

\therefore (1a) $x \text{ is thinking} \rightarrow x \text{ exists}$ 1, **UI Step 2**

Next, we decide whether Step 3 allows the *Cogito* argument to move from (1a) to (2):

(1a) $x \text{ is thinking} \rightarrow x \text{ exists}$

\therefore (2) $I \text{ am thinking} \rightarrow I \text{ exist}$.

Step 3: Write a closed wff of the form Φa from Φx of Step 2 such that:

- (i) *a* in Φa is any individual constant in the signature of QC;
- (ii) occurrences of *a* in Φa are of the same individual constant;
- (iii) occurrences of *a* in Φa replaces occurrences of the same individual variable *x* in Φx ;
- (iv) occurrences of *a* are in Φa at exactly those places where occurrences of *x* are in Φx ;
- (v) reference of occurrences of *a* in Φa has been fixed so they designate the same object in the universe of discourse (domain of quantification) of QC.

However, Step 3 can be applied to (1a) to yield (2) provided:

- (i) It can be explained how the indexical “I” can be considered an individual constant in the signature of QC, which is not the case at the moment.

²¹ Kant famously objected that existence was not a “real predicate,” though he probably meant “property.” Either way, I will not raise this issue to question a *Cogito* derivation in QC. That said, we should not confuse, as Kant seems to be doing, property existence with property exemplification. Under Platonism, a property need not be a property of anything, a view Aristotle denied. My work on property ontology stems from my Ph.D. dissertation, Cusmariu 1977. See Cusmariu 1978A, B, and C; Cusmariu 1979A and B; Cusmariu 1980; Cusmariu 1985; and Cusmariu 2016C. Platonism has had a significant influence on my artwork as a sculptor. See Cusmariu 2009; Cusmariu 2015A and B; and Cusmariu 2017A and B.

(ii) We can reasonably stipulate that the two occurrences of “I” in (2) are occurrences of the same indexical. This can be done on the familiar grounds that they replace occurrences of the same individual variable, x , in (1).

(iii) We can also reasonably stipulate that occurrences of “I” in (2) replaced occurrences of the same individual variable x in (1a). This can be done on the familiar grounds that these individual variable occurrences are bound by the same quantifier.

(iv) We can show that (2) contains occurrences of “I” at exactly those places where (1a) contains occurrences of x . Inspection shows that this is the case.

(v) It can be explained how the reference of the two “I” occurrences in (2) can be fixed so that they are co-referential.

Is “I” an Individual Constant In QC?

The answer is obviously “no” for a purely technical reason: There is no list of individual constants in any QC signatures that looks like this: $a, b, I, me, my, mine, c, d, \dots$

Philosophically speaking, the absence of “I” and its cognates is not surprising:

- Frege’s decisively refuted psychologism, rejecting the view that truths of logic and mathematics are about, or depend upon occurrent or dispositional mental states of persons. Here Frege was in agreement with Plato and Aristotle as well as many later logicians.²²
- Persons and their mental states are not in the universe of discourse of formal systems.²³ Set theory, for example, would have no use for “I” and its cognates (“me,” “my,” “mine”) as individual constants, there being nothing for them to designate, not even the empty set—already symbolized by $\{\}$ or \emptyset .

²² In a 1918 essay titled “The Thought: A Logical Inquiry,” Frege remarked (Klemke 1968, 517): “The same utterance containing the word ‘I’ will express different thoughts in the mouths of different men, of which some may be true, others false. The occurrence of the word ‘I’ in a sentence gives rise to some questions.” Unfortunately, Frege did not go on to consider whether “I” is or could be an individual constant and if so under what conditions. Frege also did not explore connections between his view of “I” and the *Cogito*. The focus of his logic treatises of decades earlier was entirely different, so the issues did not arise.

²³ The only exception of which I am aware is my own work. In Cusmariu 2012 and Cusmariu 2016, I introduce semantic evidence predicates into the metalanguage of science and mathematics, which means that persons and their beliefs are included in the universe of discourse of these disciplines. However, this is consistent with Frege’s critique of psychologism. Neither the truth value nor the epistemic value of scientific and mathematical beliefs is in any sense “subjective.”

"I" As an Individual Constant

First-person indexicals (FPIs) such as "I," "me," "my" and "mine" can be included as individual constants in a QC signature under the same conditions that allow ordinary proper names to be included. The same three questions apply.

Question 1: How do FPIs acquire a designative function in QC?

Answer: It is sufficient for present purposes to stipulate that FPIs have a designative function in QC if and only if there is a definite description of form "the x such that Fx " such that an FPI designates the object that satisfies "the x such that Fx ." What this definite description might be is context dependent, as many have pointed out, so it can be left to the context to specify it. For example, for *Discourse Cogito* purposes, this definite description can have the form "the person uttering or thinking sentence P to himself." Readers may pick whatever analysis of indexicals they wish.²⁴

Question 2: Must the reference of each FPI occurrence of be fixed separately?

Answer: No. One and the same definite description can be used to fix the reference of all FPI occurrences.

Question 3: How can the reference of FPI occurrences be fixed so they are co-referential?

Answer: Provided the same definite description is used each time in the context at hand, co-referentiality will follow as a matter of course.

Thus, we can sidestep worries such as whether FPIs have sense as well as reference; and if they have both, what relations hold between them; whether they are directly or indirectly referential (Kaplan 1989, 523); and so on.

The QC Component of the *Discourse Cogito* as a Sentence

(1) $(x)(x \text{ is thinking} \rightarrow x \text{ exists})$ Assumption

\therefore (2) I am thinking \rightarrow I exist. 1, UI

We start by writing first (1) \therefore (2) horizontally as one sentence:

(3) $(x)(x \text{ is thinking} \rightarrow x \text{ exists}) / \therefore$ I am thinking \rightarrow I exist.

A material conditional also corresponds to the argument in (3), whose antecedent is the premise of the argument, (1), and whose consequent is the conclusion of the argument, (2), so we can write (4)

(4) $(x)(x \text{ is thinking} \rightarrow x \text{ exists}) \rightarrow (\text{I am thinking} \rightarrow \text{I exist}).$

²⁴ See, for example, Reichenbach 1947; Burks 1949; Castaneda 1966; Kamp 1971; Kaplan 1989; García-Carpintero 1998; King 2001; Perry 2001; Salmon 2005; Soames 2005; and Cappelen and Dever 2013. Rosenkrantz 1993 discusses the broader metaphysical issues involved.

Syllogistic Form Is Inadequate

Descartes would have been able to recognize (4) only if stated in syllogistic form. To see if this is possible, we first do away with quantificational notation and the material conditional in the antecedent of (4) and write the result as an **A**-proposition,

(4a) All persons who are thinking are persons who exist.

Next, we do away with the material conditional in the consequent of (4) and shoehorn the rest into an **A**-proposition as well,

(4b) All persons identical with me who are thinking are persons identical with me who exist.

However, the two truth-functional components of (4) are connected by means of a material conditional. Replacing it with a copula to connect (4a) and (4b) results in gibberish, as the reader can readily see, proving the inability of syllogistic logic to capture anything more complicated than subject-predicate sentences. Syllogistic logic is inadequate for mathematical purposes in the sense that the sixteen valid syllogisms (rules) of syllogistic logic are insufficient to allow mathematical proofs to go through.²⁵

Eliminating Quantificational Notation

An adequate paraphrase of (4) must capture its three material conditionals:

(5) If it is true that if anyone who is thinking, then that person exists, then, it is true that if I am thinking, then I exist.

Applying "*ferme et assurée*" to (5) yields (6):

(6) It is certain for a person that if it is true that if anyone who is thinking, then that person exists, then, it is true that, if I am thinking, then I exist.

A Doxastic Burden Principle

The doxastic burden principle, DBP,

(DBP) *P* has an epistemic property for person *S* only if *S* is able to accept what is true of *P* and reject what is not true of *P*.

can be applied to show that (6) is false. Inspection shows that I-IX below are true of (5)

(5) If it is true that if anyone who is thinking, then that person exists, then, it is true that if I am thinking, then I exist,

²⁵ Capturing Euclid's well-known proof that the square root of 2 is irrational requires QC with the equality symbol, as the reader can easily verify. I have not studied Descartes' proofs and cannot say whether they are stated in syllogistic logic.

while the negations of I-IX are not. Applying DBP yields that (5) has an epistemic property for *S* only if *S* is able to accept I-IX and reject the negations of I-IX:

- I. The three occurrences of “if” are synonymous.
- II. The first and third occurrences of “that” are synonymous as propositional prefixes.
- III. The second occurrence of “that” is a demonstrative and is not synonymous with the first and third occurrences of “that.”
- IV. The two occurrences of “true,” “it” and “thinking” are synonymous.
- V. The three occurrences of “then” are synonymous.
- VI. The first occurrence of “if” is paired off with the second occurrence of “then.”
- VII. The second occurrence of “if” is paired off with the first occurrence of “then.”
- VIII. The two copula occurrences, “is” and “am,” are synonymous despite grammatical differences.
- IX. “Exist” and “exists” are synonymous despite grammatical differences.

The reader can easily imagine ways of expanding this list.

The Evil Genius

Recall that Descartes’ description at the end of *Meditation I* suggests that the EG has the power and intent to bring about four logically independent states of affairs:

- (i) a person accepting what is not true;
- (ii) a person rejecting what is true;
- (iii) a person being unable to accept what is true;
- (iv) a person being unable to reject what is not true.

Three Scenarios

An EG that has the power to impact a person’s mental life in such a way as to bring about (i)-(iv), presumably has the power to impact a person’s mental life in such a way as to cause a person to experience memory lapses while uttering or thinking (5) to himself,

- (5) If it is true that if anyone who is thinking, then that person exists, then, it is true that if I am thinking, then I exist.

Here are three memory-lapse scenarios under states of affairs (iii) and (iv):

Scenario 1: *S* utters or thinks to himself “If” at the beginning of (5), then continues with “it is true that,” but just as he is about to utter or think to himself the “if” after “that,” the EG immediately causes *S* to experience a memory lapse so that when *S* utters or thinks to himself the “if” after “that,” *S* is unable to remember that this “if” is synonymous with “If” at the beginning of (5) he already

uttered or thought to himself and as a result *S* is unable (iii) to accept what is true of (5), that the two occurrences of “if” are synonymous; and (iv) reject what is not true of (5), that the two occurrences of “if” are not synonymous.

Scenario 2: *S* utters or thinks to himself “that” after “true” in (5), then continues with “if anyone who is thinking, then” but just as he is about to utter or think to himself “that” after “then,” the EG, who has been eavesdropping, immediately causes *S* to experience a memory lapse so that when *S* utters or thinks to himself the “that” after “then,” *S* is unable to remember that the “that” after “true” is a propositional prefix while the “that” after “then” is a demonstrative, and as a result *S* is unable to (iii) accept what is true of (5), that the two occurrences of “that” are not synonymous; and (iv) reject what is not true of (5), that the two occurrences of “that” are synonymous.

Scenario 3: *S* utters or thinks to himself the first occurrences of “am,” “thinking,” “then,” and “exist” but then the EG, who has been eavesdropping, immediately causes *S* to experience memory lapses such that when *S* comes to the point of uttering or thinking to himself the second occurrences of these terms in (5), *S* is unable to remember that they are synonymous with their first occurrences and as a result *S* is unable to (iii) accept what is true of (5), that the two occurrences of each of these terms are synonymous; and (iv) reject what is not true of (5), that the two occurrences of these terms are not synonymous.

Cogito Certainty Has Not Been Achieved

Inspection shows that the doxastic burden principle DBP together with inability to accept what is true and reject what is not true of (5)

(5) If it is true that if anyone who is thinking, then that person exists, then, it is true that if I am thinking, then I exist,

imply that (6) is false,

(6) It is certain (for me) that if it is true that if anyone who is thinking, then that person exists, then, it is true that if I am thinking, then I exist.

In Conclusion

The short-term memory lapses I have described are not about whether the EG can bring it about that “*S* is nothing as long as *S* thinks that he is something.” Descartes is right about that. The scenarios raise an objection to the certainty of the *Discourse Cogito*. Indeed, *cette vérité*, “I am thinking, therefore I exist,” formulated as a valid argument and then as a sentence is a tautology and by implication a necessary truth; so, of course “the most extravagant suppositions of the skeptics could not shake it” for a reason already noted: Epistemic change does not entail semantic change. This does not mean, however, that skeptics cannot shake the *ferme et assurée* of *cette vérité*. They can, needing to be no more extravagant than Descartes’ own Evil Genius, aided and abetted by a principle derived from Descartes’ principle of clear and distinct perceptions. Without certainty, the *Discourse Cogito* is a failed response to skepticism.

Objections and Replies

Objection 1: Attributing truth and justification to arguments rather than their components is at relatively minor error. Descartes' readers would not have been confused about what he meant. In any case, he can live with a bit of unclarity given that the alternative is abandoning the *Cogito* and with it a response to skepticism. Your analysis is a case of "the cure is worse than the disease."

Reply: The doxastic burden principle DBP together with the EG's power to bring about memory lapses are sufficient to defeat the certainty of the *Discourse Cogito* in its original formulation, "I am thinking, therefore I exist."

Thus, consider this scenario: *S* utters or thinks to himself "I am" at the beginning of the *Cogito* sentence, then utters "thinking," then utters "therefore" and at that very instant, the EG, immediately causes *S* to be unable to remember as he utters or thinks to himself the "I" in "I exist" that it is co-referential with the "I" in "I am" and as a result *S* is unable (iii) to accept what is true of "I am thinking, therefore I exist," that the two occurrences of "I" are co-referential; and (iv) to reject what is not true of "I am thinking, therefore I exist," that the two occurrences of "I" are not co-referential. It follows that Descartes' "first principle of the philosophy I was seeking" is not certain, which dooms his response to skepticism. Needless to add, Descartes scholars would have regarded such a brief refutation of the *Discourse Cogito's* certainty as giving the matter, and Descartes himself by implication, short shrift and probably would have dismissed it as frivolous. They are less likely to do that after working through the technical details above, I hope.

Objection 2: Your critique may well be just a "one-off." Being the earliest (1637), perhaps the *Discourse Cogito* was in a sense superseded by later versions in *Meditations* (1641) and *Principles* (Latin, 1644; French 1647). Does your critique apply to those versions?

Reply: A thorough analytical comparison of the five *Cogito* versions²⁶ in light of results in this article would require book-length treatment. For now, I'd like to present textual evidence that *Cogito* versions in *Discourse* and *Principles* stand or fall together.

- The Latin original (AT VIII, 7) of *Principles* uses the epistemic terms terms "*hæc cognitio*" ("this knowledge") and "*est omnium prima & certissima*" ("is the first and most certain") in reference to a *Cogito* also stated in argument form: "*Ac proinde hæc cognitio, ego cogito, ergo sum, est omnium & certissima, quæ cuilibet ordine philosophanti occurat.*" Knowledge implies truth, so ...
- The French translation²⁷ (*Œuvres Complètes*, Kindle Edition, 696) is even closer to the *Discourse* version, using the terms "*cette conclusion*" ("this inference,")

²⁶ One in French (*Discourse*); one and Latin and one in French (*Meditations*; the French translation by de Luynes appeared in 1647); and one in Latin and one in French (*Principles*.)

²⁷ Descartes' friend Claude Picot translated *Principles* into French. See Gaukroger 1995, 138 and 386-7. In his preface to the French translation, titled "Author's letter to the translator," Descartes refers to it as "*nette et accomplie*," rendered variously as "elegant and finished"

"vraie" ("true,") and "la plus certaine" ("the most certain") in reference to the *Cogito*:²⁸ ... nous ne saurions nous empêcher de croire que cette conclusion: Je pense, donc je suis, ne soit vraie, et par conséquent la première et la plus certaine qui se présente à celui qui conduit ses pensées par ordre.

- A little later, the Latin original (AT VIII, 7) presents an inference formally similar to the *Cogito*, that proceeds from "seeing and walking" to the familiar ending "ergo sum" ("therefore I exist,") followed by a comment that uses an epistemic term "conclusio non est absolute certa" ("the conclusion is not absolutely certain"): *Nam si dicam, ego video, vel ego ambulo, ergo sum; & hoc intelligam de visione, aut ambulatione, quæ corpore peragitur, conclusio non est absolute certa.*
- The French translation (*Œuvres Complètes*, Kindle Edition, 698) also implies that inference was involved in the *Cogito*. Verb ("j'infère") ("I infer") and noun ("conclusion") ("inference") forms are used, as is the epistemic term "infaillible" ("certain," "infallible"): *Car si je dis que je vois ou que je marche, et que j'infère de là que je suis; si j'entends parler de l'action qui se fait avec mes yeux ou avec mes jambes, cette conclusion n'est pas tellement infaillible.*

Objection 3: What about the *Cogito* version in *Meditation II*?

Reply: This version reads (Cress 1979, 17) "... the statement 'I am, I exist' is necessarily true every time it is uttered by me or conceived in my mind." The doxastic burden principle DBP and the EG can jointly defeat the *Meditations Cogito* only if this version can be paraphrased into a sentence in epistemic form, by (a) replacing "necessarily true" with "certain" and (b) having "it is certain that" prefix the conditional "if I utter or conceive the statement 'I am, I exist,' then the statement 'I am, I exist' is true." However, it is beyond the scope of this paper to determine whether the *Meditations Cogito* text can support (a) and (b), or whether (a) and (b) are consistent with Descartes' other views.

Objection 4: You have shown only that Descartes failed to achieve certainty if his *Discourse Cogito* is interpreted as an argument; not that this *Cogito* must be interpreted as an argument. Thus, you have not ruled out interpretations of this *Cogito* that construe it as something other than an argument. Katz 1986 develops an interpretation according to which the *Cogito* (generically) is an example of "linguistic entailment"; while Sarkar 2003 thinks the *Cogito* is an "experiment."

Reply: Katz and Sarkar are book-length studies of approximately 200 and 300 pages, respectively, a luxury I do not have here. Accordingly, I have deliberately avoided sparring with competing views, concentrating instead on developing my own views and leaving polemical considerations for another time. The Katz and Sarkar interpretations seem to me mistaken but that is not something I can explain in an article of this scope. Briefly, however, we are asked to believe that Descartes had in mind "linguistic entailment" or "an experiment" when he said that the

(Veitch 1912, v), "polished and well-finished" (Haldane 1970, 203), and "polished and thorough" (Cottingham 1985, 179). Picot published two more translations after Descartes' death, in 1651 and 1657. I have not studied them.

²⁸ Haldane notes (1970, 202) that "the French version frequently differs considerably from the Latin." For details on differences between the two versions, see the AT IXb.

Cogito was “the first principle of the philosophy I was seeking,” which quite frankly seems incredible.

Objection 5: For Descartes, “certainty” has an epistemic sense, which rules out grounds for doubt, as well as a metaphysical sense, which rules out the very possibility of such grounds. How does metaphysical certainty affect your arguments?

Reply: It has no effect. The argument used to derive the doxastic burden principle DBP from an epistemic version of Descartes’ principle of clear and distinct perceptions (CDP) can also be used to derive a metaphysical certainty version of DBP, MDBP, from a metaphysical certainty version of CDP. It is a topic for another paper to decide whether the EG and MDBT together can also defeat the metaphysical certainty of *Discourse Cogito* sentences in SC and QC.

Objection 6: Descartes offers an alternative to deduction – intuition -- in Rule Three of *Rules for Guiding One’s Intelligence*, which he explains as (Clarke 1985, 14)

“... the conception of a clear and attentive mind, which is so easy and so distinct that there can be no room for doubt about what we are understanding. ... Thus everyone can mentally intuit that he exists, that he is thinking, that a triangle is bounded by just three lines, and a sphere by a single surface, and the like.”

What about this?

Reply: Descartes abandoned writing *Rules* in 1628, having completed eighteen rules that have a summary followed by text; and three rules that only have a summary. This material was published in 1701, 51 years after Descartes died.

- Because Descartes left *Rules* unfinished, its status as a text that potentially overrides what he says about thinking and existence in works he did finish, such as *Discourse*, *Meditations* and *Principles*, is hardly definitive.
- None of the *Cogito* versions in *Discourse*, *Meditations* and *Principles* mention intuition, suggesting Descartes had abandoned his earlier view.
- Why bother with a *Cogito* in argument form if “everyone can mentally intuit that he exists, that he is thinking”?
- The science-fiction film *The Matrix* shows virtual beings whose “clear and attentive minds” are unaware they are in a computer-generated world.

Objection 7: There is a sense of “certain” in which someone might say “I am certain that the Theory of Relativity is true even though I don’t understand what an inertial frame of reference is,” basing certainty on the opinion of a professional physicist. So, your doxastic burden principle DBP is false.

Reply: “No, you’re not certain that the Theory of Relativity is true. You’re just taking someone else’s word for it.” In any case, this is not an objection Descartes could raise. The concept of certainty he claims for the *Discourse Cogito* is not the hearsay concept. “I am certain that the *Cogito* is true because someone else thinks so” is absurd.

Objection 8: Your doxastic burden principle DBP potentially rules out certainty for any sentence whatever. It is easy to compile lists of items comparable to those

in your scenarios, go through each item, and show how the Evil Genius can sow confusion.

Reply: The scenarios I presented can indeed be modified to apply to a greater range of situations, significantly increasing the reach of the skeptical conclusion. For example, as someone utters or thinks to himself the last five letters of a ten-letter word, the EG interferes and causes that person to be unable to remember the first five letters of that word; or as someone types the first five letters of a word, the EG interferes and causes the person to be unable to remember the last five letters of the word; and so on. Loss of certainty would follow as above.

Descartes himself opened the door to radical skepticism, to which I do not have an answer at the moment. Not everyone, however, walked through the door. Back in my day as a graduate student at Brown University, whenever a member of Roderick Chisholm's epistemology seminar seemed about to raise a skeptical issue, Chisholm would cut him off and ask, "Are you a skeptic?" If the answer was "yes," he'd tell the person he didn't want to discuss skepticism and would simply move on.²⁹

Bibliography

- Ariew, Roger. 2011. *Descartes Among the Scholastics*, Ch. 10. Brill E-Book.
- Broughton, Janet. 2002. *Descartes's Method of Doubt*. Princeton: Princeton University Press.
- Burks, Arthur W. 1949. "Icon, Index, and Symbol." *Philosophy and Phenomenological Research* 9 (4): 673–689.
- Carnap, Rudolf. 1958. *Introduction to Symbolic Logic and its Applications*. New York: Dover Publications. This is a translation by William H. Meyer and John Wilkinson of *Einführung in die symbolische Logik* published by Springer in 1954.
- Cappelen, H. and Dever, J., 2013. *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. New York: Oxford University Press.
- Castañeda, Hector-Neri. 1966. "'He': A Study in the Logic of Self-Consciousness." *Ratio* 8: 130–157.
- Church, Alonzo. 1956. *Introduction to Mathematical Logic*. Princeton: Princeton University Press.

²⁹ My teachers at the City College of New York, Michael Levin, Arthur Collins and Charles Evans, were not sympathetic to Cartesian dualism. That changed in graduate school at Brown University, where I studied with Roderick Chisholm, Ernest Sosa and James Van Cleve. A youtube lecture on the *Cogito* by Sosa led to this paper. Fred Adams, Roger Ariew, Geoff Giorgi, Gary Rosenkrantz, and Ernest Sosa supplied helpful comments early on. My former colleague at the University of Rhode Island, John Peterson, was the most helpful of all, providing valuable criticisms on the paper's many drafts that led to significant improvements. Thanks to all. I'd also like to thank my former colleague at Northern Illinois University, Don Cress, for providing such clear and readable translations of *Discourse* and *Meditations*.

- Clarke, Desmond. 1992. "Descartes' Philosophy of Science and the Scientific Revolution." In *The Cambridge Companion to Descartes*, edited by John Cottingham, 258-285. Cambridge: Cambridge University Press.
- Copi, Irving. 1967. *Symbolic Logic*. Third Edition. New York: Macmillan.
- Cottingham, John (ed). 1992. Cambridge: Cambridge University Press.
- Curley, E.M. *Descartes Against the Skeptics*. 1978. Boston: Harvard University Press.
- Curry, Haskell B. 1977. *Foundations of Mathematical Logic*. New York: Dover Publications.
- Cusmariu, Arnold. 1977. *A Platonist Theory of Properties*. Unpublished Ph.D. dissertation written at Brown University under the direction of Ernest Sosa, Roderick M. Chisholm and James Van Cleve.
- Cusmariu, Arnold. 1978A. "Nonexistence without Nonexistents." *Philosophical Studies* 33: 409-412.
- Cusmariu, Arnold. 1978B. "About Property Identity." *Auslegung* 5 (3): 139-146.
- Cusmariu, Arnold. 1978C. "Self-Relations." *Southern Journal of Philosophy* 16 (4): 321-327.
- Cusmariu, Arnold. 1979A. "On an Aristotelian Theory of Universals." *Australasian Journal of Philosophy* 57 (1): 51-58.
- Cusmariu, Arnold. 1979B. "Russell's Paradox Re-examined." *Erkenntnis* 14: 365-370.
- Cusmariu, Arnold. 1980. "Ryle's Paradox and the Concept of Exemplification." *Grazer Philosophische Studien* 10: 65-71.
- Cusmariu, Arnold. 1982. "Translation and Belief." *Analysis* 42 (1): 12-16.
- Cusmariu, Arnold. 1983. "Translation and Belief Again." *Analysis* 43 (1): 23-25.
- Cusmariu, Arnold. 1985. Self-Predication and the "Third Man." *Grazer Philosophische Studien* 23: 105-118.
- Cusmariu, Arnold. 2009. "The Structure of an Aesthetic Revolution." *Journal of Visual Arts Practice* 8 (3): 163-179.
- Cusmariu, Arnold. 2012. Toward a Semantic Approach in Epistemology. *Logos & Episteme. An International Journal of Epistemology* III (4): 531-543.
- Cusmariu, Arnold. 2015A. "Baudelaire's Critique of Sculpture." *Journal of Aesthetic Education* 49 (3): 96-124.
- Cusmariu, Arnold. 2015B. "The Perils of Aphrodite: A New Take on Star Theory." *Film International* 13.3: 97-116.
- Cusmariu, Arnold. 2016A. "A Methodology for Teaching Logic-Based Skills to Mathematics Students." *Symposion. Theoretical and Applied Inquiries in Philosophy and Social Sciences* 3 (3): 259-292.
- Cusmariu, Arnold. 2016B. "Semantic Epistemology Redux: Proof and Validity in Quantum Mechanics." *Logos & Episteme. An International Journal of Epistemology* VII (3): 287-304.

- Cusmariu, Arnold. 2016C. "Toward an Epistemology of Art." *Symposion. Theoretical and Applied Inquiries in Philosophy and Social Sciences* 3 (1): 37-64.
- Cusmariu, Arnold. 2017A. "The Prometheus Challenge." *Symposion. Theoretical and Applied Inquiries in Philosophy and Social Sciences* 4 (1): 17-47.
- Cusmariu, Arnold. 2017B. "The Prometheus Challenge Redux." *Symposion. Theoretical and Applied Inquiries in Philosophy and Social Sciences* 4 (2): 175-209.
- Descartes, René. 1897–1913. *Œuvres de Descartes*. 13 volumes. Edited by Charles Adam and Paul Tannery. Paris: Léopold Cerf.
- Descartes, René. *Œuvres Complètes*. Kindle Edition. Text based on the Levrault edition of Descartes' works published in 1874 by Victor Cousin. References are to this edition because it is available online in electronic form and is searchable.
- Descartes, René. *Discourse on Method*. 1970 [1911]. In *The Philosophical Works of Descartes*. Translated by E. S. Haldane, vol. I: 81-130. Cambridge: Cambridge University Press.
- Descartes, René. 1912. *Discourse on Method*. Translated by John Veitch. Publisher unknown.
- Descartes, René. *Discourse on Method*. 1985. In *The Philosophical Writings of Descartes*. Translated by John Cottingham, Robert Stoothoff & Dugald Murdoch, vol. I, 111-151. Cambridge: Cambridge University Press.
- Descartes, René. *Discourse on Method*. 1998. Translated by Donald A. Cress. Indianapolis: Hackett Publishing Company.
- Descartes, René. *Discourse on Method*. 1999. Translated by Desmond M. Clarke. New York: Penguin Books.
- Descartes, René. *Discourse on Method*. 2006. Translated by Ian Maclean. New York: Oxford University Press.
- Descartes, René. *Discourse on Method*. 2007. Translated by Richard Kennington. Newburyport MA: Focus Publishing.
- Descartes, René. *Discourse on Method*. 2017. Translated by Jonathan Bennett. Online text <http://www.earlymoderntexts.com/assets/pdfs/descartes1637.pdf>. Accessed 11-1-2020.
- Descartes, René. 1970 [1911]. *Meditations on First Philosophy*. In *The Philosophical Works of Descartes*. Translated by E. S. Haldane, vol. I, 133-199. Cambridge: Cambridge University Press.
- Descartes, René. 1979. *Meditations on First Philosophy*. Translated by Donald A. Cress. Indianapolis: Hackett Publishing Company.
- Descartes, René. *Meditations on First Philosophy*. 1984. In *The Philosophical Writings of Descartes*. Translated by John Cottingham, Robert Stoothoff & Dugald Murdoch, vol. II, 1-62. Cambridge: Cambridge University Press.
- Descartes, René. 1998. *Meditations and Other Metaphysical Writings*. Translated by Desmond M. Clarke. London: Penguin Books.

- Descartes, René. 2008. *Meditations on First Philosophy*. Translated by Michael Moriarty. Oxford: Oxford University Press.
- Descartes, René. *Meditations on First Philosophy*. 1990. Bilingual Edition, edited and translated by George Heffernan. South Bend IN: University of Notre Dame Press.
- Descartes, René. *Principles of Philosophy*. 1970 [1911]. In *The Philosophical Works of Descartes*. Translated by E. S. Haldane, vol. I, 202-302. Cambridge: Cambridge University Press.
- Descartes, René. *Principles of Philosophy*. 1912. Translated by John Veitch. Publisher unknown.
- Descartes, René. *Principles of Philosophy*. 1985. In *The Philosophical Writings of Descartes*. Translated by John Cottingham, Robert Stoothoff & Dugald Murdoch, vol. I, 177-291. Cambridge: Cambridge University Press.
- Doney, Willis (ed). 1967. *Descartes: A Collection of Critical Essays*. New York: Doubleday.
- Enderton, Herbert B. 2001. *A Mathematical Introduction to Logic*. Second Edition. New York: Academic Press.
- Ferreiros, Jose. 2001. "The Road to Modern Logic-An Interpretation." *The Bulletin of Symbolic Logic* 7 (4): 441-484.
- Frankfurt, Harry G. 1970. *Demons, Dreamers and Madmen*. Princeton: Princeton University Press.
- Frege, Gottlob. 1967 [1884]. *The Basic Laws of Arithmetic*. Translated and edited with an introduction by Montgomery Furth. Berkeley and Los Angeles: University of California Press.
- García-Carpintero, Manuel. 1998. "Indexicals as Token-Reflexives." *Mind* 107: 529-563.
- Gaukroger, Stephen. 1995. *Descartes: An Intellectual Biography*. Oxford: Oxford University Press.
- Georgi, Geoff. "Demonstratives and Indexicals." *Internet Encyclopedia of Philosophy*. Accessed 10-11-2020.
- Gödel, Kurt. 1944. "Russell's Mathematical Logic." In *The Philosophy of Bertrand Russell*, edited by P. A. Schilpp, 125-153. La Salle: Open Court.
- Hilbert, David and Wilhelm Ackermann. 1950. *Principles of Mathematical Logic*. Translated by L.M. Hammond, G.G. Leckie and F. Steinhardt, and edited by Robert E. Luce. New York: Chelsea. The German edition titled *Grundzügen der theoretischen Logik* was published by Springer in 1928.
- Hunter, Geoffrey. 1971. *Metalogic*. Berkeley: University of California Press.
- Hurley, Patrick J. 2008. *A Concise Introduction to Logic*. Tenth Edition. Belmont CA: Wadsworth.
- Kalish, Donald, Richard Montague, and Gary Mar. 1980. *Logic: Techniques of Formal Reasoning*. 2nd edition. New York: Harcourt Brace Jovanovich.
- Kamp, Hans. 1971. "Formal Properties of 'Now.'" *Theoria* 37: 237-273.

- Kaplan, David. 1989. "Demonstratives." In *Themes from Kaplan*, edited by J. Almog, J. Perry, and H. Wettstein, 481–563. New York: Oxford University Press.
- Katz, Jerrold J. 1986. *Cogitations*. New York: Oxford University Press.
- Kenny, Anthony. 1968. *Descartes: A Study of His Philosophy*. New York: Random House.
- King, Jeffrey C. 2001. *Complex Demonstratives*. Cambridge, MA: MIT Press.
- Kleene, Stephen Cole. 1950. *Introduction to Metamathematics*. New York: Van Nostrand.
- Kleene, Stephen Cole. 1967. *Mathematical Logic*. New York: John Wiley & Sons.
- Klemke, E.D. (ed.) 1968. *Essays on Frege*. Chicago: University of Illinois Press.
- LeBlanc, Hughes. 1966. *Techniques of Deductive Inference*. Englewood Cliffs NJ: Prentice Hall.
- Markie, Peter J. 1986. *Descartes' Gambit*. Ithaca, NY: Cornell University Press.
- Mendelson, Elliott. 1979. *Introduction to Mathematical Logic*. New York: Van Nostrand.
- Peacocke, Christopher. 1976. "What is a Logical Constant?" *Journal of Philosophy* Vol. LXXIII No. 9: 221-239.
- Perry, John. 2001. *Reference and Reflexivity*. Stanford, CA: CSLI Publications.
- Quine, Willard Van Orman. 1940. *Mathematical Logic*. Cambridge MA: Harvard University Press.
- Quine, Willard Van Orman. 1950. *Methods of Logic*. New York: Holt, Reinhardt & Winston.
- Quine, Willard Van Orman. 1951. "Two Dogmas of Empiricism." *The Philosophical Review* 60: 20-43.
- Ramsey, Frank P. 1965 [1931]. *The Foundations of Mathematics*. Totowa NJ: Littlefield, Adams & Company.
- Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: Macmillan.
- Rosenkrantz, Gary S. 1993. *Haecceity: An Ontological Essay*. Dordrecht: Kluwer.
- Russell, Bertrand. 1905. "On Denoting." *Mind*, New Series, Vol. 14, No. 56. (Oct. 1905): 479–493. Reprinted in *Logic and Knowledge*. 1956. Edited by Robert C. Marsh 41-56. New York: Macmillan. Page references are to this reprint.
- Russell, Bertrand and Whitehead, Alfred North. 1911. *Principia Mathematica*, Vol. 1. Cambridge: Cambridge University Press.
- Russell, Bertrand. 1975. *Autobiography*. London and New York: Routledge.
- Salmon, Nathan. 2005. *Reference and Essence*, 2nd Edition. Amherst, NY: Prometheus Books.
- Sarkar, Husain. 2003. *Descartes Cogito*. New York: Oxford University Press.
- Schoenfield, Joseph R. 1967. *Mathematical Logic*. Reading, MA: Addison-Wesley.
- Simco, Nancy and Gene James. 1983. *Elementary Logic*. Second edition. Belmont CA: Wadsworth.
- Smullyan, Raymond M. 2014. *A Beginner's Guide to Mathematical Logic*. New York: Dover Publications.

- Soames, Scott. 2005. *Reference and Description*. Princeton: Princeton University Press.
- Takeuti, Gaisi. 1975. *Proof Theory*. New York: Dover Publications.
- Tarski, Alfred. 1944. "The Semantic Conception of Truth." *Philosophy and Phenomenological Research*, 4 (3): 341–376.
- Tarski, Alfred. 1983 [1933]. "On the Concept of Logical Consequence: In *Logic, Semantics, Meta-Mathematics*, 409-420. Indianapolis: Hackett.
- Thomason, Richmond H. 1970. *Symbolic Logic: An Introduction*. London: Macmillan.
- Williams, Bernard. 1978. *Descartes: The Project of Pure Enquiry*. New York: Penguin Books.
- Wilson, Margaret D. *Descartes*. 1978. London: Routledge.
- Wittgenstein, Ludwig. 1963 [1921]. *Tractatus Logico-Philosophicus*. Translated by David F. Pears and Brian F. McGuinness. London: Routledge & Kegan Paul.
- Zalta, Edward N. 2018A. "Frege's Rule of Substitution." <https://plato.stanford.edu/entries/frege-theorem/rule-substitution.html/>. Accessed 17-7-2020.
- Zalta, Edward N. 2018B. "Frege's Theorem and Foundations for Arithmetic." <https://plato.stanford.edu/entries/frege-theorem/>. Accessed 17-7-2020.

LOOKING FOR BLACK SWANS: CRITICAL ELIMINATION AND HISTORY

Michael F. Duggan

Abstract: This article examines the basis for testing historical claims and proffers the observation that the historical method is akin to the scientific method in that it utilizes critical elimination rather than justification. Building on the critical rationalism of Karl Popper – and specifically the deductive component of the scientific method called *falsification* – I examine his tetradic schema and adapt it for the specific purpose of historical analysis by making explicit a discrete step of critical testing, even though the schema is adequate as Popper expresses it and the elimination of error occurs at all steps of analysis. I also add a discrete step of critical elimination to Popper’s schema even though the elimination of error occurs at every step of analysis. The basis for critical elimination history is the demonstrable counterexample. The study of history will never approach the precision of science – history deals with open systems that cannot be replicated like experiments guided by fundamental laws. But just because we cannot know something with the rigor of science does not mean that we cannot know it better than we do. There may be no objective truth in an absolute sense, but there is a distinction to be made between well-tested and poorly tested theories and therefore between history done well and history done with less analytical rigor. What I hope to show is how our historical knowledge may progress through good faith critical discussion – history is discussion – and the elimination of error.

Keywords: critical rationalism, Karl Popper, black swans.

It is easy to obtain confirmations, or verifications, for nearly every theory – if we look for confirmations. (Popper 1965, 36)

No number of sightings of white swans will ever prove the theory that all swans are white, but the sighting of just one black swan may disprove it. (Popper 1935, 27)¹

Introduction: Back to Popper

What is the basis for preferring one historical theory or claim over another?² What is the dominant method in the selection and analyses of historical statements and conjectures? Is it the testing and critical discussion of competing ideas generally associated with science, or is it the partisan selection and defense of evidence

¹ Another way of stating this is “no amount of observed instances can have the slightest bearing upon unobserved instances.” (Miller 1985, 107)

² Popper asks a broader version of this question in *The Logic of Scientific Discovery* (1935/59, 108).

supportive or sympathetic to one's asserted position more typical of adversarial activities like forensics and litigation? If some historical discussions are merely contests of *verification* or *justification* rather than attempts to get to truer answers, then what might be done to transform such discourse into actual critical discussions? The purpose of participating in historical discourse should never be to engage in polemics, or trying to win for the sake of winning, but rather the attempt to arrive at truer and better-tested explanations and interpretations (Popper 1994, 160). There may never be a completely objective understanding of history in an absolute sense, but it is possible to progress toward it and there is a distinction to be made between theories that are well-tested and those that are less well-tested and therefore between history that is well done and history that is poorly done.

The study and writing of history, although in part an empirical endeavor, typically involves subjects with a greater degree of causal openness, thus allowing for a wider range of interpretations than do the more purely rational-empirical activities of the physical sciences.³ It is also more difficult to limit – isolate – the parameters of historical events and therefore questions about them. As one of the humanities, the larger part of history – addressing the 'why' questions – is interpretive, valuative, rather than narrowly factual and quantitative like science, but its method, as with that of science, is critical-rational, as well as intuitive and interpretive.⁴ Even with its dual nature, history is a part of the greater enterprise of the pursuit, increase, refinement, and testing of knowledge – of discovery. As with all empirical endeavors, and because we may never *justify* a claim with positive instances, there is only one method: the elimination of error, testing. Some interpretations are better – more rational, more accurate, better corroborated, more complete, more nuanced and insightful, better tested, more *true* – than others. On this point, I found my claims on Tarski's sophisticated realist definition of truth as the quality of theory-laden correspondence.⁵

³ Regarding causal openness, see Karl Popper, *The Open Universe* (1982) and "Clocks and Clouds" in *Objective Knowledge* (1972, 206-255). On *sense-qualia* and *quantia*, see A.J. Ayers, *Philosophy in the Twentieth Century* (1982, 66-67, 84-85, 88-90). See also Ayer, *The Central Problems of Philosophy* (1973, 71-72, 90-94, 101-102, 104, 118). Our sense perceptions tend to be more accurate in regard to quantities and in terms of quality. We may disagree on the color of a person's eyes, but not the number.

⁴ Edward O. Wilson, among others, regards the humanities to be those fields that reveal truths about human nature, where science attempts to tell us what the truths are about humans and the physical world. See Edward O. Wilson, *The Social Conquest of Life* (2012, 268-284). Also, see generally Edward O. Wilson, *The Origins of Creativity* (2017). For Popper's discussion on history's affinity with science, see "A Pluralist Approach to the Philosophy of History" in *The Myth of the Framework* (1994, 130-153).

⁵ "A true statement is one which says that a state of affairs is so and so, and the state of affairs is indeed so and so." See Alfred Tarski, *Logic, Semantics, Metamathematics* (1956/1983, 155). As Popper observes, the correspondence definition of truth is preferable to the two other definitions: truth as coherence and truth as pragmatic utility (consequentialism). (1972, 308-

The most fruitful function of history is to proffer and criticize new interpretations and to refine, modify, call into question, or refute existing interpretations via critical discussion. History *is* discussion. It is how we cull order from chaos, the cacophony of the aggregate of human interaction and the inaccessible motivations that underlie it. History is not a physical science, and it is an open question about the degree to which the component of critical elimination – the disproving of claims if incorrect, thus corroborating them if we fail to disprove them – so central to the scientific method which Karl Popper calls *falsification*⁶, can be meaningfully applied as a component of the analysis of

309) See Chapter 9, “Philosophical Comments on Tarski’s Theory of Truth” 319-340 from *Objective Knowledge*.

⁶ See generally Karl Popper, *The Logic of Scientific Discovery* (1935/1959). What I am describing here as ‘academic history’ is critical history. As one might expect, the philosophy of history and questions pertaining to the historical method and the craft of the historian have attracted both historians and philosophers and have resulted in a wide range of works on various aspects of how to approach history. These include serious philosophical treatments like R.J. Collingwood’s *The Idea of History* and surveys like Peter Novick’s *That Noble Dream*. There are also theories of singular causes of national character, like the Frontier Thesis of Frederick Jackson Turner’s “The Significance of the Frontier in American History,” and sweeping ‘big picture’ interpretations that borrow equally from anthropology, like Jared Diamond’s *Guns, Germs, and Steel*. Among the more impressive refutations of a historical outlook is Popper’s *The Poverty of Historicism*, which effectively dispatched the philosophical basis of deterministic programs like the idealistic vitalism of Hegel, the materialistic vitalism of Marx, as well as the cyclic models of Oswald Spengler (*The Decline of the West*) and Arthur Toynbee (*A Study of History*). I agree with Popper and contend that historicist and eschatological programs – to include newer incarnations like the neoliberal ‘end of history’ thesis of Francis Fukuyama – are fundamentally mistaken. John Gray’s pugnacious *Black Mass* makes a strong case against such programs and the narratives on which they are based.

Works that embrace history as a basis for living include Nietzsche’s famous Meditation, “On the Advantage and Disadvantage of History for Life,” and Margaret MacMillan’s *Dangerous Games*. I believe that all knowledge is retrospective, and reflective – historical – and therefore, it is necessary to apply lessons of the past in order to navigate through life. Given the disastrous route taken by Germany in the first half of the twentieth-century in part due to the historiography of Heinrich von Treitschke, (and in the American tradition the egregious distortions of the William Dunning School), I believe that it is also key to get history as ‘right’ as possible. As with George Kennan – a diplomat and statesman-turned-historian who successfully applied history to grand strategy – I believe that a broad and intimate understanding of history (as opposed to narrow, formal academic understanding and theory) is the most fruitful grounding for policy analysis. Some examples of works by prominent historians elaborating on the historian’s craft or merely reflecting on their subject include Stephen E. Ambrose’s *Personal Reflections of an Historian*, Bernard Bailyn’s *Sometimes and Art*, John Lewis Gaddis’s *The Landscape of History*, Edward Hallett Carr’s *What is History?* Will and Ariel Durant’s *Lessons of History*, Eric Foner’s *Who Owns History?*, Barbara Tuchman’s *Practicing History*, and Gordon Wood’s *The Purpose of the Past*.

The purpose of history in my opinion is not to strike a balance among competing views, but to tell the truth. Looking at both sides, or multiple sides, and analyzing multiple points of view is a necessary part of being a historian. But eventually he or she must make a judgment, an opinion, reflecting the formulation of a mature interpretation. The job of historians is to tell the truth as

historical explanations. This is not to say that the study of historical topics is identical to the study of natural phenomena determined by fundamental laws and represented with formulas, as with physics. It is merely the Popperian idea that through the elimination of error, of incorrect, dogmatic, delusional, misleading/propagandistic, less accurate, or less complete conjectures, our historical understanding may progress. By better testing our ideas, we may thus interpret history more accurately. Unlike science, there is no formal prediction in the historical method other than the forward-looking aspect of critical elimination and the expectation of testing a theory or interpretation of the past.

More specifically I am positing the idea that falsification or something conceptually akin to it exists in the historical method as critical elimination by historical counterexample. Popper knew of this affinity and that the historical method can be illustrated by his tetradic schema – his general formula of how knowledge progresses through the criticism and the subsequent modification of historical theories, through the elimination of error (1994, 144-145). Although Popper's schema is sufficient as he expresses it, I am also modestly suggesting that it be honed, made more specific for the purpose of historical analysis. Here, the more specific activity of critical testing (or CT) should be considered a discrete element of the schema prior to critical discussion (or CD) or else, in another version of Popper's schema, replace the even more general category of the elimination of error (EE) altogether.⁷ To be fair, historians test concepts in a less formal sense at all steps of the historical method, whether they realize it or not.

My intention then is not to blaze completely new trails, but rather to show how the ideas of critical rationalism and especially the methodological element of critical elimination are actually used in the study and analysis of history. In doing so, I hope to reintroduce critical rationalism to discourse on the method of history.

History as record is an accretion – an organic body of accumulated tentative knowledge, complementary and diverging theories, speculation, and information, rather than a singular system of analysis. It embodies a series of never-ending dialogs whose conclusions are always subject to revisiting, challenges, revision, and replacement with new interpretations. Where interpretations cannot be tested, history shares at least one of the weaknesses of the law as jury trial: it is an

they see it and not to strike an unoffending balance when the truth of competing views is not equal. This kind of surrender to moral neutrality at all costs – a 'balanced' opinion in arenas where ideas and values clash – should be avoided. Historians make mistakes of judgment all the time, but it is better to be earnestly mistaken than to have never tried to tell the truth. I have long subscribed to the idea that human history has to be interpreted within the larger context of natural history. On this point, Edward O. Wilson observes, "[h]istory makes no sense without prehistory, and prehistory makes no sense without biology." See *The Social Conquest of Earth* (2012, 287). Such a reading is more than suggestive that, spite of our impressive aesthetic, scientific, and technological accomplishments, as an overpopulated plague species that is destroying the planet, the project of human civilization has been a colossal failure.

⁷ For Popper's version of the schema in which CD is replaced with EE, see *Objective Knowledge* (1972, 119).

ongoing series of debates in which the most persuasive arguments often win out. The problem is that 'most persuasive' or 'most compelling' is not synonymous with 'truest.'

In another sense however, as an enterprise of interpreting and criticizing ideas, history is like a less rigorous version of science, although its initial conditions, assumptions, and models cannot be expressed or stated with the same degree of precision, and its predictive ability is far weaker, if extant at all. And of course historical events cannot be repeated like a scientific experiment. As with science, historical debate provides a forum for criticism and a clashing of assumptions and theories and the discussion of their consequences within a rational frame.

I hope that no one reading this will mistake my perspective for one of positivism, or conversely, postmodernist relativism or cognitive nihilist skepticism – it is a skeptical position, but one of rational rather than absolute skepticism, which is self-defeating, even in its own terms and one that Popper calls irrationalism.⁸ I am offering a critical rationalist perspective and analysis in an attempt to apply Popper's ideas to the historical method. Beyond this introduction, I will not discuss broader ontological and epistemological questions on the fundamental nature and accessibility of the physical world and events in it and its amenability to empirical description, rational analysis, and linguistic expression (both realism and anti-realism are "neither demonstrable nor refutable" and therefore questions about them are closed). (Popper 1972, 38-39) Discussions on

⁸ Critical rationalism is a form of rational skeptical philosophy – an 'attitude' more than a specific school – framed as such by Karl R. Popper. See *The Myth of the Framework* (1994, 190-191). It is based on, among other things, the idea that we learn by correcting our mistaken beliefs and by testing our ideas, rather than by shoring them up with supporting evidence. Popper believes that the critical tradition – the idea of improving an existing idea through criticism goes back to the PreSocratic philosopher, Anaximander, who criticized a cosmological theory of his mentor, Thales. See Popper, "Back to the Presocratics" in *Conjectures and Refutations* (1965, 136-165). The primary focus of Popperian critical rationalism has been in the philosophy of science and the scientific method. See generally *The Logic of Scientific Discovery* (1935/1959). This form of sophisticated realism has been embraced by major figures of science to include Albert Einstein and at one time Stephen Hawking. See Stephen Hawking, *A Brief History of Time* (1988, 10), and *Black Holes and Baby Universes* (1993, 94). See also Michael White and John Gribbin, *Stephen Hawking: A Life in Science* (1992, 102-103). By his own account, Hawking became discontented with critical rationalism and for a time cautiously embraced a view resembling instrumentalism. See Hawking, "My Position" in *Black Holes and Baby Universes* (1993, 44). More recently, Hawking had adopted a position called model-dependent realism, which appears to embody qualities of critical rationalism and pragmatism. See Stephen Hawking and Leonard Mlodinow, *The Grand Design* (2010 45-51). In addition to the philosophy of science, Popper also contributed significantly to political theory with *The Open Society and its Enemies* (1994/1945), the philosophy of history with, *The Poverty of Historicism* (1957), and the mind-body problem with, *The Self and its Brain* (1977). Popper discusses irrationalism at numerous places in *Conjectures and Refutations*, and *Objective Knowledge*. For example see Popper, *The Myth of the Framework* (1994, 180). On irrationalism in political thought, see *The Open Society and Its Enemies* (1945/1994, 430-461).

noumena and *phenomena*, are either historical reenactments or else the fruitless parlor games of academic careerists.⁹

There is no objectively justified history, only history rigorously done or not in varying degrees, and I would argue that the 'objectivity question' is in large measure misplaced and distractive, a red herring. Objective truth exists as an epistemological possibility and the most we can hope for it to get progressively closer to it. Therefore the pertinent question should never be the authoritarian dun of "how do you know?" but rather "how did you test your theory?"¹⁰ The objective basis for truth exists as physical reality as does a rational-empirical-linguistic frame by which to describe it. The ideational basis for describing objective facts exists, even if we are unaware of it.¹¹ In history as with science, the task is to discover or to move closer to truer theories and interpretations.

The methodological element in history is critical testing. One manifestation of this is critical discussion, and discussion proceeds from the proffering of interpretations. These we evaluate and test as rigorously as we can. We should resist the urge to *prove* a point or interpretation or try to *justify* a theory by its coherence with the existing state of knowledge (the coherence of a new theory with existing theories as a heuristic element is itself contingent upon testing by comparison and is the basis for the correspondence model of truth, but coherence in itself is not a basis for truth, something we might call the inductive fallacy of coherence).¹² Rather we should attempt to *corroborate* claims and interpretations

⁹ As Hume observes, absolute skepticism is an impossible position in that one would have to be skeptical of one's own skepticism, leading into an infinite regress. (1739/1888, 180-187) See also Hume's 1745 "A Letter from a Gentleman to His Friend in Edinburgh" in *An Enquiry Concerning Human Understanding*. (1993, 115-124) Regarding Popper's solution to "the not very deep methodological problem – the problem of historical relativism," see *The Myth of the Framework* (1994, 142-143).

¹⁰ As Popper notes, "[t]hus the empiricist's questions 'how do you know? What is the source of your assertion?' are wrongly put. They are not formulated in an inexact or slovenly manner, but *they are entirely misconceived*: they are questions that beg for an authoritarian answer... And I propose to replace, therefore, the question of the sources of our knowledge by an entirely different question: '*How can we hope to detect and eliminate error?*'" (1965, 25). See generally 21-27. Similarly, Popper believes that the idea of proving or verifying theories is linked with authoritarianism in science. See Popper, *The Myth of the Framework*. (1994, 94) On a related point, Popper addresses the idea of pseudo-questions such as "What is truth?" and more generally "What is?" and "What are?" – all "verbal or definitional questions" in *Objective Knowledge* (1972, 309) Popper regards these kinds of questions as unfruitful. Regarding pseudo-questions, see also *Conjectures and Refutations* (1965, 72-73).

¹¹ For Popper's 'three worlds' ontological model, see generally *The Self and Its Brain* (1977/1986). He also discusses aspects of the Three Worlds throughout *Objective Knowledge* (1972).

¹² For Popper's discussion on theories of truth, to include coherence, see *Objective Knowledge* (1972, 308-309).

via critical elimination or a kind of *progressive* or *positive negativism* – the potential progress of knowledge by eliminating error.¹³

We learn more from our mistakes than from our successes; knowledge progresses by correcting our mistaken beliefs in light of more powerful and more rigorously tested explanations, by trial and error. When a conjecture is presented with greater or truer explanatory power, we must as honest and rational¹⁴ investigators discard the discredited theory and embrace the new one rather than attempt to shore up the old one. My purpose then is primarily methodological. I am concerned with the question of how to test or better test theories; I am not interested in interpretive models. My assumptions are that the world exists as do other minds and that we may access, discuss, and come to know (or better know) aspects of it and ideas about it, however imperfectly, and how we may progress toward truer answers.¹⁵

In the spirit of critical rationalism, I would offer that history, when done well, is a process of testing through critical analysis and discourse, both internally by the individual historian and within the community of peers, reviewers, scholars, and students.¹⁶ The fundamental method of history, like all other enterprises of discovery, is the testing of ideas and the elimination of error.

In the first part of this article, I will examine the fundamental methodological categories of induction and deduction. In the second part, I will show how deduction as critical elimination can be – and in fact are – used as the decisive component of the historical method as critical discussion. I will then

¹³ I coined the term ‘positive negativism’ to characterize the progress of knowledge through critical elimination in an email correspondence with David Miller in 2012. See David Miller, “Missing the Target, The Unhappy Story of the Criticisms of Falsificationism” (2017, 13). The idea of negative empiricism was not invented by Popper, although he developed the idea more than anyone before or since. Charles Sanders Peirce and Victor Brochard both expressed this idea before him. See Nassim Nicholas Taleb, *The Black Swan* (2007, 56-57).

¹⁴ On the growth of knowledge through the correcting of mistakes, see Popper, *The Myth of the Framework* (1994, 93). On the discarding of disproved ideas, see Popper, *Conjectures and Refutations* (1965, 33-65). In *The Myth of the Framework*, Popper gives a good definition of what might be called soft rationality (as opposed to the hard reason of logic and math): “Rationality as a personal attitude is the attitude of readiness to correct one’s beliefs. In its most highly developed form, it is the readiness to discuss one’s beliefs critically, and to correct them in light of critical discussions with other people.” (1994, 181)

¹⁵ Plato, Hume, and Kant all believe that the fundamental nature of the world is beyond our ken. Popper, often regarded to be an epistemological optimist, agrees with this. See Popper, *Conjectures and Refutations* (1965, 194-195). But just because we cannot know something with absolute certainty does not mean that we cannot know it at all. See letter of Oliver Wendell Holmes, Jr. to Harold Laski dated January 11, 1929 (de Wolfe Howe 1952, 1124-1125). As regards the linguistic description of the world, if we accept a scientific view of language, such as the generative grammar model of Noam Chomsky, we can see that, contrary to postmodernist dogma, human beings are able to communicate ideas, however imperfectly. See generally Chomsky, *Language and Mind* (2006), and *Syntactic Structures* (1957).

¹⁶ Examples of internal testing would include Einstein’s famous thought experiments.

suggest an additional step of critical testing as a dynamic intermediate element between the initial theory or interpretation and critical discussion of it. It should be noted that both critical elimination (attempts to disprove an idea) and subsequent critical discussion such as peer review, when done well, are both means of testing rather than confirming conjectures as embodied by historical interpretations. I will then provide and discuss examples of testing simple historical facts, the testing of historical understanding of concepts, and macro theories. I will also discuss what might be called a 'Heisenberg Principle' of historical understanding. But first I feel compelled to discuss the current state of the philosophy of the historical methods.

The State of the Debate

The reader will quickly discern that there is scant reference to the ideas of current critical theory and philosophers in this paper. This is by design. Although I am acquainted with the current state of the philosophy of history, the various debates, and the leading players, I must confess to finding precious little of interest among them as regards the historical method. The discussions today seem to be little more than a rehashing of old debates about skepticism and relativism, terms that modern scholars often confuse or conflate.¹⁷ Some of these debates date from antiquity (debates on the centrality of language go at least as far back as Sextus Empiricus, while relativism and skepticism date back even farther to Protagoras – a Pre-Socratic philosopher – and Pyrrho).

Many recent discussions on historical methodology center around critical debates between the various incarnation of constructionism and realism.¹⁸ I see these as a continuation of the dustups between the antirealism, relativism, and irrationalist skepticism of postmodernism on the one hand, and a traditional realist epistemology of inductivism on the other.

When we read contemporary realist philosophy of history, we see valiant, if often tortured defenses of adequacy, completeness, superiority, and clarity of explanations, linguistic and cultural neutrality, and the extent or limitations of description, and coherence of interpretations with the existing body of knowledge. All of this amounts to an inductivist position – justification – and often with a tone of defensiveness, conditionality, temporizing, or apology. These were the concerns of the positivists and pragmatists of the nineteenth-century. As Popper observes, clarity is a 'moral duty,' but it has no bearing on the truth ('obscurantism' is actually intellectual obstructionism).¹⁹ Realist philosophers of history also

¹⁷ For the misuse of these words, see David Miller (2000, 156-173). Another word one hears a lot from irrationalists is subjectivism, which might be defined as relativism brought to the level of the individual.

¹⁸ For a discussion of constructivism, see Tom Rockmore, "Interpretation as Historical, Constructivism, and History" (2000, 184-199).

¹⁹ For an example of the concern for clarity among the Pragmatists, see Charles Sanders Peirce, "How to Make Our Ideas Clear" in *Philosophical Writings of Peirce* (1955/1878/1940, 23-41).

speak in terms of 'empirical justification' rather than rational-empirical corroboration and testing. I appreciate the efforts of members of the realist camp – thinkers like C. Behan McCullagh and Avirzer Tucker.²⁰ But in spite of the nobility of their efforts, some realists succumb to the problem of induction expressed by Hume 280 years ago.

As for the critics of the constructivist position who deny the possibility of truth as correspondence in history – that one's reality, to include history, is a construction of our perceptions rather than a reflection of an objective external reality – I find this claim to be immoral in a time when the liberal free press has been charged with accusations of 'fake news,' when extremist propaganda is accepted as true, and when even science is being called into question (it is also unclear how we may access ideational constructs that are external to the historian but not ideas about the external world).²¹

Curiously, both positions are wrong. Skepticism – the view that we cannot justify knowledge – is a true position, and therefore modern justificationist realists are wrong in their inductive methodology, but not in their realism. Constructionism, as a manifestation of the relativist/subjectivist position in its denial of the possibility of truth as correspondence (and therefore of truer/less true theories) is both incorrect, and in a time when the enemies of the open society are on the ascent, irresponsible. The subjective state of the investigator's mind or his or her intellectual background or outlook has no bearing on the correspondence of a theory to the external events it purports to describe. To think otherwise is to succumb to the fallacy of psychologism.²²

As regards justificationist realism being a mistaken outlook, one may ask how is it possible that objective reality exists and yet we may not support theories about it with 'evidence?'²³ Popper's answer to the problem of induction is straightforward: even though we cannot justify a theory, we may test it.

History, as with all of life, is about questions and problem solving. Questions of method imply assumptions about epistemology and knowledge (critical

On Popper's view of clarity as 'a moral duty' and 'an intellectual value,' see *Objective Knowledge* (1972, 44, 58). Regarding Popper's disdain for intentionally unclear writing, see his comments on Theodor W. Adorno in "Addendum 1974: The Frankfurt School." in *The Myth of the Framework* (1994, 78-81) and his paragraph on Hegel's writing in *The Open Society and its Enemies* (1945/1994, 243).

²⁰ See generally C. Behan McCullagh, *Justifying Historical Descriptions* (1984), and Avirzer Tucker, *Our Knowledge of the Past* (2004).

²¹ Popper expresses the dangers "of an epistemology that teaches that there are no objective facts" in his article "Source of Knowledge and Ignorance" in *Conjectures and Refutations* (1965, 5).

²² See Popper, *The Logic of Scientific Discovery*, section 25 (1935/1959). See also Popper, *The Myth of the Framework* (1994, 168-169).

²³ How may we be realists and not believe in positive reasons for our beliefs? See generally "Conjectural Knowledge: My Solution to the Problem of Induction" in *Objective Knowledge* (1972, 1-31).

rationalism assumes an outlook of sophisticated realism). Therefore questions about method are questions of philosophy and not of criticism.

Questions of relativism and subjectivity in interpretations are matters of practical concern. They should be addressed by the historian as a matter of course through a wide range of equally practical/commonsensical heuristic measures to minimize subjective bias, evaluate ideas, and eliminate error. This is a part of the commonsense 'art' of history (Popper 1994, 139). But as regards attempts to 'prove' or shore up positions with 'evidence,' these represent a backslide into inductivism. From perusing the literature, one infers that few philosophers of history today even realize that Karl Popper solved this problem more than 90 years ago (1972, 1).

I am therefore not interested in trying to justify beliefs; I am interested in testing ideas. I am not interested in arriving at more 'balanced' interpretations. Although historians may present multiple perspectives, they must ultimately tell the truth. In matters where values and truth are at stake, *balance* is for cowards, cynics, and psychopaths. I am interested in arriving at truer answers. Like Popper (by way of E.M. Forster), I "do not believe in belief." (1972, 25) Let others engage in debates gamed-out long ago to inconclusive ends about the meaning of meaning and the criteria necessary for justifying belief. As an issue of method, I could not care less about such closed questions.

I have also found that many philosophers today are shockingly unaware of modern scientific language theory – the generative language program of Noam Chomsky – preferring instead to see language in skeptical terms set within an outdated conventional understanding of language. Even a cursory understanding of Chomsky's psycholinguism makes one realize that the 'radical' linguistic beliefs and paradoxes of critical theory are little more than anachronistic distractions.²⁴

As for historians and philosophers of history offering epistemological explanations and analogs between history and science – scholars like McCullaugh and Tucker – in spite of their laudable efforts, I regard these to be foundationalist and justificationist (i.e. inductivist) in nature. As Popper observes, there are no, and can be no, ultimate foundations of knowledge, at attempts to justify beliefs are a part of "the mistaken quest for certainty." (1972, 42, 74-78)²⁵

My paper is on method and is shorn of interpretive matters. If this seems like a throwback, it is also intentional. Sometimes we must take a step backward in order to go forward. I have noticed in recent decades a lack of analytical rigor and ideological dispassion in some of the writing coming out of academic history departments and hope to reestablish a few basics given the understanding of these ideas in our own time. Those who think that discussions on falsification are of a purely historical nature – that they betray a concern for ideas whose currency is limited to the thinking of the early-to-mid-twentieth-century – are mistaken. The

²⁴ See Note 15.

²⁵ Popper attributes the term 'quest for certainty' to John Dewey (1972, 63).

first question we must ask about a philosophical concept is whether or not it is still productive. Critical elimination is one of the most productive concepts in the history of ideas and its importance is manifest in the boggling advances in the science of our time. To write-off falsification as the concern of another time would be like dismissing General Relativity or Hugh Everett's multiverse interpretation of quantum mechanics as intellectual antiques of the last century. The testing and elimination or corroboration of ideas as an element of analysis is timeless, although, like the ideas of the Presocratics or those of twentieth-century particle physicists, we may continue to build upon them. If the implications of falsification are still debated in theoretical physics, then it is certainly topical enough for the philosophy of history. If it was good enough for Einstein and Hawking, then it should be good enough for us. Popper notes that the ideas and inquiries of the early Greek thinkers are still with us today and find numerous modern analogs (e.g. models of the block, discrete, and ideational universe, the problem of motion, atomic theories, and evolution, to name a few). Because of this, Popper's exhortation 'back to the Presocratics' – also the title of one of his essays on the early Greek philosophers – is as pertinent today as ever before.²⁶ And so in a similar vein, I say 'back to Popper,' and make my case below.

I. Induction and Deduction: What We Think We are Doing Versus What We are Actually Doing

A. Induction

When we observe the world around us, we seem to be taking it in directly, literally. In fact just by thinking and observing, we are testing our assumptions and expectations. It is impossible to approach something without assumptions, and as Popper observes, just by considering a thing means that we already have "ideas and opinions about it."²⁷ In large measure, we even learn how to see.

The concept of induction for our purposes is perhaps best expressed by the cognitive model of Locke, stating that the mind is a blank slate, a *tabula rasa* and that all of our knowledge comes to us as unmediated information via the senses.²⁸ In Locke's model of simple or naïve realism, patterns and facts of the external world, expressed as law-like repetitions typical of behavior governed by laws of

²⁶ See Popper, "Back to the Presocratics" in *Conjectures and Refutations* (1965, 136-165). See also Popper, *The World of Parmenides* (1998).

²⁷ Popper's solution to the problem of induction shows that there is no such thing as unmediated perception and therefore that all observation involves the testing of assumptions – deduction. See Popper, *The Logic of Scientific Discovery* (1935/1959, 40) and *The Myth of the Framework* (1994, 145).

²⁸ See generally John Locke, "Essay Concerning Human Understanding" 1690. The idea that all observation is theory-laden is a central concept in Popper's philosophy. See *The Myth of the Framework* (1994, 145).

physics, impress themselves directly on our consciousness as pure experience and without theory, or interpretive or cognitive frames.

The biggest problem with the inductive method and one of the greatest sources of confusion surrounding it is the fact that it does not exist even though for most of us it seems to. (Popper 1935/1959, 40) The fact that people intuitively believe in induction is known as ‘the Psychological Problem of Induction’ or H_{PS} in Popper’s formulaic shorthand; the fact that induction does not really exist as a method is known as ‘the Logical Problem of Induction,’ or H_L .²⁹ Induction in logical terms means an assumption of deriving generalities from specific instances (Hurley 1988, 537).³⁰

Post-behaviorist cognitive theory tells us there is no such thing as unmediated perception.³¹ Our perceptions are routed through cognitive networks of the brain before we are even aware of them, which means, as my friend, David Isenbergh observes, we live consciously in continuous reaction in the immediate past (and therefore *all* knowledge is historical – we are blind to the future until it becomes the past), not a small consideration.³² It is through this routing that we become aware of and interpret our perceptions. Therefore all knowledge is interpretive – informed by existing knowledge and theories processed through psychological matrices – even when the interpretations are narrow, as with logic, applied mathematics, and simple sensory observations like seeing three pebbles in a jar.³³ As Popper writes in his essay, “On the Sources of Knowledge and Ignorance,” “Knowledge cannot start from nothing – from a tabula rasa – nor yet from observation. The advance of knowledge consists, mainly, in the modification

²⁹ See Popper, *Objective Knowledge* (1972, 1-31). Popper’s model shows that traditional empiricism and inductive model of perception – what he calls “the bucket theory of the mind” – in fact leads to an infinite regress. For Popper’s description of the bucket theory, see *Objective Knowledge* (1972, 60-63). How this model leads to an infinite regress, see *Conjectures and Refutations* (1965, 22-23).

³⁰ See also Popper’s discussion of the use of the word ‘induction’ by Aristotle and Bacon. (1965, 12-13)

³¹ As Popper notes, “[m]oreover, there is no such thing as an uninterpreted observation. All interpretations are interpreted in light of theories.” (1994, 145) See note 34.

³² The idea that human perception is eternally in the immediate past was suggested to me by David Isenbergh.

³³ If knowledge derived from perception was simply about pure and unmediated observation without psychological matrices to interpret and test such information, then animals with greater senses, such as dogs and cats, would presumably have a much greater understanding of the world. It would be they who dominate the world with science and abstract ideas, which is obviously not the case. Although they can commune with the natural world in ways which we cannot, we have far greater abstract knowledge of it. Regarding theories of generative grammar, see generally, Chomsky, *Language and Mind* (2006). For Popper’s theory of the mind, see generally Karl Popper and John C. Eccles, *The Self and Its Brain* (1977/1986). For a critique of generative grammar theory, see Edward O. Wilson, *The Social Conquest of Earth* (2012, 225-235).

of earlier knowledge.” (1965, 27-28) In other words, problems, theory and theoretical frameworks always precede perception.³⁴

The consequence of there being no direct perception is that there can be no unmediated interpretation of anything, including events, ideas, and texts (and people do not read texts, we read language). Therefore, deductive reason – whether it is the hard reason of formal systems of truth, the soft reason of simple open-mindedness (Popper 1994, 181) and good faith discussion, or the intermediate form of falsification in science and, in limited instances, in the study and practice of history – is the only true means of deriving knowledge. All methods, despite their distinctive disciplinary trappings and subject matter, involve the testing and criticism of ideas – and the question of academic disciplines is how to test ideas, given the character and dictates of the particular field and its constituent subjects.³⁵ They are all forms of testing and are therefore deductive.

What are the implications of the critical rationalist critique of induction on the practice of history? Unlike physicists, historians do not (and should not) look empirically for law-like patterns of the physical world in order to form their conclusions. As Popper observes in his essay, “On the Sources of Knowledge and Ignorance,” there are numerous non-scientific programs that attempt to justify their tenets ‘by *positive* reasons,’ and which are just as likely to be rationalist in nature as they are to be empirical.³⁶ Many of these programs are in the social sciences and include Freudian psychoanalysis, feminist critical theory, and Marxism. On this point, Popper quotes Bertrand Russell: “that no man’s authority can establish truth by decree; that we should submit to truth; that truth is above human authority” to include appeals to ideological premises.³⁷ (1965, 29-30)

³⁴ Popper writes, “[y]ou cannot start from observation: you have to know first what to observe. That is, you have to start from a problem. Moreover, there is no such thing as non-interpretive observation. All observations are interpreted in light of theories.” (1994, 145)

³⁵ On the idea that methodologies are determined by academic disciplines, Popper writes: “The belief that there is such a thing as physics, or biology or archaeology and that these ‘studies’ or ‘disciplines’ are distinguishable by the subject matter which they investigate, appears to me to be a residue from the time when one believed that a theory had to proceed from a definition of its own subject matter. But subject matter, or kinds of things, do not, I hold, constitute a basis for distinguishing disciplines. Disciplines are distinguished partly for historical reasons and reasons of administrative convenience (such as the organization of teaching and of appointments), and partly because the theories which we construct to solve our problems have a tendency to grow into unified systems. But all of this classification is a comparatively unimportant and superficial affair. We are not students of some subject but students of problems. And problems may cut right across the borders of any subject matter or disciplines.” (1965, 66-67)

³⁶ See “On the Sources of Knowledge and Ignorance” (Popper 1965, 3-30).

³⁷ Not only is truth above human authority, but Popper notes that “[t]here are no ultimate sources of knowledge” in general. (1965, 29) This tenet of Popper’s is conceptually related to Oliver Wendell Holmes’ observation that “[c]ertitude is not the test of certainty.” (1918, 40) As with advocates in the law, adherents to various ideological programs read observed evidence

The Problems of Induction then, are manifest in non-scientific programs as *justification*. Like scientists working under the assumptions of the Baconian inductive model, adherents to such programs look for repetitions of law-like patterns, such as 'laws of history' usually based on ideological assumptions. Unfortunately, human behavior and events arising from it preclude the predictive regularities of applied science. Unlike physics, there are no fundamental 'laws' of history.³⁸ Therefore when historians or social scientists who adopt a historicist approach look to what they perceive to be deterministic or rationalist regularities and patterns – often represented by misled attempts to apply probability and frequency ratios to human behavior – as analogs to physical laws, they err badly.³⁹ Historicists mistake human propensities for 'laws' of history or else as being reflective of the law-like predominance of reason in the human mind. They look for past examples of behavior upon which to base present conclusions as if the past truly is a historicist, or historically deterministic prologue. But the

and interpret such information as support of, or justifications of, their theories and when the evidence contradicts the theories; they simply modify them to accommodate such information. See Popper (1965, 33-65). Needless to say, we should only accept truth as the result of testing and criticism and not by the command of authority. After all, power is a characteristic of truth rather than a synonym; truth is powerful, especially when demonstrable, but not all power is truthful.

³⁸ See generally Popper, *The Poverty of Historicism*. On a broader note of historical determinism, it is probably trivial whether the world is one that includes unpredictable deterministic chaos or unpredictable indeterministic randomness.

³⁹ Various rationalist schools in the social sciences – 'rationalist' used here to mean an outlook assuming that reason is a dominant human trait or that there is a rationalist narrative to history, and as opposed to the 'weak' rationalist claim that it is simply better to be reasonable than unreasonable, but that reason is not a dominant human trait – make the curious assumption that people will generally act in their own perceived self-interest (examples would include Hegelian vitalism, Lockean libertarianism, Marxism, and the Law and Economics School). Human acts may or may not fall into general patterns of species-based behavior (to include an elasticity of behavior), but this tells us little about how individuals will act, which also varies on a cultural or social basis. What it does tell us does not have the predictive or explanatory power as the patterns of the physical world that can be tested in science. See Edward O. Wilson, *On Human Nature* (1978, 2-3), and *The Future of Life* (2002, xxi-xxii). The human brain is the product of evolution, but even if it is based on deterministic fundamental laws, human behavior is unpredictable as a chaotic phenomenon. The British philosopher and historian, John Gray, believes in behavioral determinism as did Stephen Hawking. Without entering into a discussion of free will, it would seem to be an open question about whether or not human beings can rise above their biology, their animal nature, via reason and moderation. See John Gray, *Straw Dogs* (2002, 3-17), and Stephen Hawking and Leonard Mlodinow, *The Grand Design* (2010, 30-34). Regarding fundamental laws in physics, see Max Tegmark, *Our Mathematical Universe* (2014, 134), and Roger Penrose, *The Road to Reality* (2004, 1020). On falsification itself, Penrose believes that Popper's idea that "the scientifically admissibility of a proposed theory, namely that it be observationally refutable... is too stringent a criterion, and definitely too idealistic a view of science in the modern world of 'big science.'"

past is not necessarily prologue. At best, history is an incomplete and partially obscured roadmap to a blind intersection.⁴⁰

As with inductivists in science, historicist historians look for confirmations and dismiss or explain away contradictions or inconsistencies, thus *insulating* or *immunizing* their position from criticism rather than inviting it in instances when it would risk calling their outlook into question. In doing so they embrace an attitude that is the exact opposite of critical rationalism.

In science probabilistic outliers can often be factored out. But in human events, outliers – Napoleon, Karl Marx, Thomas Edison, John Wilkes Booth, Albert Einstein, Adolph Hitler, Franklin Roosevelt – are often more influential on the course of history than individual people in the great aggregated mean of behavior. In a historicist scheme, evidence is seen as ‘proof’ of what one already believes, or else it is rejected or minimized. Needless to say, this approach to analysis has huge ramifications on the study and analysis of history.

The non-scientific nature of history is not in itself a fatal flaw. As Popper notes in *Conjectures and Refutations* and in other discussions on the Problem of Demarcation, just because a program is not scientific does not mean it cannot be important as an activity, theory, or body of ideas.⁴¹ As regards our topic, it simply means that the practice of evaluating historical interpretations is not a scientific means of discovery, although it is related to it.⁴² Popper believes that non-scientific programs, such as Freudian psychoanalysis or Darwinian evolution prior to later discoveries in genetics, are oftentimes of great importance and are perhaps even true in their claims, but they are not scientific.⁴³

⁴⁰ Regarding Popper’s views on historical determinism generally, see Karl Popper, *The Poverty of Historicism* (1957) (refuting the idea that there are determinist laws, cycles, or a narrative plot to history).

⁴¹ Popper gives Marxism and Adlerian psychoanalysis as examples of programs that utilize non-scientific justification (1965, 37).

⁴² See “The Problem of Demarcation,” in *Popper Selections* (Miller 1985, 118-30). Popper specifically notes that Freudian and Adlerian psychoanalysis may contain true ideas, even though they are not scientific programs. See also *Conjectures and Refutations* (1965, 37-38).

⁴³ Although Popper believes that disciplinary lines are largely artificial, there is a real demarcation between science and non-science. On the illusory nature of disciplinary distinctions, see *Conjectures and Refutations* (1965, 66-67). On Popper’s demarcation between science and non-science, see chapter 11 “The Demarcation between Science and Metaphysics,” in *Conjectures and Refutations* (1965, 253-303), and “The Problem of Demarcation” in *The Logic of Scientific Discovery* (1935/1959, 34-39). Popper also believes – correctly, I think – that some disciplines may straddle this demarcation between science and non-science and fall into both categories. Biology, for instance, if defined broadly enough to include both genetics and the study of animal behavior, would encompass activities that include both hard science and something closer to the social sciences. Popper likewise notes in *The Poverty of Historicism*, that evolution is an overarching meta-theory, organon, or ‘historical statement’ that characterizes the development of life on Earth and not a theory characterized by a singular ‘law of evolution’ to be tested. Even so, some aspects and claims of evolutionary theory as a body of theories as it now stands – laws of heredity or the existence of genetic mutations for example – can be tested

B. Deduction: Black Swans and Dead Roosters

How then do we progress toward more truthful interpretations?⁴⁴ Although there is no such thing as non-theory-laden observation, empiricism still plays a role. The critical rationalist solution is the idea that truthful knowledge of the physical world – especially scientific knowledge – comes from the proper amalgam of inspiration – including intuition and the rare counter-intuitive creativity that yielded ideas like Special and General Relativity and quantum mechanics – theory-laden empirical observation, and reason in the form of deductive testing.⁴⁵ From the critical rationalist model we can see that the historical distinction between the empiricism of the Anglo-American philosophical tradition and the rationalism of the Continental tradition is an artificial one. In practice, analysis of the external world requires both reason and observation. (Popper 1965, 54-55)⁴⁶

We learn not by justifying what we already believe, but by correcting our mistaken ideas and beliefs through critical analysis and discourse and in light of

and corroborated scientifically (Popper 1957, 106-107). In addition to the amenability to testing and prediction via falsification, there are other differences between science and non-science. Most prominent of these, is that history (for example) involves phenomena generally not subject to and guided by physical laws, but rather involve situations subject to the ultimate disordering factor in history: the interaction of human volition and therefore, caprice.

⁴⁴ As indicated in the introduction and in Note 5, ‘true’ and its variations (‘truer,’ ‘truest,’ ‘truthful’) are used here according to Tarski’s definition of truth as correspondence. For Popper’s views on the other two theories of truth – the coherence model, and the pragmatic model – see Popper, *Objective Knowledge* (1972, 308-309). Popper holds that correspondence is the only real theory of truth. Aristotle, and even the great skeptical empiricist, David Hume, also give definitions of truth as correspondence. See Aristotle, *Metaphysica* (1908, 7, 27); Hume, *A Treatise on Human Nature* (1739/1888, 3).

⁴⁵ Regarding Popper’s idea that theories – even scientific theories – are the ineffable products of the human imagination whose origins are irrelevant to their validity, see *The Logic of Scientific Discovery* (1935/1959, 31. A good example of the irrelevance of how a concept is inspired can be found in the example of the nineteenth-century German organic chemist, Friedrich August Kekule (1829-1896). Working with benzene cores – the central substance of many organic chemical compounds – Kekule mapped out the hexagonal model from an initial conjecture of organic molecular structure that supposedly came to him in a vivid daydream while on a London bus (presumably a horse-drawn ‘omnibus’) in 1858. The dream was of a folk dance that had a hexagonal configuration geometrically akin to the form he then hypothesized for the benzene molecule. Einstein’s imagining of riding a beam of light that helped him arrive at Special Relativity is a similar example of the creative origins of scientific theories. See *The Cambridge Encyclopedia of Scientists* (Millar et. al 1996, 180-181).

⁴⁶ Popper also quotes Russell’s famous defense of empiricism as “[i]t is therefore important to discover whether there is any answer to Hume [Problem of Induction] that is wholly or mainly empirical. If not, *there is no intellectual difference between sanity and insanity*. The lunatic who believes he is a poached egg is condemned solely on the ground that he is in a minority.” (1972, 5) This not only explains Hume’s despair at the end of his *Treatise* but Popper’s apparent pride at the beginning of his essay, *Conjectural Knowledge*, where he claims to have solved the problem of induction. See also Bertrand Russell, *History of Western Philosophy* (1946, 673).

more powerful explanations. (Popper 1994, 181)⁴⁷ Likewise, we do not ‘prove’ a point by finding a sufficient amount of evidence or ‘positive reasons’ to support or verify our position, but rather we may corroborate a theory by rigorously testing it, thus disproving and discarding it if shown to be untrue. Although some historical interpretations are stronger than others on their face, as Popper famously notes, no number of confirmations will ever prove a claim, while a single counter instance to the contrary may disprove it.⁴⁸ He writes:

There is no criterion for the truth, but there is something like a criterion of error: clashes arising within our knowledge or between our knowledge and the facts indicate that something is wrong. In this way, knowledge can grow through the critical elimination of error. This is how we can get nearer to the truth.⁴⁹ (Popper 1994, 143)

In the terms of logic, deduction is defined as a form of reasoning in which a specific conclusion must necessarily follow certain specific premises.⁵⁰ Deduction includes the syllogisms and enthymemes of logic, equations in mathematics, and in terms of physical testing, scientific falsification. A useful rule-of-thumb distinction between induction and deduction is that as a methodological process, the former is retrospective or backward-looking, while the latter is forward-looking to often unexpected results.⁵¹

To illustrate the difference between induction and deduction as tools of discovery let us consider two birds. The first is Popper’s famous black swan – the ornithological metaphor for elimination by counterexample. To illustrate induction, let us consider another well-known metaphor, that of a crowing rooster

⁴⁷ See also: “We cannot justify our theories, but we can rationally criticize them, and tentatively adopt those which seem best to withstand our criticism, and which have the greatest explanatory power.” (1972, 265)

⁴⁸ For Popper’s famous black swan, see *Popper Selections* (Miller 1985, 110). By contrast, the view that we can justify beliefs is a form of simple empiricism sometimes called positivism, and the view that we cannot justify our knowledge is called skepticism. See “Sokal & Bricmont: Back to the Frying Pan” (Miller 2000, 156-173). Critical rationalism can be characterized as a form of sophisticated realism and an outlook of rational skepticism. Admittedly, some realist, as opposed to phenomenalist, positivists embrace forward-looking experimentation and therefore – whether they knew it or not – falsification. The American philosopher Chauncey Wright may serve an example. See *Chauncey Wright and Forward-Looking Empiricism* (Duggan 2002). Wright’s student, Charles Sanders Peirce, actually articulated the concept of falsification, and Hillary Putnam suggests that Peirce anticipated falsification decades before Popper. See Hillary Putnam, *Pragmatism, an Open Question* (1996, 71) citing Charles Sanders Peirce, “Pragmatism and Pragmaticism,” (Hartshore, Weiss and Burks 1943, 443). Economist and historian Nassim Nicholas Taleb also notes that Peirce hit on the idea of negative rationality in empiricism, but believes that Victor Brochard happened on it even earlier (1879). (2007, 57)

⁴⁹ See also *Objective Knowledge* (1972, 318).

⁵⁰ A deductive argument is defined by Patrick J. Hurley as one “in which we expect the conclusion to follow necessarily from the premises.” (1999, 535)

⁵¹ For a definition of forward-looking empiricism and the forward-looking nature of deductive processes, see “Max H. Fisch: Rigorous Humanist” (Madden 1986, 375-396).

relative to the rising of the Sun. Suppose that by faithful observation we know that each morning a rooster crows and then minutes later the Sun rises in the east, something that is confirmed by 100% of observed instances on non-overcast days. We infer from these repeated observations (and without bringing in additional theories or knowledge of basic astronomy and physics) that the rooster's crowing causes the sun to rise. Here we can transform the rooster from a tool of induction into a tool of deduction, a black swan. If we are inductivists we would compile our findings that reinforce our erroneous conjecture; if we are critical rationalists, we would kill the rooster and then wait to see if the Sun rises the next morning. If it does, we would rightfully discard the conjecture of the causal rooster.

This might seem to be a frivolous illustration, but those historians who compile inductive evidence to support a premise are engaging in an approach that is conceptually identical to that of a person who believes that the crowing of a rooster causes the Sun to rise.⁵²

II. Testing Hypotheses: Popper's Tetradic Schema

In science, as with the interpretation of historical events, texts and ideas, we must choose between competing theories and critical argument, the weightiest of which are ones that can be tested or falsified (Popper 1935/1959 and 1994). A scientist begins by framing a premise, a conjecture, which is first and foremost a creative endeavor and a product of the human imagination. How and where a theory originates is insignificant.⁵³ The more narrowly-framed a theory is – the more it

⁵² For the illustration of the crowing rooster, see Rothman and Greenland (2005, 51-52). The satirical television cartoon program, *The Simpsons*, succinctly illustrated the 'specious reasoning' of inductive justification in the episode, *Much Apu About Nothing* (FOX television broadcast May 5, 1996). In that episode, the following dialog occurs between Homer Simpson and his daughter, Lisa: Homer: Not a bear in sight. The Bear Patrol must be working like a charm! Lisa: That's specious reasoning, dad. Homer: Why thank you, honey. Lisa: By your logic, I could claim that this rock keeps tigers away. Homer: How does it work? Lisa: It doesn't work; it's just a stupid rock! Homer: Uh-huh. Lisa: But I don't see any tigers around, do you? Homer: Hmm... Lisa, I want to buy your rock.

⁵³ On the view that theories are products of the imagination, Albert Einstein writes, "Physical concepts are free creations of the human mind, and are not, however it may seem, uniquely determined by the external world." (Barlett and Kaplan 2002, 683) Einstein's theories of Relativity may or may not turn out to be the last word on the physics of the macro levels of the physical universe, but, like Linus Pauling's single helix theory of genetics or Lamarck's theory of evolution, the theories are, at the very least, wonderful creations of the human mind, beautiful ideas, and important chapters in the history of ideas. Often the phenomena studied by science are as beautiful as the ideas themselves. The great American physicist of light, Albert Abraham Michelson, famous for the Michelson-Morey experiment that disproved the idea of the ether, thus setting the stage for Einstein and Special Relativity, was a scientist who appreciated the aesthetics of his subject. In the first of his collection of lectures, *Light Waves and Their Uses*, Michelson writes, "If a poet could at the same time be a physicist he might convey to others the pleasure, the satisfaction, almost the reverence, which the subject inspires. The aesthetic side of the subject is, I confess, by no means the least attractive to me." (Michelson 1903, 1-2) See

forbids – the potentially stronger the test of it may be.⁵⁴ We then attempt to test or falsify it by setting up an experiment, a true-or-false physicalization of deduction that will refute the premise if untrue and corroborate it if true (experiments are also products of human inspiration). We then submit our findings for rigorous critical discussion or peer review. The experiment can be explained and then replicated, even by people who disagree with the original conjecture. If the conjecture passes this muster, we can accept it as a conditional truth until it can be further refined or disproved or until a new theory with greater explanatory power is devised and tested. Scientific knowledge therefore, progresses by vigorously testing – corroborating a theory by attempting to disprove it or find flaws with it, and failing – which also underscores the difference between the self-critical attitude of science when done well and the selective advocacy and defense of a position typical of contests like litigation and debate. Of course the most solid knowledge of the external world is that which describes phenomena based on or guided by physical laws that can be externally framed and inter-subjectively tested. The result of this critical process is what Popper calls *objective knowledge* (1935/1959, 44; 1994, 70, 93).

To what degree can the method of science be applied to the testing of historical theses? As Popper observes in *The Myth of the Framework*, analysis and criticism may progress even in areas of investigation where the harder analytical reason of science is not possible through softer means of criticism and testing (1994, 137-153). Knowledge progresses by testing premises via rigorous critical discussion, a process he spells out in a simple formula he calls the *tetradic schema*.⁵⁵ On the progress of knowledge generally, Popper writes:

In both [science and non-science] we start from myths – from traditional prejudices, beset with error – and from these we proceed by criticism: by the critical elimination of errors. In both the role of evidence is, in the main, to correct our mistakes, our prejudices, our tentative theories – that is, to play a part in the critical discussion in the elimination of error. By correcting our mistakes, we raise new problems, we invent conjectures, that is, tentative theories, which we submit to critical discussion directed to the elimination of error. The whole process can be represented by a simplified schema which I may call the tetradic schema:

$P_1 \rightarrow TT \rightarrow CD \rightarrow P_2$

also, Norman McLean, “Billiards is a Good Game; Gamesmanship and America’s First Nobel Prize Scientist” (2008, 78-92).

⁵⁴ On the relationship of the narrowness of a theory relative to its testability, see Popper *Conjectures and Refutations* (1965, 36, 3). Narrowness in terms of limiting or framing a conjecture should not be confused with the narrowness of a sample under investigation.

⁵⁵ Popper, “A Pluralist Approach to the Philosophy of History” in *The Myth of the Framework* (1994, 140-142). Popper also discusses his schema in other contexts in “Epistemology Without a Knowing Subject” in *Objective Knowledge* (1972, 119), and in “Of Clouds and Clocks” in the same volume at 243-244.

This schema is to be understood as follows. Assume that we start with some problem P_1 – it may be either a practical, or a theoretical, or a historical problem. We then proceed to formulate a tentative solution to the problem: a conjectural or hypothetical solution – a *tentative theory*, TT . This is then submitted to *critical discussions*, CD in light of evidence, if available. As a result, new problems, P_2 arise. (1994, 140-142)

Where Popper believes science differs from non-science, in addition to the fact that physical phenomena are directly subject to fundamental laws,⁵⁶ is with the inclusion of falsification as an element of critical discussion (CD), even though discussion is a part of testing a scientific theory at all points in the process. The unvarying nature of physical laws also allows for precise prediction in physics, something that also sets it apart from history.

A. An Added Component: The “Black Swan” (Critical Testing)

Popper’s famous adage and the inspiration for the title of this paper well illustrates this method. To paraphrase: if we want to test the conjecture that all swans are white, we should not look for white swans – no number of white swan sightings will ever ‘prove’ this conjecture – but rather black ones. A single confirmed sighting of a black swan will disprove the hypothesis. A statement that has been corroborated (such as “not all swans are white” to build on Popper’s example), can be regarded as a conditional truth if it can pass muster of a replicable true-or-false experiment, assuming there is risk of being shown to be untrue if it is untrue. It must then survive critical discussion and peer review.

Popper’s illustration of the black swan (and the dead rooster) as the elimination of an untrue statement applies as much to history as it would to any kind of testing of ideas: in order to test a historical premise, we should not attempt to shore it up by finding sympathetic ‘white swan’ or ‘crowing rooster’ support and justifications, but by ‘black swans’ or counterexamples that will disprove it if untrue, thus corroborating it if true. As such, documentation is not ‘proof’ or evidence to be used in support of a position, but sources to be evaluated themselves and then a basis against which to test the hypothesis. This may result in the corroboration of the conjecture or else by its elimination. The possibility of elimination is what Popper calls the ‘risk’ of testing a theory.

⁵⁶ Popper believes that because science, physics for example, is based on fundamental laws, that there is a difference of kind between science and non-science, to include history. He notes: “[w]hen we speak of success in physics we have in mind the success of its predictions: and the success of its predictions can be said to be the same as the empirical corroboration of the laws of physics.” (Popper 1957, 35) He also writes, “[i]f we were to admit laws that are themselves subject to change, change could never be explained by laws.” (1957, 103) Even with the Problem of Demarcation, Popper believes that there are affinities between science and history. He notes, “[b]ut my thesis has been for many years: *all those historians and philosophers of history who insist on the gulf between history and the natural sciences have a radically mistaken idea of the natural sciences.*” (1994, 139)

Popper's schema is a simplified illustration that we might flesh out to provide even greater clarity about how knowledge progresses in history. We may do this by adding the element of critical testing (*CT*) to the schema between the tentative theory (*TT*) and critical discussion (*CD*) the formula is rendered as $P_1 \rightarrow TT \rightarrow CT \rightarrow CD \rightarrow P_2$, thus formally distinguishing the critical testing of the historical counterexample and subsequent critical discussion, which the tetradic schema implicitly combines.⁵⁷ One could counter that because testing is implicit in both the formulation of the tentative theory and again in the critical discussion/error elimination phase of the process, the additional critical test is therefore redundant or extraneous. To this I would say that the critical test between the tentative theory and critical discussion would formalize and make plain the key step of falsification as manifest by experimentation in the scientific method. This addition is intended just to make the illustration clearer in terms of how the process actually works. 'Testing' implies an actual operation of elimination rather than a general assertion that elimination should occur at this point. Like the historian, the scientist also tests ideas just by thinking about them both informally and as thought experiments. Moreover, critical discussions – although having elements of testing and elimination – are less singular and formal and may include discussion groups and peer review, rather than experimentation, per se. Moreover, it is conceivable that a critical discussion could be based on probability or inductive arguments. Therefore formal deductive testing can be regarded as a discrete step or element of the schema as applied to historical analysis.

How would this work in practice? If to simply consider an idea is to test our opinions and expectations about it, then the additional step would be a test of a more formal, demonstrable nature. It would be a counter-hypothetical or counterexample that would risk disproving the conjecture if shown to be untrue (subsequent counterexamples will likely arise and be a part of critical discussions). We must test interpretations by rigorously seeking the equivalent to Popper's black swans to test our tentative theories.

B. Refutation and Simple Factual Claims

Negative rationality in history can be easily illustrated in regard to simple factual claims addressing historical 'what' or 'how' questions as the 'on point' counterexample or counterclaim. For example, if such-and-such a historical personage is believed by some to have been gradually poisoned by arsenic, but a test of that person's hair shows no trace of the poison, then we can dismiss the

⁵⁷ See Popper's "Epistemology Without a Knowing Subject" in *Objective Knowledge* (1972, 119). It is a fair point that both critical discussion and experimentation both involve the testing of ideas, but it is important to distinguish the two as distinct, interrelated phases of the process of testing ideas. In another expression of the schema, Popper uses the even more general *EE* (elimination of error) in the place of *CD*. See *Objective Knowledge* (1972, 119-122, 243-244).

premise. Of course this example actually involves the use of scientific testing. Similarly, if a historian makes the claim that the image of a shadowy bearded figure in a Daguerreotype taken in Peterborough, New Hampshire on a known date in 1860 is that of Abraham Lincoln, but it is known from photographs of Lincoln from that time show him to be beardless, or if Lincoln was in Washington D.C. on the same day, we can discard the claim.⁵⁸

For a strongly empirical example, let us suppose that a skeptic makes the claim that the Apollo Moon landings were faked in a film studio. If that person is taken to an observatory with a telescope powerful enough to allow an observer to actually see the boot prints and the flag left by the astronauts at one of the landing sites, then the claim can be considered refuted.

This would not 'prove' that men walked on the Moon in an absolute sense, but it would strongly corroborate it. One could only deny such an observation with elaborate, tortured attempts to insulate the original claim from such powerful empirical refutation. If a person does not accept a claim intersubjectively demonstrated to be truer, we must ask him or her "what refutation of your position would you accept?" (and then be ready to answer the same question if posed to us).⁵⁹ If the skeptic replies that he or she will accept no position refuting their own, then this person can be considered to be ideologically deluded or wedded to the position, or else irrational. In practical terms, there is little difference between the two (Popper 1994, 180-181).

Just as many trivial objective statements can be corroborated or eliminated through documentation (the time and place of the Lincoln assassination, for example), a large class of more important historical statements can also be tested.

C. Testing Metatheories and Concepts

Although narrowly-tailored, fact-based conjectures are more testable than broad and complex explanations addressing 'why' questions, there is no reason why counterexamples cannot be used to evaluate the historical usage, development, and understanding of ideas and even macro and meta-conjectures that are more

⁵⁸ See Stefan Lorant, *Lincoln, a Picture Story of His Life* (1979, 87). Coincidentally when Lincoln was a practicing attorney, he used negative rationality to defend a neighbor of his, Duff Armstrong, against a charge of murder. A witness claimed that even though it was night, he could see the face of the accused because of the nearly full moon. Lincoln discredited the man's testimony by producing an almanac with a lunar table showing that the moon had already set by the time of the murder. Mr. Armstrong was acquitted. See David Herbert Donald, *Lincoln* (1985, 150-151). I am obviously not the first person to hit on the idea of the counterexample as a means of testing historical ideas. The idea of a counterexample – 'proof' or a claim that disproves an existing claim – should not be confused or conflated with the idea of the counterfactual – a kind of hypothetical causal thought experiment. Historians who have embraced the idea of counterfactuals as means of testing historical hypotheses include Marc Bloch and Niall Ferguson. See John Lewis Gaddis, *The Landscape of History* (2004, 100-102).

⁵⁹ "What conceivable facts would I accept as refutations or falsifications of my theory." (Popper 1976, 41-42)

interpretive than simple empirical facts. The range of theories that can be evaluated this way might be thought of as existing on a spectrum from most narrow to the most broad (again, the more a theory forbids, the potentially stronger the test of it), but for the sake of convenience, let us postulate the categories of simple, almost archaeological, facts (to include narrow conceptual counterexamples as well as empirical counter-examples), and broadly interpretive and speculative theories and explanations.

a. The Forest and the Trees. A “Heisenberg Principle” of History: Testing Macro Interpretations

Large events and historical currents involving innumerable facts, bodies of facts, and theories, necessarily involve a greater element of interpretation. This macro, or ‘big picture’ approach to large but distinctive events also involves the testing of premises, although it also implies an apparent paradox, a kind of Heisenberg principle of historical understanding. The paradox goes like this: the broader the event being described and therefore the broader the interpretive theory (the more it attempts to explain or bring together), the potentially more interesting and important, but the inherently less knowable the subject and the less powerful the explanations of it. By contrast, the smaller, more limited and more fact-oriented a conjecture, the more solid it may be, but also the less interesting. As Arthur Schlesinger observes (echoing Popper), the more a theory explains, the less it explains and that a theory that “explains everything, explains very little.” (1986, 141)⁶⁰

⁶⁰ Schlesinger notes that an overly broad thesis about the Open Door Policy actually cuts against our understanding of it. He writes “[t]he Open Door Thesis is evidently not falsifiable. Because it explains everything, it explains very little. It is not a testable historical hypothesis at all. It is theological dogma.” That said, a ‘big picture’ outlook is more important in order for history to be useful, say as a basis for foreign claim than a ‘down in the weeds’ view with no idea of the bigger picture. A very general philosophical claims like the statement “war is a function of human irrationality and denotes a failure of policy” may be true and may form a partial basis for a realistic sensibility upon which to build a historical outlook, but does not go very far to understand or explain the proximate causes of an event like the First War. Schlesinger on Popper: “theories that attempt to explain everything in fact explain nothing.” The Heisenberg principle of history underscores the fact that both detail and circumspect are necessary – details may be hard and factual, and a generalized view – a historical sense (or ‘historical consciousness’ in Gaddis’s words) – is what a historian builds over a lifetime of study. It should be noted that although the very nature of history precludes precise prediction, a broad and intimate (as opposed to a remote and formal) understanding of history is perhaps the best grounding one can have in terms of formulating effective policy. Although there are no guarantees, a policymaker with historical understanding to inform his or her intuitions has a better chance of adumbrating the chances of a policy’s success or failure based on what has worked in the past and what has not, and why. In terms of foreign policy, George F. Kennan was this kind of intuitive ‘Cassandra.’

The broader a historical trend, the more *open* it is as a system and therefore the less testable, unless simply to disprove it.⁶¹ The more open a system is, the less solid our knowledge of it. This is especially evident of simple explanations of large and complex historical currents. Here counterexamples are also highly interpretive. Hegel and Marx posit historicist programs based respectively on a vitalistic unfolding of human events based on reason and historical necessity. To test these we might find any number of counterexamples showing that humans are as much of a randomizing element as they are an ordering factor, perhaps even more so.

Open systems by their very nature are not only less predictable, they are also less knowable in retrospect, and counterexamples here are more interpretive than a simple factual elimination. For example, the claim that United States policies toward the Soviet Union were justified by a relentlessly expansionist policy on the part of Stalin during the early Cold War – ‘World Communism’ – could be countered by the following falsifying examples: 1). Stalin’s ideological and personal opposition to Trotsky’s vision for world revolution (“Socialism in One Country”); 2). Stalin’s unwillingness to provide much help for communist movements in either Greece or China, and his lukewarm support for North Korea’s plans to invade the South in 1950. These counter-observations carry a certain weight and provide a compelling basis for criticizing the initial claim, but none are definitive in refuting it.

A more testable example might involve a simple or generalized interpretive statement applied to a large category. For instance, some historicists claim that civilizations follow a life cycle like that of an organism with successive phases of birth, infancy, immaturity, adolescence, maturity, prime, decline, and death.⁶² Let us also assume that this claim is not suggesting a common trend or tendency,⁶³ but rather a process guided by ‘laws’ of history and the subsequent claim that all societies follow this pattern. Let us also suppose that the ‘death’ of a civilization does not mean the eventual extinction of the human species, but rather the demise of a distinct civilization while others continue to arise and decline in lifecycles of their own.

If we are able to find long term counterexample trends of rise and decline beyond the perceived ‘death’ of a civilization – as with the examples of China, India, Italy, and Ireland – we may consider the claim to be falsified, and therefore the general hypothesis of life cycles should be discarded as an absolute principle. The careful scholar of Chinese or Indian history will note that these civilizations have been through numerous cycles of rise and decline rather than a singular pattern

⁶¹ Regarding Popper’s views of open systems see generally “Of Clouds and Clocks” in *Objective Knowledge* (1972, 206-255). See also Karl R. Popper, *The Open Universe, an Argument for Indeterminism* (1982).

⁶² See for instance Arthur Toynbee, *A Study of History* (1946/1953) and *The New Science of Gambattista Vico* (1744/1976).

⁶³ As with Arthur Schlesinger’s *Cycles in American History*.

akin to the lifecycle of a discrete living organism. Admittedly these counterexamples are open to debate: is modern Italy a part of the same lineage that includes that Etruscans and Romans, or is the modern West a continuation of the Classical West? – and lifecycle historicists may use all manner of counter argument to insulate their position from criticism.

If we look for examples that support our premises, we will certainly find them, but rather than look for white swans, we should be looking for black swans.

b. Testing Concepts: Ideational Comparison

The testing of ideas is conceptually similar to the testing of simple facts, although as with large events, the means are more interpretive in that here we are dealing with the comparison of concepts rather than the empirical corroboration of the correspondence of subject to object. The primary difference is that qualitative ideas are framed by metaphysical – not necessarily meaningless or false – rather than scientific or deductive statements. We may not objectify – externalize – such statements. In the language of engineering, the analysis of concepts is a machine with less precise tolerances than that of phenomena guided by fundamental laws.

For example, if a historian or philosopher makes the claim that the Greeks had no understanding of the idea of consciousness, and then we come across Socrates' distinction between subjective experience and nothingness in the death scene in the *Apology* (which are actually translated as 'consciousness' and 'unconsciousness' in modern English editions) we may conclude that they understood this idea.⁶⁴ This premise is weaker than the empirical example of the Lincoln photograph, and as a refutation, is more interpretive than the simple factual refutation of the arsenic and photograph examples. It is an interpretive comparison and the basis for a position in a critical discussion.

Here too when choosing the 'black swan' counterexample, we must therefore do so with the goal of maximizing risk to our theory. This said, we must acknowledge that the testing of historical ideas is a heuristic tool involving the comparison, contrasting, and discussion of concepts. It is not a hard deductive operation.

The great danger in discussions of ideas is that they will devolve into fights over definitions. In scientific discussions, definitions are virtually irrelevant. In activities like the law and historical analysis, definitions have an unfortunate importance. Thus participants in good faith critical discussions should minimize the distraction of definitional arguments and insofar as possible, they should agree upon the meaning of concepts beforehand. Popper loathed the quibbling

⁶⁴ See Plato, *Apology* (1942, 59). Plato reports Socrates as saying: Let us reflect in another way, and we shall see that there is great reason to hope that death is good; for one of two things – either reason to hope that death is a state of nothingness and utter unconsciousness, or, as men say, there is a change and migration of the soul from this world to another. Now, if you suppose that there is no consciousness, but a sleep like the sleep of him who is undisturbed even by dreams, death will be an unspeakable gain.

over definitions and believed that the discussion of concepts was unfruitful and to be avoided.⁶⁵ As Popper observes, we should focus on problems and questions rather than concepts. With ideational testing we are actually asking questions about a historical state of understanding. The only way to prevent such discussions from devolving into fruitless debates is for both sides to enter into discussion in good faith.

Regardless of the solidity of critical elimination in historical discussions and related activities like the adversarial process of the law, and the fact that they are sometimes logical and empirical in nature, their theories are usually not based on testable physical laws like those used of physics or chemistry. They are based on often contested historical facts that cannot be replicated. In this sense they rely more on rational explanation and interpretation of events than on the objectively testable nature of phenomena represented by physical laws. The most solid of refutations in history are those of ideas that are shown to be logically or physically impossible.

c. Coherence as Testing

In *Objective Knowledge*, Popper explains that there are 'three main theories of truth.' Of these,

The oldest was the correspondence theory, the theory that a statement is true if (and only if) it corresponds with the facts, or if it adequately describes the facts. This is the theory which I think Tarski has rehabilitated. Second is the so-called coherence theory: a statement is regarded as true if (and only if) it coheres with the rest of our knowledge. The third theory is that truth is pragmatic utility or pragmatic usefulness. (Popper 1972, 308)

In history we are interested in truth as correspondence. Falsification to determine truthful correspondence should never be confused with coherence – which for our purposes is the confirming of historical claims with the existing state of accepted knowledge. This is conformity and not correspondence. Although historical coherence takes the form of a test – that of comparison – it is a form of justification and therefore tells us nothing new (at best, it is testing against tested knowledge). A theory must correspond with an event in the real world rather than cohere with what is merely accepted. In this sense, we may think of coherence as the 'lateral' testing of ideas (the comparison of one idea with another idea or ideas believed to be true), and correspondence as 'vertical' testing of idea against object. As we have seen, the object may be a physical event or an idea.

This is why historical research using secondary sources is a justificationist activity, unless we add the element of *evaluation* or the process by which historians constantly test their claims in an informal sense. As such, the evaluation of sources involves the testing of both the existing source and the new claim.

⁶⁵ For Popper's comments on the discussion of definitions, see *Objective Knowledge* (1972, 28, 58, 124, 310-312, 327-328).

Evaluation is testing and therefore saves research from being just another form of justification. Even here we must concede that the process of evaluating primary and secondary sources is still closer to coherence than to correspondence and is therefore inherently weaker than purely empirical corroboration. Secondary sources are also a shortcut to primary sources. All of this underscores Paul Feyerabend's observation that the real practice of science is much less neat than Popper's distillation of the process would have us think. Feyerabend's claim is even more applicable to the analysis of history.

There is of course a practical problem – that of taking accepted or 'certified' knowledge as a given and the assumption that it has already reasonably tested. The danger is that certifying a new theory based on its coherence with the existing state of knowledge may actually perpetuate mistaken ideas.⁶⁶ Even though there was a time when the consensus of informed opinion held that the Earth was the flat center of the universe, we must make an assumption of the current state of knowledge as a starting point but not as authority. After all, most of us know of scientific truths from having read about them rather than having done the experiments ourselves (Miller 1985, 50). This is even truer of our historical understanding.⁶⁷

An idea may not be justified by comparing it against the current state of knowledge, but the current state of knowledge can be brought into question with a new theory, or else may be regarded as a baseline to be corroborated, added to, modified, or refuted. This is how our understanding progresses. Therefore, as with scientists, the greatest historians as truth-seekers are those who smash paradigms rather than shore them up. "For the problem is, of course, whether 'the unanimous testimony of historians' is to be accepted, or whether it is, perhaps, to be rejected as the result of their reliance on a common yet spurious source." (Popper 1965, 24) We can see that authority has no bearing on the truth, and thus we are back to the elimination of error.

We must also be careful not to confuse or conflate coherence testing with ideational testing. The testing of one idea with another in order to corroborate a claim is a kind of soft critical comparison. Although highly interpretive, it is a form of testing in that we are comparing a historical idea relative to the current understanding of the same idea. Idea is therefore contrasted with idea. This is opposed to coherence, where a new interpretation is compared for its conformity with the existing state of accepted knowledge.

⁶⁶ For example, an attorney friend once told me that historians of the American Civil War who were not lawyers tended to rely uncritically on J. G. Randall's *Constitutional Problems under Lincoln*. Because, he believed, that this book was incorrect at points in its constitutional analysis, the reliance of later historians on this book enshrined these errors into the historiography of the period.

⁶⁷ See Popper, "Sources of Knowledge and Ignorance" *Conjectures and Refutations* (1965, 21-29).

d. Pragmatic “Truth” (Utility)

The only time we are interested in pragmatic ‘truth’ (utility) as a measure for historical interpretations is when dealing with an incomplete record (Sub-Roman Britain for example) and must conjecture working theories out of ignorance. Here archaeological speculation replaces or fills in the blanks of a dearth of historical knowledge and where the sources are dubious, inaccurate or contradictory. For questions of chronology and the chronological sequence of events and objects, archaeological techniques, like dendrochronology and seriation may help provide important basic information.⁶⁸

Here we find another parallel between history and the physical sciences. Physicists are able to make precise predictions in quantum mechanics; they can answer the ‘what’ questions without knowing ‘why.’ The results of quantum mechanical formalism are knowable in a probabilistic (as opposed to a one-to-one deterministic sense), but we have little understanding of the actual physical phenomena we are describing. As Roger Penrose observes, “[i]t is a common view among many of today’s scientists that quantum mechanics provides us with *no* picture of ‘reality’ at all!” (2005, 782)⁶⁹ Because of this, and because of the incompatibility of quantum mechanics with Special and General Relativity, we live in an age of roadblocks in physics that is equally a Golden Age of cosmological speculation not seen since the Pre-Socratics. The point remains: we may not know what is going on in the quantum world, but we can make predictions about quantum outcomes. The purpose of history is not predictive in nature but we may seek to modestly fill in the gaps in our understanding with theories that provide utility rather than factual correspondence. Where structural/realist explanations exist, they are to be preferred to pragmatic explanations – even in the terms of pragmatic utility. Where explanations of correspondence are not possible, utility must suffice.

D. “Facts”

The fact that science and history can both be described by Popper’s schema suggests an affinity. The primary distinction arises from the intrinsic differences

⁶⁸ On the archeological dating technique of seriation, see James Deetz, *In Small Things Forgotten* (1977, 64-90).

⁶⁹ Thus we have a range of cosmological models attempting to account for the phenomena of quantum mechanics from those positing that the nature of reality is indeterministic (like that of the Copenhagen interpretation) to a kind of ultimate determinism claiming that every possibility happens and therefore every possible universe will exist, such as in the “Many Worlds” interpretation of Hugh Everett. On the different interpretations of quantum mechanics, see generally Max Tegmark, *Our Mathematical Universe* (2014). For a popular primer on quantum mechanics, see David Z. Albert, *Quantum Mechanics and Experience* (1992). For the basis of the “Many Worlds” Interpretation, see Hugh Everett, III. (1957), ‘Relative State’ Formulation of Quantum Mechanics (1957, 454-462). For a present-day critique of Everett’s multiverse, see Lee Smolin, *Einstein’s Unfinished Revolution* (2019 153-180).

between behavior of phenomena and process of the natural sciences governed by fundamental laws, and broader human events that are not and therefore the distinction between what we call scientific facts versus what we call facts in history. What then are facts, and what are their respective roles in the interpretive model sketched above?

In popular usage, facts are supposed to embody indisputable, immutable, and universal truths (perhaps akin to the 'atoms of truth' and protocol statements of the logical positivists), such as the fundamental laws of physics.⁷⁰ The idea is that reality can presumably be broken down into such statements.

Facts as observation statements have a different role in the realms of history, business, psychology, culture, economics, politics, and the law, than they do in the study of purely physical interaction. In science, facts represent quantifiable patterns based on or governed by fundamental laws that can often be demonstrated inter-subjectively.⁷¹ This is not the case in discussions in history, where 'facts' purport to represent alleged events and sequences of events that in most instances cannot be replicated.⁷² Consequently, so much of historical discussion is characterized by educated conjectures. Given that all history is selective and the great majority of facts can never be known, it is easy to see why historical conjectures can pull in so many different directions; once we agree upon the terms, the dates and places of Lincoln's birth and death can be known with relative certainty, but his innermost thoughts on race will be the topic of never ending discussion.

Even with the limitations of historical facts, we can in some instances test interpretive (as opposed to narrowly factual) historical conjectures.

III. Conclusion

Error is the mother of Knowledge, and the history of the birth of Knowledge out of Error is the history of the human race.⁷³ (Richard Wagner)

⁷⁰ On logical positivism and protocol statements, see Popper, *Conjectures and Refutations* (1965, 39-41), *The Logic of Scientific Discovery* (1935/1959, 95-97), and A.J. Ayer, *Philosophy in the Twentieth Century* (1982, 199-200). See also, A.J. Ayer, ed., *Logical Positivism* (1959), A.J. Ayer, *Language, Truth and Logic* (1946) and Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (1961).

⁷¹ See note 38.

⁷² The late Stephen Hawking and Leonard Mlodinow have suggested that John Wheeler's delayed-choice experiment indicates that events in the past may not be 'fixed.' See *The Grand Design* (2010, 82-83).

⁷³ In the spirit of critical rationalism, I do not accept Popper and his ideas uncritically. For instance, I think that the criticisms of Paul Feyerabend, one of Popper's most brilliant students and a philosopher widely regarded to be an apostate of the Popperian outlook, have a good deal of validity. In my opinion, Feyerabend's criticisms make critical rationalism more nuanced, more careful, and less strident. They strengthen Popper's program. Feyerabend is correct that Popper distills the scientific method to an outline of clarity and simplicity that rarely exists in the real world. But Feyerabend's anarchistic 'anything goes' approach to science – while

Out of error comes knowledge, and from testing, this knowledge progresses toward truer answers. Even if the heart of the cosmos is one of epistemological darkness, tested knowledge is real knowledge and critical elimination is a valid, if limited, method. It is *the* method. Even if all investigation ultimately arrives at a position of Kantian doubt holding that there are limits to what we can know, or that in most cases we cannot know at all, we will still come to know this great negative truth through reason, through testing.⁷⁴

In practical terms, the question for the historian is: how do we eliminate error, or rather, how do we test our theories? Popper says that we arrive at truer explanations through the failure of vigorous attempts to disprove our own ideas.

In history we must resist the urge to confirm what we already believe and must never insulate theories from criticism – things that require moral courage and integrity. This is a view that requires and inspires honesty with oneself as regards the topic of investigation. No pet theory or favorite interpretation should ever be so dear as to deter us from the most rigorous attempts to disprove it in the strongest way possible through the use of the historical counterexample, historical falsification. As Nietzsche observes, “[a] very popular error: having the courage of one’s convictions; rather it is a matter of having the courage for an

perhaps more reflective of how science really works – can be completely accounted for in Popperian terms in that it still involves testing and the elimination of error even if it is by less formal, *ad hoc*, unsystematic, or accidental means. When boiled down to its elements, the method is still one of trial and error and Feyerabend’s depiction of how science works still requires inspiration, conjecture, and refutation. If not, then there would be no difference between science and pseudoscience (and non-science generally), and yet this does not seem to be the case. In the spirit of critical rationalism, I do not accept Popper and his ideas uncritically. For instance, I think that the criticisms of Paul Feyerabend, one of Popper’s most brilliant students and a philosopher who is widely regarded to be an apostate of the Popperian outlook, have a great deal of validity. Feyerabend’s criticisms, in my opinion, make critical rationalism more nuanced, more careful, and less strident. They strengthen Popper’s program. Of course the great irony of all this is the lesson that critical rationalism should not be taken uncritically for an authority. And yet the idea that no body of knowledge is authoritative and that no theory or sets of theories should ever be beyond earnest revisiting are themselves foundational tenets of critical rationalism. I should also note that although I am a realist and a rational skeptic, I am not as optimistic as Popper about what we can know. There are several cosmological models that appeal to me – the ideational universe of Plato, the elegant classical model of Special and General Relativity, and the brilliant outline of Leibniz’s *Monadology*. There are also models that I find unsettling, but which I would accept as true, if shown to be. The “Many Worlds” model of Hugh Everett, III, is the paragon example of these. And yet as I have grown older, I have fallen into a Kantian doubt; perhaps the world is fundamentally beyond our kin. But even with our weak tools of reason and observation, we must do the best we can. We cannot know everything but perhaps we can know more, no matter how tenuous and conditional. My point is that I believe that science gives us something real and that its knowledge is progressive. And that is something. See, Paul Feyerabend, *Against Method* (1993). On Popper’s agreement with Kantian doubt, see *Conjectures and Refutations* (1965, 194).

⁷⁴ Richard Wagner, quoted by R.J. Hollingdale in *Nietzsche, the Man and his Philosophy* (1965/1999, 61).

attack on one's convictions!!" (Kaufmann 1974, 19) As scholars committed to truth, we must lead the attack ourselves. We must invite honest criticism. Like the combat officer whose position has been overrun, we must call in fire on our own coordinates.

No historical interpretation is ever final and no historical information no matter how well tested should ever be beyond revisiting. History should not be "argument without end," (Geyl 1955) but rather good faith discussion without end. As Popper notes, "I may be wrong, and you may be right and by effort we may get nearer to the truth." (1945/1994, 431)⁷⁵ As it is with history, so it is with the philosophy of history, and if I am lucky, somebody will take notice of the premises of this article and try to disprove them, so that by effort we may get nearer to the truth.

Bibliography

- Albert, David Z. 1992. *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press.
- Aristotle. 1908. *Metaphysica*, translated by W.D. Ross. Oxford: Clarendon Press.
- Ayer, Alfred Jules. 1946. *Language, Truth and Logic*, 2nd ed. New York: Dover Publications.
- , ed. 1959. *Logical Positivism*. New York: MacMillan Co.
- . 1973. *The Central Problems of Philosophy*. New York: Holt, Rinehart and Winston.
- . 1982. *Philosophy in the Twentieth Century*. New York: Random House.
- Barlett, John and Justin Kaplan, general ed. 2002. *Bartlett's Familiar Quotations*. 17th ed. Boston: Little, Brown and Company.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Moulton Publishers.
- . 2006. *Language and Mind*. 3rd ed. New York: Cambridge University Press.
- Deetz, James. 1977. *In Small Things Forgotten*. New York: Doubleday.
- Donald, David Herbert. 1985. *Lincoln*. New York: Simon & Schuster.
- Duggan, Michael F. 2002. *Chauncey Wright and Forward-Looking Empiricism*. unpublished Ph.D. dissertation, Georgetown University.
- Eccles, John C. 1977/1986. *The Self and Its Brain*. New York: Routledge.
- Everett, Hugh III. 1957. "'Relative State' Formulation of Quantum Mechanics." *Reviews of Modern Physics* 29: 454-462.
- Feyerabend, Paul. 1993. *Against Method*. New York: Verso, 3rd ed.
- Gaddis, John Lewis. 2004. *The Landscape of History*. New York: Oxford University Press.
- Geyl, Pieter. 1955. *Debates with Historians*. The Hague: Martinus Nijhoff.
- Gray, John. 2002. *Straw Dogs*. London: Granta Books.

⁷⁵ Quoted again in *The Myth of the Framework* (1994, xii).

Michael F. Duggan

- Hartshore, Charles, Paul Weiss and Arthur W. Burks. eds. 1943. *Collected Papers of Charles Sanders Peirce*. Cambridge: Harvard University Press.
- Hawking, Stephen. 1988. *A Brief History of Time*. New York: Bantam.
- . 1993. *Black Holes and Baby Universes*. New York: Bantam.
- Hawking, Stephen, and Leonard Mlodinow. 2010. *The Grand Design*. New York: Bantam.
- Hollingdale, Reginald John. 1965/ 1999. *Nietzsche, the Man and his Philosophy*, Cambridge: Cambridge University Press.
- Holmes, Oliver Wendell. 1918. "Natural Law." *Harvard Law Review*, 32 (1): 40-44.
- Hume, David. 1739/1888. *A Treatise of Human Nature*. Oxford: Clarendon Press.
- . 1993. "A Letter from a Gentleman to His Friend in Edinburgh." In *An Enquiry Concerning Human Understanding*, 115-124. Indianapolis, IN: Hackett Publishing Company.
- Hurley, Patrick J. 1988. *Logic*. Belmont, CA: Wadsworth Publishing Company, 3d ed.
- . 1999. *A Concise Introduction to Logic*. Belmont, CA: Wadsworth Pub. Co.
- Kaufmann, Walter. 1974. *Nietzsche: Philosopher, Psychologist, Antichrist* (4th ed.). Princeton, New Jersey: Princeton University Press.
- Lorant, Stefan. 1979. *Lincoln, a Picture Story of His Life*. New York: Bonanza Books.
- Madden, Edward H. 1986. "Max H. Fisch: Rigorous Humanist." *The Transactions of the Charles S. Peirce Society* 22: 375-396.
- McCullagh, Christopher Behan. 1984. *Justifying Historical Descriptions*. Cambridge: Cambridge University Press.
- McLean, Norman. 2008. "Billiards is a Good Game; Gamesmanship and America's First Nobel Prize Scientist." In *The Norman McLean Reader*, 78-92. Chicago: University of Chicago Press.
- Michelson, Albert Abraham. 1903. *Light Waves and Their Uses*. Chicago: University of Chicago Press.
- Millar, David, Jan Millar, John Millar, and Margaret Millar. 1996. *The Cambridge Encyclopedia of Scientists*. Cambridge: Cambridge University Press.
- Miller, David, ed. 1985. *Popper Selections*. Princeton: Princeton University Press.
- . 2000. "Sokal & Bricmont: Back to the Frying Pan." *Pli* 9: 156-173.
- . 2017. "Missing the Target, The Unhappy Story of the Criticisms of Falsificationism." <http://www.warwick.ac.uk/go/dwmiller>.
- Peirce, Charles Sanders. 1878/1940/1955. "How to Make Our Ideas Clear," In *Philosophical Writings of Peirce*, 23-41. New York: Dover.
- Penrose, Roger. 2004. *The Road to Reality*. New York: Random House.
- . 2005. *The Road to Reality*. New York: Alfred A. Knopf.
- Plato. 1942. "Apology." In *Plato*, edited by Benjamin Jowett. Roslyn, NY: Walter J. Black, Inc.
- Popper, Karl. 1935/1959. *The Logic of Scientific Discovery*. New York: Routledge.
- . 1972. *Objective Knowledge*. Oxford: Clarendon Press.

- . 1982. *The Open Universe, an Argument for Indeterminism*. New York: Routledge.
- . 1994. *The Myth of the Framework*. New York: Routledge.
- . 1965. *Conjectures and Refutations*. New York: Basic Books, Inc., 2nd ed.
- . 1945/1994. *The Open Society and its Enemies*. Princeton: Princeton University Press.
- . 1957. *The Poverty of Historicism*. London: Routledge.
- . 1977. *The Self and its Brain*. New York: Routledge.
- . 1976. *The Unended Quest: An Intellectual Autobiography*. London: Routledge.
- . 1998. *The World of Parmenides*. London: Routledge.
- Putnam, Hillary. 1996. *Pragmatism, an Open Question*. Oxford: Blackwell.
- Rockmore, Tom. 2000. "Interpretation as Historical, Constructivism, and History." *Metaphilosophy* 1/2: 184-199.
- Rothman, Kenneth L., and Sander Greenland. 2005. "Philosophy of Scientific Inference." In *Handbook of Epidemiology*, edited by Wolfgang Ahres and Iris Pigeot, 51-52. Berlin, Heidelberg: Springer-Verlag.
- Russell, Bertrand. 1946. *History of Western Philosophy*. New York: Simon and Schuster.
- Schlesinger, Arthur M. Jr. 1986. *The Cycles of American History*. Boston: Houghton Mifflin Company.
- Smolin, Lee. 2019. *Einstein's Unfinished Revolution*. New York: Random House.
- Taleb, Nassim Nicholas. 2007. *The Black Swan*, New York: Random House.
- Tarski, Alfred. 1956/1983. *Logic, Semantics, Metamathematics*. 2nd ed. Indianapolis, IN: Hackett Publishing Company, Inc., J.H. Woodger ed.
- Tegmark, Max. 2014. *Our Mathematical Universe*. New York: Alfred A. Knopf.
- Toynbee, Arthur. 1946/1953. *A Study of History*. Oxford: Oxford University Press.
- . 1744/1976. *The New Science of Gambattista Vico*. 3rd ed. Ithaca: Cornell University Press.
- Tucker, Avirzer. 2004. *Our Knowledge of the Past*. Cambridge: Cambridge University Press.
- White, Michael, and John Gribbin. 1992. *Stephen Hawking: A Life in Science*. New York: Penguin Books.
- Wilson, Edward O. 1978. *On Human Nature*. Cambridge: Harvard University Press.
- . 2002. *The Future of Life*. New York: Alfred A. Knopf.
- . 2012. *The Social Conquest of Earth*. New York: W. W. Norton & Company.
- . 2017. *The Origins of Creativity*. New York: W. W. Norton & Company.
- Wittgenstein, Ludwig. 1961. *Tractatus Logico-Philosophicus*. London: Routledge.
- de Wolfe Howe, Mark, ed. 1952. *Holmes-Laski Letters* v. II. Cambridge: Harvard University Press.

ATTENTION DEFICIT: ALIENATION IN PLATFORM CAPITALISM

Samson Liberman

Abstract: The aim of this paper is a socio-philosophical analysis of attention deficit phenomenon, which is being detected at the intersection of several subject areas (psychiatry, theory of journalism, economics). The main methodological instrument of the study is a Marxist principle of alienation. Alienation of attention, which, on the one hand, is being understood as a process of producing attention as a commodity, and on the other one – as the process of producing a person as a user of the platform, provides the methodological basis, necessary for a holistic view of the phenomenon. The main differences of attention alienation from alienation of labor and desire are considered within the paper. The possibility of a modern form of alienation is associated primarily with the emergence of the new forms of capital – platforms, providing infrastructure for the interaction of other users and aimed at collection and procession of large amounts of data. The main aspects of attention management: game, content sharing and design have been distinguished within the paper. The main consequences of alienation of attention for the structure of the individual and society have been spelled out. The effects of the spread of gaming techniques of attention management and content distribution techniques specific to social networks have been considered. It being is suggested that there is a correlation between the spread of ADHD diagnosis and the spread of attention management technologies, and, as well, between the distribution of attention management technology and the ‘renaissance’ of social in the social theory.

Keywords: alienation of attention, digital capitalism, platform capitalism, social networks, attention deficit.

Introduction

Attention deficit is originally a medical term for a specific behavioral disorder. Within the confines of psychiatry, attention deficit is a low ability to concentrate while communicating or completing tasks. Attention deficit hyperactivity disorder (ADHD) in modern medicine is being classified as a nervous system disease (ICD-11 2019). Bernard Steigler (Steigler 2006) and Geert Lovink (Lovink 2012) consider the influence of the modern society to be the cause of the spread of ADHD symptoms. Such a politicization of the attention deficit discourse refers to Nick Srnicek’s program book, Platform Capitalism (Srnicek 2017).

The place of attention deficit in modern capitalism is considered thoroughly within such a research field as an ‘attention economy’. This term is commonly ascribed to Michael Goldhaber (Goldhaber 1997). Herbert Frank is also considered to be one of the founders of the direction. The latter believes that the

main sign of the modern economy is a merger of money and attention, their mutual expression through one another (Franck 1999). Ethan Zuckerman writes about the collapse of the hopes of the “globalists” and relates it, among other things, to the limitation of our attention (Zuckerman 2013). Attempts to consider the problem of attention deficit from the perspective of journalism theory have been made in the works of the Russian specialists V.V. Dekalov (Dekalov 2017) and Biryukov V.A. (Birjukov 2016).

Therefore, there are a number of attempts made to problematize the phenomenon of attention deficit from the viewpoint of particular sciences – economics, medicine, design, marketing, and journalism. These approaches to the problem are heterogeneous not only in terms of method, but also in terms of the subject: not infrequently, the notions ‘attention deficit’ and ‘attention’ imply different phenomena.

Three different concepts might be distinguished, which are denoted by the term ‘attention deficit’. In medicine and psychology the said terms imply a low ability for arbitrary concentration of attention and abstraction. For the theory of media, ‘attention economy’ is a set of approaches and techniques for retaining someone else’s attention that are used, for instance, by marketers (visual website design, sales funnel, etc.). For Goldhaber and Franck, ‘attention deficit’ is an irregular distribution of audience and recognition. Here, attention is understood as a special social capital that can be accumulated, leased and sold, and which is measured in the number of views, followers and social influence in general.

Despite the urgent need for a general analysis, there is still no philosophical or socio-philosophical approach to this problem, although attempts of coming up with such a generalization are being made by particular sciences: economics (Franck 1999) (Goldhaber 1997) and psychology (Kahneman 1973).

Alienation of Attention and Alienation of Desire

Bernard Stiegler is one of those scholars, who are trying to comprehensively approach the problem of attention deficit. For him, ADHD is not only a personality disorder or mental disorder, but also a symptom of a ‘disorder’, existing in society. B. Stiegler attributes the spread of ADHD with intoxication by consumption, depreciation of values and atrophy of the ability to desire.

In a hyper-industrial society, ‘any value must be fully quantifiable, in other words, any value is doomed to complete depreciation’ (Stiegler 2006). Total calculation of cost for any object of consumption destroys its value and, in turn, the agent, who attaches value to a particular thing through desire. Being unable to desire and to focus his own attention, a person becomes unable to communicate and socialize.

B. Stiegler applies Marxist understanding of the alienation of labor (from product, from activity, from the generic nature of a human being) to desire. Following Freud and Lacan, he understands desire as a factor constituting a person as a subject. The internal logic of this alienation can be represented

through three levels (similar to the alienation of labor in Marxist theory): alienation from objects of desire, from the ability to desire and the 'generic' nature of a person as a willing being.¹

Not only does hyper industrial society possess the industry of goods production, but also the industry of desires or needs production. The industrially produced values, that is, the objects of desire, are a kind of fetish, since the process of their production is hidden from the agents of desire. Therefore values and objects of desire are alienated – this is the first level of alienation.

Since values are alienated and do not belong directly to agents, the very ability to desire – a wish as a special process, turns out to be alienated as well: I delegate my ability to desire to someone or something else (consultant, television, etc.). This is the second level. And, finally, a person is alienated from his 'generic nature.' She desires, but not herself and not for herself. If in Marxist theory the alienation of labor produces the worker as a commodity, then according to Stiegler, the alienation of desire produces consumer demand, that is, namely, it produces a consumer as a measurable economic quantity.

B. Stiegler connects the spreading of ADHD with the inability to desire. In our opinion, this provision is not enough justified. Despite the fact that B. Stiegler's implementation of the principle of alienation extends his analysis to all three aspects of the concept of 'attention deficit' in question – ability, technique, and capital; it is desire, but not attention he refers to. Accordingly, the three aspects mentioned above are: first – the ability to desire, to want, to feel the need for something, etc.; second – techniques and approaches for the creation of desires and needs, which are described quite thoroughly both by critics of the consumer society and by applied sales manuals; third – desire capital, that is, the actual quantity of sales or consumer demand. In our opinion, the deficit of desire does not automatically transfer into a deficit of attention. In the medical aspect, deficit of desire will be expressed more through apathy and depression than through weakening of attention concentration and hyperactivity.

It is the symptoms of depression (the loss of meaning, purpose and desire) that are attributed to the consumer society by the classics of this concept: the 'mass' of J. Baudrillard, the 'automatons' of E. Fromm, the 'society of the spectacle' by G. Debord, etc. The example drawn by B. Stiegler himself also illustrates depression rather than ADHD: the Cartier couple, who tried to poison their children and commit suicide, from the point of view of B. Stiegler, took this step because of the despair and dislike that inevitably arise when intoxication by consumption. They wanted to shield their children from this despair (Stiegler 2006). Symptoms of desire deficiency – passivity and isolation are almost opposite to attention deficit symptoms – weakening of concentration and hyperactivity.

¹ This question of the production of values, symbolic value and needs is discussed in detail, for example, by J. Baudrillard, criticizing the concept of natural needs of Marx. You can also turn to the analysis of the alienation of the ability of desire from Deleuze and Guattari in the *Anti-Oedipus*.

Nevertheless, the idea of applying the scheme of analysis of alienated labor to desire can be fruitful while studying attention deficit. If we extend the logic of Stiegler's 'hyper-industrialization' to the phenomenon of attention deficit, we can obtain the concept of 'attention industry'. The difference between the attention industry and the desire industry will be represented through the emergence of other techniques and methods of constructing the subjectivity of participants in social relations. Alienated or industrialized labor produces workers as goods, alienated desire produces consumers, alienated attention, in turn, produces users. Moreover, the following forms of alienation do not abolish the previous ones: the same participant of social relations can act either as a worker, or a consumer, or a user.

Platform Capitalism as an Industry of Attention

Thus, attention deficit might be represented as the next form of exploitation and alienation of a human being, which is inherent to contemporary high-tech capitalism. However, the concept of capitalism in general and modern capitalism in particular also needs to be defined. Within the framework of the given paper, we define it as 'platform capitalism' in the understanding of Nick Srnicek (Srnicek 2017).

'Platform capitalism' is characterized by the emergence of platforms as fundamentally new forms of capital. N. Srnicek calls platforms a special type of firms that can effectively monopolize, extract, analyze and use the growing volumes of recorded data (Srnicek 2017). Platforms are also intermediaries for other economic agents: these are digital infrastructures that allow two or more groups to interact (Srnicek 2017). But, due to its digital nature, platforms have the ability to record the slightest movements and transactions within themselves; that is, to produce data. Being accumulated in sufficient quantities, these data become capital and begin generating income. They can be used to gain a competitive advantage, attract advertisers, or modify the platform itself. This is, namely, the principle of operation of the 'gang of four' (GAFA – Google, Amazon, Facebook, Apple) keeps to.

Access to the platform for users is always provided free of charge, or at a wittingly underestimated price, since the main source of income is the extraction and use of users' data. Therefore, the platform is designed to be as attractive and comfortable for using as possible. Apart from attracting users, the platform always has to deal with the task of retaining them. Therefore, they use attention management technologies. The main difference between modern attention management technologies is the focus on data collection.

Users only produce data if they perform certain actions: clicks, views, and so on. Accordingly, the goal of attention management technologies is to make the user watch, click and swipe by all possible means which we will thoroughly consider below.

Thus, the contradiction between medical (low ability to arbitrary concentration) and economic (limitation of attention as an economic resource) understanding of deficit might be overcome in the notion 'alienation of attention'. At the same time, such alienation is fulfilled in the conditions of modern capitalism, which we refer to as 'platform capitalism'. Alienation of attention is a consequence of the use of attention management technologies by platforms. Unlike narrative or ideology, which also deal with audience's attention, modern attention management technologies do not have the task of inciting action or thought. They also differ from the 'reality show', detailed by J. Baudrillard, because they have no task to arouse desires and emotions in us. Their main task is to instigate the users to perform actions on the platform (clicks, views, etc.), for what it is sufficient to simply keep their attention.

In our opinion, attention management on digital platforms has three clear directions: *game, content sharing, architecture or platform design*.

For example, various online tests always contain an element of the game: we have a problem, we are provided with a number of solutions and at the end, we gain a certain result. At the same time, data collection is practically not hidden; we may even see some results of their processing. Almost every platform uses gaming techniques: Google allows us jumping over cactus for a dinosaur, social networks give points for filling in data about ourselves, Tinder is entirely built like roulette, etc. The main principle of any game as an attention management technology is a reward system, when for a certain combination of actions we are promised a reward in the form of points, a beautiful picture with jingle or any other attribute of the winner. These efforts taken by platform developers are aimed at one goal: to retain the users' attention, to further encourage them to produce data.

Another way to make us click the mouse is to share content. After the advent of WEB 2.0 Internet resources, everyone gained access to the global audience. From the passive mass consumer, the layman turned into the author of the content. This blurring of the boundaries between production and consumption of content was earlier interpreted in the direction of consumption (for example, by J. Baudrillard): a television show is when the mass looks at itself (Baudrillard 1994). Nowadays not only does the mass watch, but also records itself. The principle of content 'sharing' lies at the heart of social networks, video hosting and similar platforms: 'Tell your friends what is new with you', 'Learn what's new with your friends!', 'Do not forget to assess and leave a comment!'. Passivity and inter-passivity transforms into activity and inter-activity.

By platform design, we mean the very form of the platform, which is also aimed at retaining user's attention. Design is an interface that the user directly comes across with, that is, the external design of the platform, for example, the color and shape of the buttons, the design of the video sequence and so on. For example, Google spends a lot of resources on identifying the so-called 'color of money', the color of that cherished action button (Holsen 2009).

Despite the fact that a certain type of attention retention technologies is inherent to a particular type of platform; platforms normally try using the entire range of capabilities available. Social networks use gaming techniques, entertainment platforms, the media often offer to register and begin sharing content, and the design of all platforms is equally tailored to hold attention.

Thus, attention deficit can be considered as an industry of alienated attention, similar to alienated labor in Marxist theory or alienated desire in the theories of consumer society. With the alienation of labor, capital produces needs for public agents; its task is to create labor force, to induce people to work and activity as production of goods. With the alienation of desire, capital produces the needs of public agents; its task is to create consumer demand. In alienating attention, the task of capital is to produce users and hold their attention. The platform does not sell anything to users, but it sells users as data and audience. The platform does not produce anything, but collects data produced by us.

In all previous forms of alienation, the management of the attention of the others is also present in a hidden or an indirect form. So, for example, in order to induce the release of a product or to overthrow the political regime, it is necessary to attract public attention (design of posters and leaflets, slogans, manifestos, narratives, etc.), but attention alone is not enough – you need to encourage people to take particular actions. Attention is being alienated, but this happens indirectly during the alienation of labor. Also, with the alienation of desire, the final objective of the mechanisms of the consumer society (mass cinematography, reality shows, etc.) is to cause certain desire and perception in a person, resulting in consumption. The alienation of attention here occurs as a concomitant process. With the advent of platforms, it has become possible to capitalize on pure attention, without prompting production or consumption.

The theory of alienation, in our opinion, has an advantage over other theories and methods of analysis of attention deficit. For example, the theory of commodification of the audience, which D. Smythe creates (Smythe 1981), quite accurately describes the economic and technological processes, but overlooks the psychological or medical aspect of attention deficit. Also, the attention economy created by M. Goldhaber (Goldhaber 1997) and G. Franck (Franck 1999) ignores the consequences that occur with a person and her ability to concentrate when she commodifies attention. Understanding attention deficit as an alienation allows one to explain, on the one hand, the emergence and spread of such a medical diagnosis as ADHD, and on the other, the advent of the new forms of capital related to the distribution of users' attention.

Consequences of Alienation of Attention

The 'renaissance' of social, which we can observe in social theory today, owes to interactivity of social networking platforms. If in the second half of the 20th century, theorists considered 'de-realization' or 'de-composition' of society to be the main line in social philosophy, (Furs 2002, 16), today 'speculative realism' is

going from strength to strength. The main accusation among modern philosophers is the lack of 'reality' and 'actant nature' of objects.

The revival of the social is as well indicated by the radicalization and politicization of public space. Syrian conflict, American wall, struggle for the rights of the feminist movement and sexual minorities etc. We are not talking about the said phenomena themselves or about similar events had happened before, but rather about the way in which they are discussed in society and what political effect they produce. As we have written above, the modern 'mass' has ceased being a 'silent majority', social networks have found a way to turn passive consumers into active users (Lovink 2012). Political identity, of course, did not supplant consumer identity, but once again became one of the most important mechanisms of socialization and social life in general. Although, as Lovink observes, the question of whether 'friends' from FB can turn into 'comrades' remains open (Lovink 2012).

In his essays 'Treatise on Comment Culture', and 'Chronic Narcissus: Minimal Selfie Technology' he thoroughly analyzes in the mechanisms of the production of social in social networks. "... Desperate attempt to be heard, to achieve any impact..." (Lovink 2012) : this, according to Lovink, is the main goal of users of social networks. In the era of the alienation of labor, manufacturing was the function of social differences production, in the era of the alienation of desire consumption played this role, and, in turn, in the era of the alienation of attention, attention and influence perform this function. A new term – influencer has appeared nowadays for making reference to the elite. Status in society is presently determined neither by the number of material goods produced nor by the number of acquired ones, but by the number of likes and followers.

The aspiration to get as much attention as possible from the other users is as well reflected on the psychological level. However, the spread of ADHD, in our opinion, is associated not only with the constant pressure of social networks. The Internet design industry is as well not a leading factor when we talk about shaping the prerequisites for ADHD. In our opinion, the main cause of the ADHD epidemic is the spread of the attention-keeping gaming techniques we have considered above.

A fascination with computer games a decade ago had been stigmatized and considered by psychologists as an addiction with the aim to escape from a traumatic reality. Today, games are integrated into almost all types of communication and social practices: lending, retirement benefits, social networks, individual purchases, education, and so on. Everywhere we are asked to have an account, being motivated with balls and special status, etc.

Here we do not consider the game foundation of culture or the game family nature of a person, but game techniques for attention management. The main feature of these techniques is an externally posed task, the number of solutions for which is externally limited by special rules. And when solving this game problem, the participant achieves a special status of the winner, member of the

community, etc. In the framework of our study, such important components of the game as the plot, role, competition, etc. are not critical.

The crucial moment of the game as a technique for attention maintaining is the reward for a user. A player's reward is usually realized procedurally in the form of a demonstration of some bright event (fireworks, fanfares, colorful action scenes) or attributively through 'icons' next to the username, special clothing items of the player's 'avatar', 'ranks' and other distinctive attributes, emphasizing his special status.

It might seem that such techniques have been used for a quite long time and in almost the entire service sector: accumulative systems of discounts and bonuses, lotteries and sweepstakes, customer ranking (silver, gold, platinum), time-limited validity of an advertising campaign, etc. Also, similar techniques are used in labor or educational activities. However, only the emergence of platforms unleashed the full potential of gaming techniques. The ultimate goal of gaming attention management techniques is only the game process itself, since it is enough for the user to produce the data the platforms need.

The game tries to do everything so that the player does not have time to get bored. At the slightest difficulty in solving the problem, a tooltip helps us, highlighting the desired button, etc. In case of a long absence of the user on the platform, notifications and invitations are being sent to him, special return conditions are offered in the form of game bonuses, etc.

ADHD is a syndrome of a human being, who was brought up by platforms that use gaming attention management technologies. The main feature of the game is a limited set of solutions to the main conflict or problem; these solutions are always open and already offered to the player as ready-made ones. Also, the problem itself has always been posed from the outside by the organizer of the game. Accordingly, the skills of analyzing a situation, finding solutions and posing a problem for such a person are atrophied to a certain degree. In the absence of a reward system, to which a person became used to by playing games, his interest cannot be constant and arbitrary. The ability to self-focus is practically absent. Hyperactivity also stems from here, as a constant need for reward and approval, which a person has become accustomed to through gaming. This need for social recognition is being also fueled by social networks, as we discussed above.

An example of the action of ADHD is not a married couple trying to commit suicide and poison their children because despair and absence of love, but the generation of people who have spent last 10 of 15 years of their lives on various platforms for several hours a day. However, they cannot deal with any operation that requires more than two consecutive clicks on the 'unlit' buttons (e-mail, Excel, etc.).

The desire for recognition and attention has always been inherent to humanity (just think of Herostratus), but nowadays, attention retention has become an entire industry. Not only gamblers, but also victims of mass shootings

by students in schools, victims of many terrorist attacks, and those, who suffered from the participation in hazardous 'challenges' might be considered as its victims.

Conclusion

The main conclusion that is to be made as a result of the research carried out, is that attention can claim to be the central category of social philosophy. It can be represented as some superindividual active substance or fundamental human activity (by analogy with Hegel's 'Geist', Marx's 'Arbeit' and 'Wille' of Schopenhauer and Nietzsche), around the distribution and alienation of which social relations and man as the agent of these relations emerge.

It is noteworthy that the previous tradition tends to regard desire as such a category. This is, for example, the basic premise of Freudian Marxism, taken very broadly as a socio-philosophical mainstream. We have made an attempt to show the historical limitations of this approach and its inapplicability to certain phenomena of modern life (political populism, activism, ADHD and so on).

However, within the framework of the given paper we were interested not so much in metaphysics but in the consideration of historical-specific mechanisms of alienation of attention that are characteristic to the current stage of platform capitalism instead. We have tried to present the alienation of attention as a kind of industry characterized by special forms of social relations and human subjectivity. In our opinion, the basic technologies of social and human production nowadays are: content sharing, game and design of Internet platforms.

Initially, the article was conceived as an attempt to build a socio-philosophical justification for a number of concepts that have a common position about attention deficit (economics of attention, psychology of attention, attention management). Ontological and metaphysical justification of the category of attention may become a further direction for scientific research.

Bibliography

- Birjukov, Vladimir A. 2016. "Sredstva massovoj informacii v uslovijah razvitiya fenomena 'jekonomiki vnimanija.'" *Izvestija vysshih uchebnyh zavedenij. Problemy poligrafii i izdatel'skogo dela* 3: 91-95.
- Baudrillard, Jean 1994. *Simulacra and Simulation*. Michigan: University of Michigan Press.
- Zuckerman, Ethan 2013. *Rewire: Digital Cosmopolitans in the Age of Connection*. Danvers: WW Norton & Co.
- Dekalov, Vladislav V. 2017. "Vnimanie kak bazovyj resurs kommunikativnogo kapitalizma." *Rossijskaja shkola svyazej s obshchestvennost'ju: ezhegodnyj al'manah* 10. 27-39.
- Franck, Gerbert 1999. "The Economy of Attention." *Telepolis* 7 December.
- Furs, Vladimir N. 2002. *Kontury sovremennoj kriticheskoj teorii*. Minsk: EGU.

Samson Liberman

- Goldhaber, Michael H. 1997. "The Attention Economy and the Net." *First Day* 2. Number 4–7 April.
- Holson, Laura M. 2009. "Putting a Bolder Face on Google." URL: <https://www.nytimes.com/2009/03/01/business/01marissa.html>. Accessed 06.06.2019.
- ICD-11 for Mortality and Morbidity Statistics 2019. URL: <https://icd.who.int/browse11/l-m/en#/http%3a%2f%2fid.who.int%2fd%2fentity%2f821852937>. Accessed 06.06.2019.
- Lovink, Geert 2019. *Kriticheskaja teorija interneta*. Moscow: Ad Marginem Press.
- Smythe, Dallas W. 1981. "On the Audience Commodity and its Work." *Media and Cultural Studies: KeyWorks* 230. Accessed 06.06.2019.
- Srnicek, Nick. 2017. *Platform Capitalism*. New York: John Wiley & Sons.
- Kahneman, Daniel. 1973. "Attention and Effort." In *Englewood Cliffs 1063*, 218-226. NJ: Prentice-Hall.
- Stiegler, Bernard. 2006. *The Disaffected Individual in the Process of Psychic and Collective Disindividuation*. URL: <https://arsindustrialis.org/disaffected-individual-process-psychic-and-collective-disindividuation>. Accessed 25.03.2020.

Perfect Happiness

Daniel Rønnedal

Abstract: In this paper, I will develop a new theory of the nature of happiness, or “perfect happiness.” I will examine what perfect happiness *is* and what it *is not* and I will try to answer some fundamental questions about this property. According to the theory, which I shall call “the fulfillment theory,” perfect happiness is perfect fulfillment. The analysis of happiness in this paper is a development of the old idea that happiness is getting what you want and can be classified as a kind of desire-satisfaction theory. According to the fulfillment theory of happiness, it is necessarily the case that an individual *x* is perfectly happy if and only if all *x*’s wants are fulfilled. The interpretation of this basic definition is important, since the consequences of the particular version defended in this essay are radically different from the consequences of many other popular theories of happiness. The fulfillment theory is also quite different from most other desire-satisfaction theories of happiness. We will see that it has many interesting consequences and that it can be defended against some potentially serious counterarguments. The upshot is that the analysis of (perfect) happiness developed in the present paper is quite attractive.

Keywords: happiness, perfect fulfillment, desire-satisfaction theories, final ends, The Rational Will.

1. Introduction

Happiness has been studied for thousands of years by philosophers, poets, religious thinkers and theologians.¹ More recently, scientists have turned their eye to this phenomenon: psychologists,² economists,³ biologists⁴ and many others. There is even a whole journal devoted to the study of happiness: the *Journal of Happiness Studies*.⁵ Some recent philosophical contributions to the study of this subject include Feldman (2010), Haybron (2008), Martin (2012), and Russell (2012). Nevertheless, there is no consensus on what happiness is. In this paper, I will first briefly mention some different views on this issue. Then I will develop a new theory of the nature of happiness, or rather – what I shall call – “perfect happiness.” According to this theory, perfect happiness is perfect fulfillment. I will call this theory “the fulfillment theory of (perfect) happiness.” I will examine what perfect happiness *is* and what it *is not* according to this theory, I will prove some theorems that follow from it, and I will defend the theory against some possible

¹ Annas (1993), Bok (2010), McMahon (2005), White (2006).

² Boniwell, David, and Ayers (2013).

³ Bruni and Porta (2005).

⁴ Grinde (2012), Nes (2014).

⁵ For a collection of some papers published in this journal, see Delle Fave (2013).

counterarguments. The upshot is that the fulfillment theory of (perfect) happiness is quite attractive.⁶

The paper is divided into six sections. In Section 2, I discuss some preliminary linguistic and methodological questions and I propose a classification of various approaches to the topic of happiness. In Section 3, I turn to the explication of the theory developed in the present paper: the fulfillment theory of happiness. I try to render what I mean by this theory more precise by answering some fundamental questions. In Section 4, I will prove some interesting theorems that follow from our definition. I will focus on some of the most interesting consequences and try to explain their significance. In Section 5, I defend the theory against some possible counterarguments that are potentially quite serious. I try to argue that these problems cannot be used to refute the fulfillment theory of happiness. Finally, Section 6 contains a brief summary of the paper and a conclusion.

2. Theories of happiness

Different theories of happiness try to answer different kinds of questions: metaphysical (What is the nature of happiness? What kind of 'thing' is it?), linguistic (What does 'happiness' mean? Do sentences that include 'happy,' 'happiness,' etc. have truth-values?), scientific (What are the sources, causes and effects of happiness?), epistemological and methodological (How should we study happiness, what methods should we use?), ethical and metaethical (What is the value of happiness and how is happiness related to morality and rationality?), and so on.

In this paper, I will primarily be interested in the metaphysical questions. I will develop a theory of the *nature* of happiness. However, I will first say a few words about the meaning of 'happiness' and about my methods. Later, I will also consider some ethical questions.

Most people – for example, most thinkers mentioned in the introduction and later in this section – seem to agree that 'happiness,' 'happy' and similar words are ambiguous in natural languages.⁷ I share this view. At least, it seems obvious to me that various philosophers and scientists use these words to refer to different phenomena. When I use the term 'happiness' and speak about 'perfect happiness,' I am focusing on one important aspect of this concept. I do not deny that 'happiness' can be used in other interesting senses. Nowadays, for example, we often seem to use 'happy' as synonymous with 'feeling happy.' According to this interpretation, someone *is* happy if and only if (iff) she is *feeling* happy. But this does not seem to be the only sense of the word. My theory is therefore partly

⁶ For more on some empirical research on happiness, see, for example, Diener and Diener (1996), Easterlin (2003), Kahneman, Diener, Schwarz (1999), Myers and Diener (1995) and Seligman (2002).

⁷ See also Davis (1981) and Thomas (1968).

‘stipulative’ in the sense that I focus on one reading of ‘happiness.’ Still, even stipulative definitions of various concepts should not depart too much from our ordinary languages. The way I use the term is not arbitrary; the idea that happiness is getting what you want is very old and has been defended, in one form or another, by many philosophers throughout history. To be able to say something interesting about happiness, we should define what we mean by the concept. I try to do this in the present paper.⁸

According to the theory presented in this paper, the concept of happiness is not the concept of a purely mental phenomenon. We cannot immediately observe with our senses whether someone is happy and we cannot use introspection to decide whether or not we are happy, at least not in most cases. Happiness is usually not something immediately given in experience. It is not something in the mind (in contrast to the feeling of happiness), it is not something in the head and it is not something in the body. At least, it is not always and necessarily something *only* in the mind, head or body. Even if we had perfect knowledge of someone’s mental and bodily states, we would not normally know whether she is happy or not. Since perfect happiness is perfect fulfillment, according to our theory, we cannot know whether someone is perfectly happy or not without knowing whether the things she wants are true or not. If a mother wants her daughter not to suffer from some illness, for example, she is perfectly happy only if her daughter is *in fact* not suffering from some illness. Therefore, to know whether the mother is perfectly happy or not, it is not enough to know everything about the mother, we also need to know something about an objective state of the world – the health of the daughter.

Whether an individual is perfectly happy or not at a particular moment in time cannot usually be decided empirically at that time, for the content of what someone wants might be about another time, for example a state of affairs in the future. We cannot *now* directly observe the future, even though it might be possible *in the future* to observe what is the case *then*. Sometimes it is in principle impossible to decide whether someone is perfectly happy only by empirical methods, for someone can want abstract things that cannot be observed. Someone might, for instance, want a mathematical theorem, say the Goldbach conjecture, to be true. We cannot observe whether this theorem is true or not. Hence, we cannot observe whether this person is perfectly happy or not.

⁸ It is not obvious that ‘happiness’ is ambiguous in English. Feldman (2010, Appendix C) argues against this view. Suppose that he is correct about this. Even if it were true that ‘happiness’ is not ambiguous in English, we do not have to go on using this word in the same sense that we have so far. Words in natural languages do change meaning over time, and there may be good reasons for introducing a new reading. So, the fact that some scientific or philosophical theory about happiness uses the term ‘happiness’ in some unusual sense, is not a decisive argument against such a theory. For more on the use of ‘happiness’ in this paper, see Section 5 and footnote 36.

It is conceivable that it is in some cases possible to empirically decide whether an individual is perfectly happy. If all of an individual's wants are about present states of affairs that are empirically observable at a particular moment in time, then it is in principle possible to decide whether or not this individual is perfectly happy at this time. It might in principle even be possible for an individual to decide by introspection whether she is perfectly happy or not. If all an individual's wants are about her own present introspectable mental states and all her wants are introspectable, she might use introspection to 'observe' whether she is perfectly happy or not. Even though cases of this kind are possible, it does not follow that there are any actual examples of such conceivable situations.

It is, of course, possible to empirically study the feelings and physical, biological and psychological processes that go on in someone who is perfectly happy (in our sense) or who is happy in some other sense of 'happy,' for example, someone who is feeling happy or is 'satisfied' with her life. Nevertheless, according to the current theory, this is not the same thing as studying happiness itself. Perhaps one aspect of happiness can be studied by introspection and, in the future, even by brain scans: namely, what someone wants. Yet, to know whether these wants are satisfied or not, something more is usually needed.

'Happiness' is neither an evaluative term nor a normative term according to the fulfillment theory. Sentences that include 'happiness' are normally used to express beliefs that are true or false; they are not used simply to express our feelings or to prescribe certain things. The concept of happiness is a purely formal, intellectual or theoretical concept. 'Happiness' has more in common with logical expressions, such as 'everything,' 'something' and 'nothing,' than with empirical words such as 'red' or 'sweet,' according to the fulfillment theory. This does not entail that no aspects of what it means to be happy can be investigated empirically and it does not entail that empirical sciences do not have anything interesting to say about happiness, as obviously they do. Still, it is very difficult to study the kind of happiness that is described in this paper by empirical methods alone.

There are many kinds of theories of happiness: hedonistic theories,⁹ desire-satisfaction theories,¹⁰ life-satisfaction theories,¹¹ final end theories, eudaimonistic and well-being theories, and functional and self-realization theories,¹² subjective well-being theories,¹³ virtue theories,¹⁴ emotional state

⁹ Bentham (1781/1988), Feldman (2004, 2010), Mill (1863/1987), Sidgwick (1907/1981), Sprigge (1991), Tännsjö (2007).

¹⁰ Chekola (1974, 2007), Davis (1981b), McGill (1967), Perry (1926), Rawls (1972), Solomon (1976).

¹¹ Benditt (1974), Brandt (1967), Kekes (1982), Martin (2012), Nozick (1989), Suikkanen (2011), Sumner (1996, 2000), Tatarkiewicz (1978), Telfer (1980), Thomas (1968), van Praag and Ferrerer-i-Carbonell (2004), Veenhoven (1984, 1984b), von Wright (1963), Wilson (1968).

¹² Aristotle's *Nicomachean Ethics*, Franklin (2010), Russell (2012).

¹³ See Pavot and Diener (2013) for an introduction to theories of this kind, which are currently very popular among social scientists.

¹⁴ The ancient Stoics, see Annas (1993), Becker (1998).

theories,¹⁵ harmony theories,¹⁶ “supernatural” theories,¹⁷ pluralistic (objective list) theories,¹⁸ etc.¹⁹ The theory that I introduce in this paper can be classified as a kind of desire-satisfaction theory. Now, let us turn to this approach.

3. The fulfillment theory of perfect happiness

In this section I will develop the fulfillment theory of happiness, or perfect happiness. It is important to emphasize that it is a theory of *perfect* happiness and not of happiness, since it is possible to be happy without being perfectly happy and since it seems possible to talk about different degrees of happiness. ‘Happy,’ in contrast to ‘perfectly happy,’ is vague. In this sense the theory is about an “ideal.”²⁰ Hence, when I speak of ‘happiness’ in this paper, I usually mean ‘perfect happiness.’ I will not try to define what it means to be less than perfectly happy.

According to the theory, perfect happiness is perfect satisfaction or perfect fulfillment. More precisely, we shall use any of the following equivalent definitions:

Definition of perfect happiness

D1. It is necessary that, for every individual x : x is perfectly happy iff all x ’s wants are fulfilled (satisfied).

D2. It is necessary that, for every individual x : x is perfectly happy iff everything x wants is the case.

D3. It is necessary that, for every individual x : x is perfectly happy iff for every A , if x wants it to be the case that A , then A is the case.

¹⁵ Haybron (2001, 2005, 2008), Sizer (2010).

¹⁶ Perhaps Plato’s *Republic*.

¹⁷ St Thomas Aquinas’ *Summa Theologiae*.

¹⁸ Aristotle’s *Nicomachean Ethics*, Montague (1967).

¹⁹ It is not obvious how various philosophers and scientists should be classified. Some thinkers might belong in several categories. For example, Aristotle sometimes seems to defend a well-being theory of happiness, sometimes a functional theory, sometimes a virtue theory and sometimes a pluralistic theory (*Nicomachean Ethics*, 11–17; 1097a15–1099b10). The picture is complicated by the fact that not all individuals use the same language. Aristotle, for example, uses the Greek term “*eudaimonia*,” which is often translated as ‘happiness.’ Some seem to think that this is reasonable, e.g. Kraut (1979) and Annas (1993), others, that it is misleading, e.g. Sumner (1996) and Haybron (2008). How various thinkers should be classified will, of course, depend on exactly how the different theories are formulated. They are not necessarily defined in such a way that they are mutually exclusive. It is beyond the scope of this paper to discuss all of these various views about happiness. Note also that not all thinkers intend to study the same ‘thing’ when they study ‘the phenomenon’ they call ‘happiness.’ Therefore, these theories are not necessarily inconsistent with the fulfillment theory of happiness or with each other.

²⁰ One could argue that how happy someone is, is determined by how many of this person’s wants are satisfied. The more wants that are satisfied, the happier the person is. This is perhaps approximately true, but it does not seem entirely right to me. Some wants are more important than others. Having one fundamental desire fulfilled might be much more important for overall happiness than having many unimportant wants fulfilled.

These definitions can be interpreted in many different ways depending on what we mean by 'necessary,' 'every individual,' 'iff,' 'everything,' 'wants' and 'fulfilled.' To try to make the theory more precise, I will answer some questions about it.

Q1. What kind of theory is this and how is it related to other similar theories of happiness in the literature?

A. The fulfillment theory of happiness can be classified as a desire-satisfaction theory of happiness. According to theories of this kind, happiness is satisfaction of desires or wants or inclinations in some sense. However, there are many different versions of this type. Let us consider three important distinctions.

Firstly, there are *subjective* and *objective* forms. According to the subjective forms, the important thing is that we *believe* that our desires are satisfied, not that they *in fact are* satisfied. If someone *believes* that her desires are satisfied (even when they *in fact are* not), then she *is* happy. Some subjective theories emphasize the pleasure we often *feel* when our desires are satisfied or when we believe that they are satisfied. According to those theories *feeling* satisfied is a necessary and perhaps sufficient condition for happiness. Davis (1981b) is an example of a subjective form of desire-satisfaction theory of happiness. According to objective desire-satisfaction theories, the important thing is that our desires *in fact are* satisfied, not that we *believe* that they are satisfied or that we *feel* satisfied. It is neither necessary nor sufficient that we *feel* satisfied or fulfilled to *be* satisfied or fulfilled, according to objective forms, and it is neither necessary nor sufficient that we *believe* that we are satisfied or fulfilled; we can *be* satisfied without *feeling* satisfied and without *believing* that we are satisfied and we can *feel* satisfied and *believe* that we are satisfied without *being* satisfied. Chekola (1974, 2007) and Solomon (1976) are examples of objective forms. The fulfillment theory of happiness in this paper is an objective kind of desire-satisfaction theory.

Secondly, there are *actual* and *ideal* versions of desire-satisfaction theories. According to actual forms it is our *actual* desires that must be satisfied for us to be happy, and according to ideal forms it is our *rational* desires (or perhaps the desires we would have if we were perfectly wise) that must be satisfied. Chekola (1974, 2007) can be classified as an actual and Rawls (1972) as an ideal desire-satisfaction theory. The fulfillment theory of happiness in this paper is an actual kind.

Thirdly, there are *restricted* and *unrestricted* forms. According to unrestricted forms, *all* desires must be satisfied for an individual to be (perfectly) happy; according to restricted forms, only *some* desires must be satisfied, for instance, our most 'important' desires, or our 'now-for-now' desires, or our desires based on true, justified, rational beliefs, or desires about our own lives or about our own subjective, conscious experiences, etc. Chekola (1974, 2007) and Solomon (1976) are examples of a kind of restricted version; Chekola, for example, focuses on "global" desires (relatively permanent, comprehensive and important desires), not "local" desires. Kant, in some places, appears to defend an

unrestricted form (see, for example, Kant (1788/1997, 104; 5:124)). Even the expression “all desires” can be interpreted in many ways. Let us consider one important distinction. There are *eternal* unrestricted and *temporal* unrestricted forms. According to the eternal version “all desires” means “all desires at all times”; according to the temporal version “all desires” means “all desires that the individual *c* has at the particular moment when the sentence ‘*c* is (perfectly) happy’ is evaluated.” The fulfillment theory of perfect happiness is a temporal, unrestricted form. This means that individual *x* is perfectly happy at time *t* iff *absolutely all* wants *x* has at *t* (but not necessarily at other times) are satisfied at *t*.²¹ An individual can, therefore, be perfectly happy at one moment in time even though she is not perfectly happy at some other time. At any time she is either perfectly happy or not and at no time is she both perfectly happy and not perfectly happy.

Q2. Has anyone else defended a similar theory? How is the fulfillment theory different from these theories?

A. Chekola (1974, 2007) and Solomon (1976), as we have seen, defend theories that are close to the theory introduced in this paper, but they do not argue for an unrestricted form. Kant, in some places, appears to express a view of happiness that is very close to the one defended in this paper (see, for example, Kant (1785/2002, 15; Ak 4:399), Kant (1788/1997, 104; 5:124) and Wike (1994)). But it is difficult to interpret the Prussian philosopher and perhaps his version is more similar to an eternal form.

The theory in this paper is unrestricted. Few desire-satisfaction theories of happiness are. This is reasonable since we are speaking of *perfect* happiness. It is possible to be happy without being *perfectly* happy, but someone is *perfectly* happy only if *all* her wants are fulfilled. In this sense, the fulfillment theory is more ambitious than many other similar theories. It is very difficult to be perfectly happy (see **Q8** below). Most desire-satisfaction theories in the literature have little to say about what it means for a desire to be satisfied. As far as I know, the analysis in this paper has not been defended by anyone else in the literature, at least not explicitly (see **Q7** below). Yet, it seems to be intuitively very plausible.

Q3. To what category does happiness belong?

A. Most people seem to assume, often without much discussion, that happiness is some kind of property. This is natural, since ‘happy’ is often used as an ordinary predicate. We say, for example, that Susan is happy or that Jones is happy, or that someone has the property of being happy, and so on. However, sometimes philosophers and scientists have located happiness in some other category. Diener and Biswas-Diener (2008) appear to think of happiness as a process. Zamagni (2005) speaks about happiness as an interpersonal relation. Sentences such as

²¹ However, to be able to respond to argument five in Section 5, this proposition must be qualified somewhat. The quantifier in the definition of perfect happiness is a propositional or sentential quantifier. When this quantifier is instantiated, the instances will be “quantifier-free,” that is, free from propositional quantifiers.

“Peter is happy with his wife” and “Jenny is happy with her teacher” suggest that this might be a fruitful way of looking at happiness. According to some life-satisfaction theories, happiness might be a relation between a person and that person’s life. A person stands in this relation to her life iff she is satisfied or happy “with her life.”²² According to some views, happiness might be a relation between a person and a state of affairs (or a proposition). We say such things as, “He is happy that he has a job,” “She is happy that her children are healthy,” etc. In these cases, happiness appears to be a relation between a person and a state of affairs. Both Aristotle (*Nicomachean Ethics*) and Russell (2012) seem to think that happiness (eudaimonia) is an activity. According to the fulfillment theory, happiness is a property. An individual *x* has this property iff all *x*’s wants are satisfied.²³

Q4. What does ‘necessarily’ mean?

A. The theory is a definition of perfect happiness; it is supposed to tell us something about the essence of perfect happiness. The necessity is, therefore, an absolute necessity. The equivalence is supposed to be true in every possible world at every point in time.

Q5. What kinds of things are happy? What does ‘every individual’ mean?

A. We speak of many things as being happy: a happy person, a happy individual, a happy dog, a happy life, a happy marriage, a happy day, month or year, a happy moment, a happy time, a happy feeling, and so on. In this paper, I am primarily interested in happiness as a property of individuals. The expression ‘every individual’ could mean every individual whatsoever of any kind or it could range over some subset of individuals. If we assume that ‘every individual’ ranges over absolutely everything, then everything that does not want anything will be perfectly happy. This view may shed some light on certain eastern philosophies and religions, such as Buddhism and Taoism, and on asceticism as an ideal. If we could get rid of all our wants (desires, inclinations), we would become perfectly happy according to the fulfillment theory. Nevertheless, as long as we are alive, it is probably impossible not to want anything at all. Furthermore, if dead people do not want anything (and dead people can have properties), everyone who is dead will be perfectly happy. This might seem to be a comforting view. If we will be perfectly happy being dead, why fear death? Yet, some might think that these consequences are counterintuitive. It seems somewhat strange to call such things

²² Note that I do not want to suggest that everyone who says that happiness is ‘life-satisfaction’ thinks that happiness is a relation between a person and his life. Such a philosopher might, for example, think that happiness is a monadic property, but that a person has that property iff she is satisfied with her life as a whole.

²³ Elsewhere, I speak about individuals as perfectly happy at particular moments in time (see, for example, **Q1** above) or as perfectly happy in a possible world or in a possible world at a particular time (see, for example, **Q11** below). So, it is possible to think of perfect happiness as a relation that involves time (and/or possible worlds). However, in this sense, all ordinary properties can be interpreted as relations. Therefore, I shall continue to speak about perfect happiness as a property.

as stones, raindrops and carbon atoms ‘perfectly happy.’ Still, if such entities do not want anything, it follows that they are perfectly happy according to the theory. If we want to avoid these consequences, we can restrict the expression ‘every individual’ to every individual who wants something (or are able to want something). This will include all (or most) humans and many animals; it might include aliens and supernatural beings (if there are any), and in the future perhaps various artificial agents. Things that do not (are not able to) want anything will then not count as perfectly happy.

Q6. What does ‘wants’ mean?

A. A want is a kind of attitude, often called a ‘propositional attitude,’ since it is supposed to have a proposition or state of affairs as its content or object. Let us abbreviate the expression “Individual *c* wants it to be the case that *A*” in the following way: *WcA*. If *c* wants it to be the case that *A*, we can say that *c* has the property of wanting it to be the case that *A*. Here are some other words that are often used as synonyms for ‘want’: ‘desire,’ ‘inclination,’ ‘urge,’ ‘propensity,’ ‘wish,’ ‘love.’ I will sometimes use ‘desire’ as an alternative to ‘want’ for linguistic variation. An individual *c* has a desire or want for *A* iff *c* wants *A* to be the case.

In this paper, wanting something means wanting it all things considered. It is possible to want something without feeling like doing it. One can, for example, want to go to the dentist without being particularly keen on doing it. One wants to go to the dentist because going to the dentist is a (necessary) means to having healthy teeth and avoiding toothache. So, it is possible to want something as a means to something else and it is possible to want something in itself.

According to the fulfillment theory of happiness, it is possible to want ‘anything.’ ‘*A*’ in the expression “*WcA*” can be replaced by any well-formed sentence whatsoever. ‘*A*’ can be about the present time (I want to talk to you now)²⁴, about the future (She wants to write a book [sometime in the future]) or about the past (I want [hope, desire, wish] that I made the right choice yesterday (Feldman, 2004, 2)); it can be about a contingent state of affairs (She wants to go to Europe) or a necessary state of affairs (He wants the Goldbach conjecture to be true); it can be about facts concerning nature (He wants the sun to shine tomorrow) or about various mental states (She wants to feel the pleasure of eating an apple pie); it can be about *c* (He wants to be perfectly happy) or about some other individual or individuals (She wants her daughter to be happy); it can be about things within *c*’s control (He wants to climb Mount Everest) or about things that are not within *c*’s control (She wants there to be peace in the Middle East); it can be about something *c* wants to *do* (She wants to play tennis) or about something *c* wants to *be* (He wants to be a member of the group); and so on. It is even possible to want impossible states of affairs. Someone can both love and hate

²⁴ Let ‘*A*’ stand for “I talk to you now.” Then “I want to talk to you now” can be symbolized in the following way: *WiA* (read: “I want it to be the case that I talk to you now”), etc.

something at the same time according to the theory (He wants to be married to her and he wants not to be married to her.).²⁵

Q7. What does it mean to say that a want is fulfilled?

A. Here is a first preliminary answer. If individual *c* wants it to be the case that *A*, then *c*'s want is fulfilled (or satisfied) iff *A*. In other words, if *c* wants it to be the case that *A* and *A* is in fact the case, then *c*'s want is fulfilled, and if *c* wants it to be the case that *A* and *A* is in fact not the case, then *c*'s want is not fulfilled. If *c*'s want is not fulfilled, we can say that it is frustrated. However, things get more complicated when we consider the fact that the content of a want can be about the future, for example as in the following scenario: on Monday (t_1), I want you to meet me here on Friday (t_3). On Monday and on Wednesday (t_2) it is not the case that you meet me here. Suppose that you in fact meet me here at t_3 . Then, my want seems to be fulfilled. But *when* is it fulfilled? At t_1 or at t_3 ? And is the want *not* fulfilled and thus frustrated at t_1 and at t_2 ? Suppose, instead, that you do not meet me here at t_3 . Then, my want seems to be frustrated. But *when* is it frustrated? At t_1 , t_2 or t_3 ?

According to the fulfillment theory in this paper we shall use the following terminology, which I think makes the fulfillment theory unique and quite different from other desire-satisfaction theories in the literature. In the possible world(s) (if there are any) where you do meet me here at t_3 , my want (at t_1) is fulfilled at t_1 (not at t_3), and in the possible world(s) (if there are any) where you do not meet me here at t_3 , my want (at t_1) is not fulfilled, and hence frustrated, at t_1 (not at t_3). It is not until t_3 we can know for sure whether or not my want is satisfied at t_1 . If you do in fact meet me here, my want is satisfied at t_1 , and if you do not in fact meet me here, my want is frustrated. We do not say that my want is frustrated at t_1 and we do not say that it is frustrated at t_2 due to the fact that you do not meet me here at either t_1 or t_2 , because what I want at t_1 is not that you meet me here at t_1 and it is not that you meet me here at t_2 , it is that you meet me here at t_3 . It is not until t_3 (or until it is settled that it will be the case that you will not meet me here at t_3) that it is settled that my want is frustrated at t_1 . All of this is compatible with the idea that I might change my mind. At t_3 , I might no longer want you to meet me here now (at t_3). In fact, suppose that it is true at t_3 that I want it to be the case that you do not meet me here now (at t_3). Furthermore, suppose that you do in fact meet me here at t_3 . Then, my want at t_3 is frustrated, even though my want at t_1 is fulfilled. Suppose, instead, that you do not meet me here at t_3 . Then, my want at t_3 is fulfilled, but my want at t_1 is frustrated. These clarifications will become important in Section 5.

Q8. Is it possible to be perfectly happy in this life? Are there any individuals who are perfectly happy?

²⁵ The theory of wants that is used in the proofs in Section 4 and throughout the present article is developed in more detail in the paper Rønnedal (2020). See also Rønnedal (2019b) and Rønnedal (2019c).

A. Some people seem to think that it is impossible to be perfectly happy in this life. According to St Thomas Aquinas, for example, a certain participation in happiness is possible but perfect and true happiness cannot be had in this life; we have to wait for the afterlife to experience that kind of happiness (*Summa Theologiae*, First Part of the Second Part, Question 5, Article 3). Of course, Thomas does not use the expression 'perfect happiness' in the same sense as in this paper. Yet, he seems to think that perfect happiness entails perfect fulfillment. So, he would perhaps also reject the idea that we can be perfectly fulfilled in this life. According to the fulfillment theory, however, it is in principle possible to be perfectly happy in this life. It is possible that all a person's wants are fulfilled. Nevertheless, it seems to be very difficult for most humans. Perhaps no living person has ever been perfectly happy and perhaps no living person will ever be. Although it is not absolutely or logically impossible to be perfectly fulfilled, it might be historically impossible for some individuals.²⁶ We cannot know a priori whether there has ever been a living individual who was perfectly happy.

Q9. What does it mean to say that the concept of perfect happiness is a purely formal concept?

A. The fact that the concept of perfect happiness is a purely formal concept means that there are no restrictions on the contents of the wants that must be fulfilled for an individual to be perfectly happy according to the theory (however, see footnote 21). The concept itself has no matter. Almost any kind of individual living almost any kind of life can in principle be perfectly happy. Someone living a quiet life in solitude can be perfectly happy. Someone living an active, hectic, social life can be perfectly happy. Ascetics and hedonists, introverts and extroverts, married and unmarried people, active and contemplative individuals can all be perfectly happy. It does not matter if you are male or female, young or old, rich or poor. You can be perfectly happy no matter what your social class, ethnicity or sexual orientation is. It is possible that there are perfectly happy bus drivers, businessmen, nurses and philosophy teachers. Perfect happiness is, in principle, compatible with almost any kind of job. Even slaves and wicked villains can, in principle, be perfectly happy according to the theory. However, it will be more difficult in some positions than in others, given the way we are constituted physically and mentally. For example, most people do in fact want certain things. We want to drink when we are thirsty, eat when we are hungry, sleep when we are tired; we want to feel secure and live in peace with our neighbors; we want to be with other people when we are lonely and have at least some good friends; we want to be free and healthy; we want to have a meaningful job and develop our talents; we want to feel pleasure (at least sometimes) and we do not want to feel pain; and so on. If we want those things and do not get them, we will not be perfectly happy. A slave who wants to be free, for example, will not be perfectly happy. It will be very difficult for someone who suffers from chronic pain to be

²⁶ For more on the concepts of historical possibility, impossibility and necessity, see Section 4.

completely fulfilled. And, even though it is not *logically* impossible, it is probably *historically* impossible for all (or at least most) people who live as wicked villains to be perfectly happy given the way we are constituted and what it means to live a life of this kind.

Q10. Is happiness good? Is it good in itself? Does it have intrinsic value?

A. Nothing in the fulfillment theory in itself entails that the property of perfect happiness is good, good in itself or intrinsically good. Nor is it necessarily the case that it is good that someone is happy. As we have seen, it is possible that even wicked villains are perfectly happy. It is doubtful that it is good that such people are completely fulfilled; it is probably bad (at least 'all things considered' and in most cases). But perfect happiness might have positive value given certain conditions. Whether or not it has will depend on what value theory is correct. It would take me too far from the main topic to pursue this question in the present paper.²⁷

Q11. How is the fulfillment theory of happiness related to other non-desire-satisfaction theories of happiness?

A. In many respects the fulfillment theory is radically different from other popular views of happiness. Perfect happiness is not pleasure. It is not life-satisfaction. It is not a harmonious life, it is not an activity (according to virtue). It is not an emotion or a mood or a feeling or a sensation, or a disposition to feel certain things or be in certain moods. It is not a mental state, or a type of mental state. It is not a property of a mental state or of a type of mental state. Happiness is not in the head. It is not in the body. It is not the same thing as well-being or virtue. One can be happy without faring well and without being morally good, and one can fare well and be morally good without being happy. A happy person is not necessarily functioning well or developing her talents or (human) dispositions.²⁸

Even though happiness is not essentially connected to pleasure, life-satisfaction and similar phenomena, according to the fulfillment theory, it is likely that many individuals will not be perfectly happy if they experience a lot of pain, are dissatisfied with their lives, and so on. For many people probably do in fact want to be satisfied with their lives and do not want to feel pain, they do want to be good persons and develop their talents and dispositions, they do want to be in

²⁷ The theory of happiness that is developed in this paper is part of a larger project where I try to construct a formal ethical system in a Kantian spirit. According to this system, everyone who is perfectly virtuous (and hence deserves to be happy) ought to be perfectly happy. So, the concept of perfect happiness can play an important role in ethics. However, this is not the place to defend this view. For more on this, see Rønnedal, forthcoming.

²⁸ Of course, it is possible to combine the fulfillment theory of happiness with a happiness theory of well-being. Then, one could argue that an individual *x* has a life that is perfectly good for *x* iff all *x*'s wants are fulfilled. But to show this would require some extra arguments and I am not sure that this theory of well-being is correct. In any case, it is likely that there is a positive correlation between fulfillment and well-being. For more on happiness theories of the good life, see, for example, Brülde (2007).

a 'happy' mood, and so on. If those wants are not fulfilled, they will not be perfectly happy.

The matter of x's happiness may consist in pleasure, the matter of y's happiness in satisfaction with life, the matter of z's happiness in pleasure, satisfaction with life, virtue and friends, and so on, because x, y and z want different things. Even though this is possible, it does not follow that the *nature* of x's happiness is something other than the *nature* of y's happiness, and so on, and it does not follow that 'happiness' is ambiguous.²⁹ The nature of happiness is still the fulfillment of wants. Other theories of happiness are interesting for the fulfillment theory of perfect happiness because they tell us something important about what kinds of things many people in fact do want. If a person in fact wants those things, she will not be perfectly happy without them.

Q12. Is there a maximum degree of happiness?

A. It is plausible to claim that there is in principle a maximum degree of happiness according to the fulfillment theory; one cannot be more happy than perfectly happy. If someone is perfectly happy, there is *absolutely nothing* that she wants that is not the case. In this sense, she lacks *absolutely nothing* and is perfectly 'self-sufficient.' *Absolutely all* her desires are fulfilled, no matter how trivial or insignificant they might seem. Getting more money will not make her happier since she does not want more money. Being more famous will not make her happier since she does not want to be more famous. Having more power will not make her happier because she does not want more power. Not even feeling more pleasure, less pain or being more healthy will make her happier since she does not want to feel more pleasure, less pain or be more healthy. Therefore, it is reasonable to assert that if someone is perfectly happy, she cannot be happier.³⁰ According to some other theories, there is no maximum: no matter how happy someone is, he could conceivably be happier (Davis (1981b)).

4. Arguments for the fulfillment theory and some theorems

There are many possible arguments for the fulfillment theory of happiness. It is an intuitively plausible, simple and elegant theory. The version developed in this paper is more precise than many similar theories. Therefore, it is easier to decide what follows and what does not follow from it. Some theories of happiness are so vague that they are almost unfalsifiable in principle. The fulfillment theory is the development of an idea that seems to have been around for more than two thousand years and which has been attractive to many thinkers from various backgrounds. Some kind of desire-satisfaction theory appears to have already been considered by the ancient Greeks – see, for example, Plato's *Gorgias* 491e–

²⁹ Even though it does not follow from our theory that 'happiness' is ambiguous, I am inclined to believe that this word can be used in several different senses (see Section 2).

³⁰ However, note that this conclusion does not strictly follow from the fulfillment theory in itself as it has been defined in this paper. For more on this, see Section 5.

494c. According to the medieval theologian and philosopher St Augustine, “he alone is blessed [happy] who has all that he wills, and wills nothing wrongly” (*The Trinity*, Book XIII, Chapter 5). Augustine is approvingly quoted by the scholastic thinker Thomas: “Augustine says (De Trin. Xiii, 5) that ‘happy is he who has whatever he desires, and desires nothing amiss’” (*Summa Theologiae*, First Part of the Second Part, Question 3, Article 4). The enlightenment philosopher Kant appears to defend a theory that is very similar to the one in this paper. According to him, “... all human beings always have of themselves the most powerful and inward inclination to happiness, because precisely in this idea all inclinations are united in a sum” (Kant (1785/2002, 15; Ak 4:399)). In *Critique of Practical Reason*, he expresses what is basically the same idea: “Happiness is the state of a rational being in the world in the whole of whose existence *everything goes according to his wish and will*” (Kant (1788/1997, 104; 5:124)). *The Metaphysics of Morals* contains a similar characterization: “That *everything* should *always* go the way you would like it to. ...What is such a condition called?... It is called *happiness*” (Kant (1797/2017), 6:480). In this paper, I will concentrate on one kind of argument, the fruitfulness of the theory. I will show that we can use the theory to prove several interesting theorems that follow from it. Since we have used a very precise definition of ‘perfect happiness,’ all arguments are without doubt deductively valid. This means that the conclusions must be true if the premises are true. In so far as the theorems are intuitively reasonable, we can also read the arguments in the other direction, that is, as abductive arguments in support of the fulfillment theory.

The idea that happiness (eudaimonia), well-being or blessedness is something everyone, or at least everyone who is rational, wants, is old. It has, for example, been expressed by Seneca the Younger. “To live happily... is the desire of all people” says the Stoic philosopher (*De Vita Beata (On the Happy Life)*, 99). According to Augustine, the roman philosopher Cicero, asserted that “[a]ll of us certainly will to be blessed [happy]” (*Hortensius*³¹). Augustine himself defends this proposition. According to him, “... [a]ll of you wish to be blessed [happy]; you do not wish to be miserable... whatever else it is that anyone secretly wills, he does not withdraw from this will which is sufficiently known to all and is in all men” (*The Trinity*, Book XIII, Chapter 3; see also Book XIII, Chapter 7). Thomas expresses a closely related idea in *Summa Theologiae*, First Part of the Second Part, Question 3, Article 1: “Happiness is the last end, to which man’s will tends naturally.” Kant, as we have seen, defends a similar claim. According to him, “... all human beings always have of themselves the most powerful and inward inclination to happiness” (Kant (1785/2002, 15; Ak 4:399)), and “There is *one* end... that one can presuppose as actual for all rational beings... and thus one aim that they not merely *can* have, but of which one can safely presuppose that without exception they *do have* it in accordance with a natural necessity, and that

³¹ Only fragments of this work are still available.

is the aim at *happiness*" (Kant (1785/2002, 32; Ak 4:415)). In *Critique of Practical Reason* he expresses the same view: "To be happy is necessarily the demand of every rational but finite being..." (Kant (1788/1997, 23; 5:25)). If the fulfillment theory is true, we can prove that this intuition, in a certain sense, is correct. Before we establish this, we will consider how we can make this intuition more precise.

If someone is not perfectly rational, almost nothing of interest follows from the fact that she wants something. So, we cannot prove that anyone whatsoever wants to be happy. If this proposition is true, it is not logically true. If someone is perfectly rational, however, we will assume that 'wanting' functions as a kind of modal operator in normal modal logic.³² We will say that it is true that a perfectly rational individual, *c*, wants something, *A*, in a possible world, *w*, iff *A* is true in every possible world, *w'*, that is acceptable to *c* in *w*. We will also assume that if a possible world *w'* is acceptable to *c* in *w*, then *w'* is acceptable to *c* in *w'*. Furthermore, we will, as is standard, assume that it is true that it is (historically) necessary that *A* in a possible world, *w*, iff *A* is true in every possible world that is alethically accessible from *w*, and that the alethic accessibility relation is an equivalence relation. In addition, we shall assume that if the possible world *w'* is acceptable to individual *c* in the possible world *w*, then *w'* is alethically accessible from *w*, and if *c* wants it to be the case that *A* in a possible world, *w*, then *c* wants it to be the case that *A* in every possible world that is alethically accessible from *w*. Given these assumptions, which are plausible, we can now prove our first theorem **T1**.³³

T1. *It is necessary that every perfectly rational individual wants to be perfectly happy.*

Proof. Suppose that **T1** is not valid. Then there is some possible world, *w*₁, in which there is some perfectly rational individual, *c*, that does not want to be perfectly happy. Hence, *c* is perfectly rational in *w*₁ and it is false that *c* wants to be perfectly happy in *w*₁. It follows that there is a possible world, *w*₂, that is acceptable to *c* in

³² For some introductions to modal logic, see, for example, Blackburn, de Rijke and Venema (2001), Chellas (1980), Garson (2006) and Hughes and Cresswell (1968).

³³ I cannot defend all these assumptions in the present paper. For more details about the background theory, see Rønnefeldt (2020), (2019b) and (2019c). In a more developed theory, the accessibility relations can be 'relativized' to time. Intuitively, *A* is historically possible in a possible world *w* at a certain moment in time *t* iff *A* is still possible at *t* given the history of *w* and the laws of nature that hold in *w*, and it is historically necessary that *A* in *w* at *t* iff *A* is true at *t* in every possible world that is still possible at *t* given the history of *w* and the laws of nature that hold in *w*. In the present paper, when we say that *w'* is alethically accessible from *w*, we mean that *w'* is alethically accessible from *w* at a particular moment in time (and similarly for the acceptability relation). However, for our current purposes, we do not need to introduce any moments in time in our models. So, this element is suppressed in the present paper. Intuitively, *w'* is alethically accessible from *w* at *t* iff *w'* is still possible given the history of *w* and the laws of nature that hold in *w* (in a tree-like structure *w* and *w'* have not yet branched off at *t*). Furthermore, in this paper we assume that if *x* is perfectly rational, it is necessary that *x* is perfectly rational (this assumption is not necessary to prove all theorems).

w_1 in which c is not perfectly happy. Since c is not perfectly happy in w_2 , it is not the case that everything c wants in w_2 is true. Accordingly, there is something, X , that c wants in w_2 that is not the case. Consequently, it is true in w_2 that c wants it to be the case that X and it is false in w_2 that X . The world w_2 is acceptable to c in w_2 [by assumption and the fact that w_2 is acceptable to c in w_1]. It follows that X is true in w_2 , for in w_2 c wants it to be the case that X . Yet, this is absurd. Q.E.D.

In conclusion, the fulfillment theory of happiness does not entail that *everyone* wants to be happy, but it does follow from the theory that *everyone who is perfectly rational* wants to be perfectly happy (given our assumptions). Accordingly, if someone does not want to be perfectly happy, she is not perfectly rational. We now turn to the next theorem.

Let us say that a theory of happiness is a 'harmony theory' just in case it identifies happiness with harmony, consistency, unity, (mental) health or some similar property. According to a theory of this kind, happiness is coherence, inner peace, integration, tranquillity, harmony, psychological freedom, consistency, unity. Someone is happy iff she has a healthy, well-ordered, well-structured mind, a mind that is integrated and at peace with itself. A happy person is a whole person, a complete person, and a happy soul is a soul where every part of the soul is in harmony with every part of the soul and with the whole soul. The opposite of happiness, according to a theory of this kind, is inconsistency, strife, disunity, incoherence, sickness, inner war. An unhappy soul is a soul that is at war with itself; it is a disintegrated soul, a soul without unity; it is a soul where different parts pull in different directions. It is unclear whether anyone has defended a theory of this kind. Plato occasionally appears to come close to arguing for some kind of harmony theory (see, for example, *Republic*). According to the fulfillment theory, happiness is not *identical* to consistency, etc. However, we will see that consistency, in a broad sense, is a *necessary* but not sufficient condition for perfect happiness. In other words, it is possible to be consistent without being perfectly happy, but it is not possible to be perfectly happy without being consistent. So, even though happiness is not the same thing as harmony, it is closely related to such properties as coherence, integration and peace of mind.

Consider the following definition:

D4. Individual c 's will is free from contradictions iff it is not the case that there is something, X , such that c wants it to be the case that X and c wants it to be the case that not- X . If there is something, X , such that c wants it to be the case that X and c wants it to be the case that not- X , then c 's will is contradictory, and vice versa.

We are now in a position to state our next theorem **T2**.

T2. *It is necessary that someone is perfectly happy only if her will is free from contradictions.*

Proof. Straightforward.

From theorem **T2** we can immediately derive some corollaries, for example, it is necessary that if someone's will is not free from contradictions, she is not

perfectly happy and it is impossible that someone with a contradictory will is perfectly happy.

As usual, we shall assume that it is true that it is (historically) possible that A in a possible world, w , iff A is true in at least one possible world that is alethically accessible from w . It is now easy to establish theorem **T3**.

T3. *It is necessary that someone is perfectly happy only if everything she wants is possible.*

Proof. Left to the reader.

Consider the following definition:

D5. An individual c 's will is free from dilemmas iff there is no A and B such that it is impossible that A-and-B and c wants A to be the case and c wants B to be the case.

Given this definition, we can prove theorem **T4**.

T4. *It is necessary that someone is perfectly happy only if her will is free from dilemmas.*

Proof. Suppose that **T4** is not valid. Then there is a possible world, w_1 , where someone, c , is perfectly happy and in which there is an A and a B such that it is impossible that A-and-B and c wants it to be the case that A and c wants it to be the case that B. Accordingly, it is impossible that X-and-Y in w_1 and c wants it to be the case that X in w_1 and c wants it to be the case that Y in w_1 . Since it is impossible that X-and-Y in w_1 and w_1 is alethically accessible from itself, X-and-Y is false in w_1 . Since c is perfectly happy in w_1 , everything c wants in w_1 is true in w_1 . So, if c wants it to be the case that X in w_1 , then X is true in w_1 ; and if c wants it to be the case that Y in w_1 , then Y is true in w_1 . Consequently, X is true in w_1 and Y is true in w_1 . Hence, X-and-Y is true in w_1 . But this is absurd. Q.E.D.

Consistency, in a broad sense, seems to be the very essence of rationality. If you want to be rational, you should try to be consistent; you should try to avoid not only contradictory beliefs but also contradictions of the will. But why should one be rational and consistent? Our theorems above provide us with one very interesting reason. If you are perfectly rational, then you want to be perfectly happy. And you cannot be perfectly happy if you are not consistent (in a broad sense). Hence, if you are perfectly rational you want to be consistent. In other words, being consistent is a necessary condition for perfect happiness. If you are not consistent, you cannot be perfectly happy. In fact, it is plausible to assume that no perfectly rational individual will have inconsistent desires.

Before we establish our next theorem, we will introduce a definition and prove a lemma. We shall say that B is a necessary means to A iff it is historically necessary that A implies B. The so-called "hypothetical imperative" is a principle that is defended by many philosophers, including Kant (Kant (1785/2002, 34; Ak 4:417)). There are several possible interpretations of this principle. According to the reading we will use in the proof of **T5** below, it is necessary that if a perfectly rational individual, x , wants it to be the case that A, and B is a necessary means to A, then x also wants it to be the case that B. We can now prove our lemma.

L1. *The hypothetical imperative is valid. In other words, if x is perfectly rational, then if x wants it to be the case that A , and B is a necessary means to A , then x wants it to be the case that B .*

Proof. Suppose that **L1** is not valid. Then there is some possible world, w_1 , where some perfectly rational individual, c , wants it to be the case that something, X , is the case, and where something, Y , is necessarily implied by X at the same time that c does not want Y to be the case. Consequently, there is a possible world, w_2 , that is acceptable to c in w_1 in which Y is false. Hence, X is true in w_2 . Since w_2 is acceptable to c in w_1 , w_2 is alethically accessible from w_1 [by assumption]. It follows that it is true that X implies Y in w_2 . Therefore, Y is true in w_2 . But this is absurd. Q.E.D.

Now, let us spell out the details of the argument for theorem **T5**.

T5. *It is necessary that everyone who is perfectly rational wants to have a will that is free from contradictions.*

Proof. It is necessary that every perfectly rational individual wants to be perfectly happy (**T1**). It is necessary that someone is perfectly happy only if her will is free from contradictions (**T2**). Given **T1** and **T2**, it is easy to derive **T5** by using the hypothetical imperative. For the following proposition is an instance of this principle: it is necessary that if a perfectly rational individual, c , wants to be perfectly happy and having a will free from contradictions is a necessary means to perfect happiness, then c also wants to have a will free from contradictions. Consequently, **T5** is valid. Q.E.D.

We can also prove several similar theorems, for example, that it is necessary that everyone who is perfectly rational wants it to be the case that everything she wants is possible and that it is necessary that everyone who is perfectly rational wants to have a will that is free from dilemmas.

Let us now turn to our last two theorems, which are perhaps the most interesting. The idea that happiness is the final, complete, last, supreme, or highest end in some sense is an old idea. Aristotle was perhaps the first philosopher to clearly express this view. According to him, "it is for the sake of [happiness] that we all do everything else [we do]." (*Nicomachean Ethics* 1.12, 1102a2-3.) Thomas, as we have seen, expresses a closely related idea in *Summa Theologiae*, First Part of the Second Part, Question 3, Article 1: "Happiness is the last end, to which man's will tends naturally." If the fulfillment theory of happiness is true, we can show that this intuition, in a certain sense, is true. To establish this result, we must first define what we mean by "a final, complete, last, supreme, or highest end." This concept can be defined in many different ways. In this paper, I will use the following definition:

D6. Something A is a final end for the individual x iff x wants A and everything x wants is a necessary means to A .

Given this definition, we can prove our last two theorems (**T6** and **T7**).

T6. *For every individual x , there are no two distinct (not necessarily equivalent) final ends for x . In other words, if A is a final end for x and B is a final end for x , then A and B are necessarily equivalent.*

Proof. Left to the reader.

T7. *Perfect happiness is a final, complete, last, supreme, highest end for every perfectly rational individual. In other words, it is necessary that if x is perfectly rational then x wants to be perfectly happy and everything x wants is a necessary means to x 's perfect happiness.*

Proof. Suppose **T7** is not valid. Then there is a possible world, w_1 , in which there is a perfectly rational individual, c , that does not want to be perfectly happy or else it is not true in w_1 that everything c wants is a necessary means to c 's perfect happiness. But we have already shown that it is necessary that every perfectly rational individual wants to be perfectly happy (**T1**). Hence, c wants to be perfectly happy in w_1 . Accordingly, it is not true in w_1 that everything c wants is a necessary means to c 's perfect happiness. Consequently, c wants something, X , in w_1 that is not a necessary means to c 's perfect happiness in w_1 . It follows that there is a possible world, w_2 , that is alethically accessible from w_1 in which c is perfectly happy and X is false. Since c is perfectly happy in w_2 , everything c wants is in fact the case in w_2 . So, it is true in w_2 that if c wants it to be the case that X , then X . Since X is false in w_2 , it follows that it is false that c wants it to be the case that X in w_2 . Therefore, X is false in some possible world, say w_3 , that c accepts in w_2 . By assumption c wants it to be the case that X in w_2 , for c wants it to be the case that X in w_1 and w_2 is alethically accessible from w_1 . It follows that X is true in w_3 . Yet, this is absurd. Q.E.D.

Philosophers have for a long time thought that there is something special about happiness. Happiness is not just an end among other ends, like power, money or fame; it is a higher order end or an all-inclusive end, an end that includes all other ends. If happiness is perfect fulfillment, we can show that this is the case (for all rational individuals). This fact makes the fulfillment theory of happiness very attractive.³⁴

³⁴ Before I end this section, I would like to discuss one possible argument against **T1**. Suppose c is perfectly rational. Then c wants to be perfectly happy (from **T1**). Imagine that c has several wants (or desires), W_1 , W_2 and W_3 , and so on, that take a lot of efforts to satisfy, including a desire to write a book on happiness. Furthermore, suppose that a neuroscientist offers c some surgery that will rid c of all c 's desires, except some very basic desires (like the desires to eat and sleep), which will allow c to become perfectly happy after the surgery. Since c wants to be perfectly happy, c will accept this offer. But this is absurd. It is clear that rationality does not compel c to accept the scientist's offer. For accepting the scientist's offer will prevent c from ever fulfilling his actual wants (W_1 , W_2 , W_3 , etc., including the desire to write a book on happiness). Hence, **T1** is false.

However, this is not a serious problem for the theory in this paper. For the theory does not entail that a perfectly rational individual would accept the scientist's offer. Suppose that it is necessary that if the scientist performs the surgery, then c will not write a book on happiness (which c wants to do). Assume that c wants to go through with the surgery (for reductio). Then c wants

5. Arguments against the fulfillment theory

There are many potential arguments against the fulfillment theory of happiness. It is beyond the scope of this paper to discuss all possible problems. However, as far as I know there are no conclusive arguments against the theory. Still, let me briefly address some possible objections.³⁵

Firstly, we should not call the theory in this paper a theory of (perfect) happiness because happiness is not perfect fulfillment. Fulfillment does not 'deserve' the label 'happiness,' for happiness is clearly something purely mental. However, this objection seems to beg the question. If the fulfillment theory is correct, then happiness is not something purely mental. The following argument suggests that it is reasonable to call perfect fulfillment 'perfect happiness.' Happiness is the final end. It is that 'thing' for the sake of which we want everything we want. This idea goes back at least to Aristotle (see Section 4 above). But we have proved that perfect fulfillment is a final end for every perfectly rational individual (T7). In the light of theorem T6, we can see that perfect fulfillment is not just *a* final end, but *the* final end for every perfectly rational individual. Therefore, (perfect) happiness is (perfect) fulfillment. This does not entail that we cannot use 'happiness' in other senses too, for example for some kind on mental state or property of a mental state. Furthermore, I am not alone in using the term in this sense. Many other philosophers (see Q1 above) have used 'happiness' in a similar sense throughout history.³⁶

Secondly, it is unreasonable to claim that every individual that does not want anything is perfectly happy. But we have already seen how we can respond to this argument (see Q5 above). If we want to avoid this conclusion, we can restrict our theory to things that have (or can have) desires. Then things that do not (are not able to) want anything are not perfectly happy.

Thirdly, it is unclear why we should focus on 'perfect happiness.' Since nobody (or almost nobody) is perfectly happy, those who wish to use happiness as a measure of the success of a policy or a society will have little use for the notion of perfect happiness. This is perhaps true, but happiness is not only interesting as a measure of the success of a policy or a society. The notion of happiness can have

it to be the case that he will not write a book on happiness [from our assumptions and L1]. Hence, c wants to write a book on happiness at the same that he wants it to be the case that he will not write a book on happiness. But it is reasonable to assume that no perfectly rational individual has inconsistent desires of this kind. Hence, our assumption is false. It is not the case that c wants to go through with the surgery (even though c wants to be perfectly happy).

³⁵ The first four arguments in this section are arguments that colleagues have raised when they have been confronted by the ideas in this paper, the last two (or versions of the last two) are arguments that can be found in the literature.

³⁶ We do not have to be dogmatic about this. If some reader insists on using 'happiness' in some other sense, we can replace all talk about 'perfect happiness' in this paper with 'perfect fulfillment.' The important thing is that (perfect) fulfillment plays many of the roles that traditionally have been ascribed to (perfect) happiness, eudaimonia or blessedness.

many functions. For example, we want to know if there is a final end and what this final end is, and we have seen that perfect happiness is a final end for every perfectly rational individual (see **T7** above). We have shown that all perfectly rational individuals want to be perfectly happy (see **T1** above). Hence, the notion can play an important role in a theory of rationality. The concept of perfect happiness can also play an important role in various ethical theories, for example as an element in the highest good (see footnote 27 and Wike (1994)). Again, this does not entail that we cannot use 'happiness' in other senses too, for example, to refer to someone's life-satisfaction. But this does not make the concept of perfect happiness useless.

Fourthly, it has been suggested to me that the fulfillment theory has counterintuitive consequences. Consider a very short life A, that contains just one easily satisfied desire, and a very long life B, that contains a great many satisfied desires and just one fairly trivial frustrated desire. The fulfillment theory entails that life A is happier than life B. But this is implausible and, hence, the theory must be false. This would perhaps be a serious problem for the theory if it were true that it entails that life A is happier than life B, but it does not. Perfect happiness is not a property of whole lives according to the theory. It is a property that an individual can have at a particular moment in time. The theory in itself does not even entail that it makes sense to say that one individual is happier than another. *Perfect happiness* is a property, *happier than* is a relation. It is perhaps independently plausible to claim that if individual c is perfectly happy at a particular moment in time t and individual d is not perfectly happy at t, then c is happier than d at t. Suppose that this is the case, that c is perfectly happy at t, that d is not perfectly happy at t and that d has many more desires that are fulfilled than c at t. Then it is the case that c is happier than d at t. Some might perhaps think that this conclusion is counterintuitive, but to me it seems reasonable. Individual d has more things than c at t but also wants more out of life than c, which is satisfied with what she has. Hence, it makes sense to say that c is happier than d at t.³⁷

Fifthly, the fulfillment theory might seem to be inconsistent. Bradley (2007) and Feldman (2004, 17, 2010, Sec 4.5) discuss a potential problem or paradox for "preferentism" that might also be a problem for the fulfillment theory of happiness. In the specific forms discussed in the literature, the paradox is not explicitly

³⁷ However, note that these conclusions do not follow from the fulfillment theory in itself. It is possible to deny them and still hold on to the theory. Note also that the theory does not entail that it is better to be c than d, that c's life is better than d's life or that we should choose to be c if we must choose to be c or d. The theory has no such evaluative or normative consequences. Even if it were possible to find some alternative definition of the relative happiness of whole lives such that life A is happier than life B, I am inclined to believe that this is not a conclusive argument against the fulfillment theory. Therefore, I will not consider any such definitions in this paper.

focused on the concept of perfect happiness, but it is easy to see the relevance of the problem to the theory presented in this paper.

Here is the puzzle. Suppose some person, *c*, is not perfectly rational. Suppose he has only one desire – the desire not to be perfectly happy. Then (according to the theory) if he is not perfectly happy, then his sole desire has been satisfied and so he is perfectly happy. But if he is perfectly happy, then his desire has been frustrated and he is not perfectly happy. Then his sole desire has been satisfied, and so on. The theory seems to imply that such a person, with the stipulated desire, would be perfectly happy if and only if he is not perfectly happy. Thus we seem to have a paradox that arises for the theory of perfect happiness defended in this paper. Let ‘*Pc*’ stand for “*c* is perfectly happy.” Then the argument can be symbolized in the following way:

- | | |
|-------------------------------------|---|
| 1. $Wc \neg Pc$ | [Assumption] |
| 2. Pc | [Assumption] |
| 3. $\forall A(WcA \rightarrow A)$ | [From 2 by the definition of perfect happiness] |
| 4. $Wc \neg Pc \rightarrow \neg Pc$ | [Instance of 3, $\neg Pc/A$] |
| 5. $\neg Pc$ | [1, 4, Modus Ponens] |
| 6. $Pc \rightarrow \neg Pc$ | [2–5, discharging the assumption] |
| 7. $\neg Pc$ | [Assumption] |
| 8. $Wc \neg Pc \rightarrow \neg Pc$ | [1, 7, Propositional logic] |
| 9. $\forall A(WcA \rightarrow A)$ | [8, <i>c</i> only has one desire] |
| 10. Pc | [From 9 by the definition of perfect happiness] |
| 11. $\neg Pc \rightarrow Pc$ | [7–10, discharging the assumption] |
| 12. $Pc \leftrightarrow \neg Pc$ | [6, 11, Propositional logic] |

This might seem to be a devastating argument against the fulfillment theory. But when we symbolize the derivation as above it is easy to see a serious problem with the argument. The universal quantifier in 3 is a propositional quantifier and in 4 we have instantiated *A* with $\neg Pc$. But $\neg Pc$ is an abbreviation of $\neg \forall A(WcA \rightarrow A)$ and this sentence includes a propositional quantifier. It is well-known that it is problematic to allow universally quantified sentences to be instantiated with universally quantified sentences when we use \forall -elimination for propositional quantifiers. To see one of the problems, let $A = \forall XX$ and assume that our substitution-instances can include any formula whatsoever. Then $A[A/X] = A$, where $A[B/X]$ is the result of replacing all free occurrences of the variable *X* in *A* by *B*, for $\forall XX[\forall XX/X] = \forall XX$. So, to know if $\forall XX$ is true or not we must first know the truth-value of $\forall XX$. This clearly seems to generate a vicious circle. To avoid this problem such instances are often prohibited (for more on some systems that solve this problem, see Rönndal (2019)). Hence, step 4 in the derivation above is

not permitted. But then the conclusion does not follow from our assumption. Consequently, we can avoid this 'paradox.'³⁸

The arguments I have discussed so far do not strike me as particularly strong. Nevertheless, the discussion has hopefully made the fulfillment theory somewhat clearer. I will now discuss one of the most serious arguments against the theory.

According to this argument, getting what you want will not make you happy, and hence there must be something wrong with the fulfillment theory of happiness. This argument has been nicely expressed by Kekes (1982):

Consider a man who has all he wants. ... we must concentrate on a man who has all he seriously wants. But suppose that he wants only one thing; he pursues it single-mindedly, to the exclusion of everything else, and while he gets it, his soul shrivels. Rich misers, successful avengers, triumphant climbers of greasy poles notoriously find themselves empty, once their obsessions are satisfied. Or a man may want only what he does not have, and when he gets it, like Don Juan, he no longer wants it. Yet others are mistaken in thinking that what they want will satisfy them. The glittering sophistication of an inner circle may pale once the outsider finds himself accepted. Having what one wants, therefore, is no guarantee of happiness. Nor should it be supposed that doing all that one really wants leads to happiness.

Does this argument refute the fulfillment theory? I will now try to argue that it does not. Consider the following scenario. At time t_1 , individual c wants to be a millionaire and this is the only thing c wants at t_1 . At time t_2 , c becomes a millionaire. But at t_2 this is not enough for c anymore. At t_2 , c wants to be a billionaire, or at t_2 , c 'realizes' that there is more to life than money. Now, at t_2 , c wants friends and family, for example, something c has 'neglected' up until t_2 . So, c is not perfectly happy at t_2 since everything c wants at t_2 is not the case. We can still say that c at t_1 is perfectly happy (given, the unrealistic assumption, that the only thing c wanted at t_1 was to be a millionaire). Again, consider the following scenario. At t_1 , c wants to become a member of an inner circle, and at t_2 , c finds himself accepted. But at t_2 (or soon after) c does not want to be a member any longer, and it is even the case that c wants to leave the circle. Then c is not perfectly happy at t_2 . The following scenario is similar. At t_1 , c wants to become famous, and at t_2 , her dream comes true. But c soon discovers all sorts of negative effects of being famous; it was nothing like she expected it to be. Being famous, c has little privacy and is constantly chased by paparazzi; people lie about her in the media;

³⁸ Someone who wants to use this argument against the fulfillment theory must show why we should accept step 4 in the derivation above and how we can avoid all the problems that follow if we allow universally quantified sentences to be instantiated with universally quantified sentences when we use \forall -elimination for propositional quantifiers. Note that this kind of response also takes care of other similar potential 'paradoxes.' For example, we do not have to assume that we must first (in some sense) 'know' that a perfectly rational individual is perfectly happy to decide that a perfectly rational individual is perfectly happy, even though every perfectly rational individual, according to our theory, wants to be perfectly happy.

she receives anonymous death threats, and is stalked by some fanatic fan; etc. As a consequence, *c* no longer wants to be famous. Then, *c* is not perfectly happy at t_2 . In these cases, we also assume that it is true at t_1 that *c* believes that being a member of the inner circle or being famous will make *c* *feel* satisfied and fulfilled (at t_2) and that *c* *wants* to feel satisfied and fulfilled. However, in these scenarios, *c* does *not* feel satisfied or fulfilled at t_2 . Hence, *c* is neither happy at t_1 nor at t_2 . So, these thought-experiments are no threat to the fulfillment theory. Other cases are similar, for example the scenarios that concern what someone wants to do. In some cases, it might be correct to say that *c* was happy at t_1 (given that all *c*'s wants at t_1 are fulfilled). In some cases, *c* is neither perfectly happy at t_1 nor at t_2 . If *c* expects and wants to *feel* happy when *c*'s goal is satisfied at t_2 and *c* does not experience this kind of satisfaction at t_2 , then we cannot say that *c* was perfectly happy at t_1 .

It should also be noted that some of these scenarios are unrealistic. Few people only want *one* thing, for example to be rich, famous, a member of some inner circle, etc. People usually want many things. True, a lot of individuals want to be rich and famous, but they usually also want other things: friends, family, respect, security, love. Someone might, for example, want to be rich at t_1 because he thinks that then he will be respected by everyone in some group, and he wants to be respected. But at a later time, t_2 , when he is rich, he finds out that even though he is now rich everyone in the group does not respect him. Therefore, we cannot say that *c* is perfectly happy either at t_1 or at t_2 . Single-mindedly striving for only one thing in life, may lead to the fact that many of our desires will never be fulfilled.

There is wisdom in the advice that you should be careful what you wish for and in the proverb that all that glitters is not gold. Nothing of this, however, can be used as a refutation of the fulfillment theory of happiness.

6. Conclusion

In this paper, I have developed a new theory of the nature of happiness, or perfect happiness. I have tried to explain what perfect happiness *is* and what it *is not* according to this theory. The basic idea is a development of the old view that happiness is getting what you want and can be classified as a kind of desire-satisfaction theory. According to the theory, which we called "the fulfillment theory," perfect happiness is perfect fulfillment. Even though this idea is old, we have seen that the particular version developed in the present paper is new. According to the fulfillment theory of happiness, it is necessarily the case that an individual *x* is perfectly happy iff all *x*'s wants are fulfilled. Desire-satisfaction theorists have usually not said much about what it means for a desire to be satisfied. In the present paper, I have tried to make this idea more precise. I have proved several interesting theorems that follow from the theory and I have tried to defend it against some potentially serious counterarguments. We have seen that perfect happiness is a final end for everyone who is perfectly rational according to the theory, and that everyone who is perfectly rational wants to be

perfectly happy and consistent (since it is not possible to be perfectly happy without being consistent). The theory has many other interesting consequences and more could be said to defend it against various possible problems. Elsewhere, I explore some relationships between the concepts of virtue, perfect happiness and the highest good and try to show that the concept of perfect fulfillment can play an important role in a certain Kantian ethical theory (Rønnedal, forthcoming). The upshot is that the analysis of the nature of (perfect) happiness developed in the present paper is quite attractive.³⁹

Bibliography

- Annas, Julia. 1993. *The Morality of Happiness*. New York/Oxford: Oxford University Press.
- Aquinas, Thomas. 1265–73. *Summa Theologiae*. Translated by Fathers of the English Dominican Province. 1911; (reprint, Allen, TX: Christian Classics, 1981).
- Aristotle. 1992. *Nicomachean Ethics*. Many editions. Translated by D. Ross. Oxford/New York: Oxford University Press.
- Augustine. 1963. *The Trinity* (The Fathers of the Church Series). Translated by Stephen McKenna, Washington, D. C., Catholic University of America Press.
- Becker, Lawrence C. 1998. *A New Stoicism*. Princeton, New Jersey: Princeton University Press.
- Benditt, Theodore M. 1974. "Happiness." *Philosophical Studies*, 25: 1–20.
- Bentham, Jeremy. 1789. *The Principles of Morals and Legislation*. Amherst, New York: Prometheus Books, 1988.
- Blackburn, Patrick, de Rijke, Maarten and Venema, Yde. 2001. *Modal Logic*. Cambridge University Press.
- Bok, Sissela. 2010. *Exploring Happiness: From Aristotle to Brain Science*. New Haven: Yale University Press.
- Boniwell, Ilona, David, Susan, and Ayers, Amanda Conley (eds.). 2013. *The Oxford Handbook of Happiness*. New York/Oxford: Oxford University Press.
- Bradley, Ben. 2007. "A Paradox for Some Theories of Welfare." *Philosophical Studies*, Vol. 133 (1): 45–53.
- Brandt, Richard Booker. 1967. "Happiness." In *The Encyclopedia of Philosophy*, edited by P. Edwards, 413–414, New York: Macmillan Publishing Co, Inc. and The Free Press.
- Bruni, Luigino and Pier Luigi Porta. 2005. *Economics and Happiness: Framing the Analysis*. Oxford University Press.
- Brülde, Bengt. 2007. "Happiness Theories of the Good Life." *Journal of Happiness Studies*, 8 (1): 15–49.
- Chekola, Mark. 1974. *The Concept of Happiness*. University of Michigan.

³⁹ **Acknowledgements:** I began to work on this paper in 2016. The first version was finished early in 2018. I would like to thank everyone who has commented on the text since then.

- Chekola, Mark. 2007. "Happiness, Rationality, Autonomy and the Good Life." *Journal of Happiness Studies*, 8 (1): 51–78.
- Chellas, Brian F. 1980. *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.
- Davis, Wayne. 1981. "Pleasure and Happiness." *Philosophical Studies*, 39: 305–318.
- Davis, Wayne. 1981b. "A Theory of Happiness." *American Philosophical Quarterly*, 18: 111–120.
- Delle Fave, Antonella. (ed.). 2013. *The Exploration of Happiness: Present and Future Perspectives*. Springer.
- Den Uyl, Douglas and Machan, Tibor R. 1983. "Recent Work on the Concept of Happiness." *American Philosophical Quarterly*, 20: 115–134.
- Diener, Ed and Biswas-Diener, Robert. 2008. *Happiness: unlocking the mysteries of psychological wealth*. Malden, MA: Blackwell.
- Diener, Ed and Diener, Carol. 1996. "Most People Are Happy." *Psychological Science*, 7 (3): 181–185.
- Easterlin, Richard A. 2003. "Explaining Happiness." *Proceedings of the National Academy of Sciences of the United States of America*, 100 (19): 11176–11183.
- Feldman, Fred. 2004. *Pleasure and the Good Life*. New York/Oxford: Oxford University Press.
- Feldman, Fred. 2010. *What Is This Thing Called Happiness?* New York/Oxford: Oxford University Press.
- Franklin, Samuel S. 2010. *The Psychology of Happiness*. Cambridge University Press.
- Garson, James W. 2006. *Modal Logic for Philosophers*. New York: Cambridge University Press.
- Grinde, Bjørn. 2012. *The Biology of Happiness*. Springer.
- Haybron, Daniel. M. 2001. "Happiness and Pleasure." *Philosophy and Phenomenological Research*, 62 (3): 501–528.
- Haybron, Daniel. M. 2005. "On Being Happy or Unhappy." *Philosophy and Phenomenological Research*, 71 (2): 287–317.
- Haybron, Daniel. M. 2008. *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: Oxford University Press.
- Hughes, George Edward and Cresswell, Max J. 1968. *An Introduction to Modal Logic*. London: Routledge, reprinted 1990.
- Kahneman, Daniel, Diener, Ed and Schwarz, Norbert. (eds.). 1999. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation Press.
- Kant, Immanuel. 1785/2002. *Groundwork for the Metaphysics of Morals*. Edited and translated by A. W. Wood. New Haven and London: Yale University Press.
- Kant, Immanuel. 1788/1997. *Critique of Practical Reason*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press.

- Kant, Immanuel. 1797/2017. *The Metaphysics of Morals*. Edited by L. Denis, translated by M. Gregor. Cambridge: Cambridge University Press.
- Kekes, John. 1982. "Happiness." *Mind*, 91: 358–376.
- Kraut, Richard. 1979. "Two Conceptions of Happiness." *The Philosophical Review*, 138: 167–197.
- Martin, Mike W. 2012. *Happiness and the Good Life*. Oxford: Oxford University Press.
- McGill, Vivian Jerauld. 1967. *The Idea of Happiness*. New York: Frederick A. Praeger.
- McMahon, Darrin M. 2005. *Happiness: A History*. New York: Atlantic Monthly Press.
- Michalos, Alex C. (ed.). 2014. *Encyclopedia of Quality of Life and Well-Being Research*. Springer.
- Mill, John Stuart. 1863. *Utilitarianism*. Buffalo, New York: Prometheus Books, reprinted 1987.
- Montague, Roger. 1967. "Happiness." *Proceedings of the Aristotelian Society*, 67: 87–102.
- Myers, David G. and Diener, Ed. 1995. "Who Is Happy?" *Psychological Science*, 6 (1): 10–19.
- Nes, Ragnhild Bang. 2014. "Happiness and Behavior Genetics." In *Encyclopedia of Quality of Life and Well-Being Research*, edited by A. C. Michalos, 2643–2650.
- Nozick, Robert. 1989. *The Examined Life: Philosophical Meditations*. New York: Simon & Schuster.
- Pavot, William and Diener, Ed. 2013. "Happiness Experienced: The Science of Subjective Well-being." In *Oxford handbook of Happiness*, edited by I. Boniwell, S. David and A. C. Ayers, 134–152.
- Perry, Ralph Barton. 1926. *General Theory of Value: Its Meaning and Basic Principles Construed in Terms of Interest*. Cambridge, Massachusetts: Harvard University Press, reprinted 1967.
- Plato. 1925. *Lysis. Symposium. Gorgias*. Translated by W. R. M. Lamb. Loeb Classical Library 166. Cambridge, MA: Harvard University Press.
- Plato. 2013. *Republic*, Volume I and Volume II: Books 1–10. Edited and translated by Christopher Emlyn-Jones, William Preddy. Loeb Classical Library 237 and 276. Cambridge, MA: Harvard University Press.
- Rawls, John. 1972. *A Theory of Justice*. Oxford: Oxford University Press.
- Russell, Daniel C. 2012. *Happiness for Humans*. Oxford: Oxford University Press.
- Rönnedal, Daniel. 2019. "Semantic Tableau Versions of Some Normal Modal Systems with Propositional Quantifiers." *Organon F*, 26 (3): 505–536.
- Rönnedal, Daniel. 2019b. "Boulesic-Doxastic Logic." *Australasian Journal of Logic*, 16 (3): 83–132.
- Rönnedal, Daniel. 2019c. "Quantified Temporal Alethic Boulesic Deontic Logic." *Filosofiska Notiser*, 6 (1): 209–269.
- Rönnedal, Daniel. 2020. "Boulesic Logic, Deontic Logic and the Structure of a Perfectly Rational Will." *Organon F* 27 (2): 187–262.

- Rönndal, Daniel. Forthcoming. "The Highest Good and the Relation between Virtue and Happiness: A Kantian Approach".
- Seligman, Martin. 2002. *Authentic Happiness*. London: Nicholas Brealey Publishing.
- Seneca, Lucius Annaeus. 1932. *De Vita Beata (On the Happy Life)*. Translated by John W. Basore. Harvard University Press (Loeb edition).
- Sidgwick, Henry. 1907. *The Methods of Ethics*. Indianapolis: Hackett, reprinted 1981.
- Sizer, Laura. 2010. "Good and Good for You: An Affect Theory of Happiness." *Philosophy and Phenomenological Research*, 80 (1): 133–163.
- Solomon, Robert C. 1976. "Is There Happiness After Death?" *Philosophy* 51: 189–193.
- Sprigge, Timothy Lauro Squire. 1991. "The Greatest Happiness Principle." *Utilitas*, 3 (1): 37–51.
- Suikkanen, Jussi. 2011. "An Improved Whole Life Satisfaction Theory of Happiness." *International Journal of Wellbeing*, 1 (1): 149–166.
- Sumner, Leonard Wayne. 1996. *Welfare, Happiness, and Ethics*. New York: Oxford University Press.
- Sumner, Leonard Wayne. 2000. "Something In Between." In *Well-Being and Morality*, edited by R. Crisp and B. Hooker, 1–19, New York: Oxford University Press.
- Tännsjö, Torbjörn. 2007. "Narrow Hedonism." *Journal of Happiness Studies*, 8: 79–98.
- Tatarkiewicz, Władysław. (1976). *Analysis of Happiness*. The Hague: Martinus Nijhoff.
- Telfer, Elizabeth. 1980. *Happiness*. New York: St. Martin's Press.
- Thomas, D. A. Lloyd. 1968. "Happiness." *Philosophical Quarterly*, 18: 97–113.
- van Praag, Bernard and Ferrerer-i-Carbonell, Ada. 2004. *Happiness Quantified: A Satisfaction Calculus Approach*. Oxford and New York: Oxford University Press.
- Veenhoven, Ruut. 1984. *Conditions of Happiness*. Dordrecht: Reidel Publishing.
- Veenhoven, Ruut. 1984b. *Data-Book of Happiness: A Complementary Reference Work to 'Conditions of Happiness'*. Springer.
- von Wright, Georg Henrik. 1963. *The Varieties of Goodness*. London: Routledge.
- White, Nicholas P. 2006. *A Brief History of Happiness*. Oxford: Blackwell.
- Wike, Victoria S. 1994. *Kant on Happiness in Ethics*. State University of New York Press.
- Wilson, John. 1968. "Happiness." *Analysis*, 29: 13–21.
- Zamagni, Stefano. 2005. "Happiness and Individualism: A Very Difficult Union." In *Economics and Happiness: Framing the Analysis*, edited by Bruni and Porta, Chapter 12, 303–335.

The Question as to Why We Have to Live Out the Agony of Our Epoch and its Fundamental Un-Answerability: A Reading of the Preface to the 1967 Edition of Klossowski's' Original 1947 *Sade My Neighbor*

Rajesh Sampath

Abstract: This paper excavates certain impulses that are buried in Pierre Klossowski's 1968 edition of his original 1947 work, *Sade My Neighbor*. We argue that the self-suffocating nature of our historical present reveals the problem of an epochal threshold: in which twenty-first century democracy itself is threatened with death and violence in delusional neofascist attempts at national self-preservation. This speaks to a deeper enigma of time, epochal shifts, and the mystery of historical time; but it does so in a manner that escapes classical problems in the philosophy of history. Rather, by returning to Klossowski's late 1940s and late 1960s contexts while reoccupying the New Testament question of Jesus's forsakenness on the Cross, we unravel a series of paradoxes and aporias that attempt to deepen metaphysical problems of time, death, and the sovereign autonomy of human freedom and existence. Ultimately the paper concludes by offering certain speculative philosophical constructions on why today's self-cannibalization of democracy has its roots in unresolved tensions that span these two poles: a.) the primordial secret of early Christian proclamation of Jesus's death and b.) the post-Christian Sadean experiment of a philosophical revolution that was doomed to implode when the valorization of pain, suffering, and death fails to fill the vacuum left behind by atheism.

Keywords: metaphysics, democracy, death, time, epochal shifts.

Introduction

In the 1991 English translation (Klossowski 1991) of the 1967 French Edition of Klossowski's *Sade My Neighbor*, originally published in 1947, one is startled about something as simple as the succession of the publication dates. Here we are in 2021, thirty years after the English translation, which was almost twenty-five years after the Second Edition of Klossowski's *Sade My Neighbor*, which was twenty years after the original publication in 1947. Through these long arcs of historical time, we will try to link Klossowski's 1947 present passing through the 1967 Edition to our 2021 present without collapsing one into the other, conflating them or superficially dividing them in chorological time. This is not the work of

empirical intellectual history either, if the reader is curious why we are invoking chronological dates as signifiers of massive epochal thresholds in Western modernity. We are interested in the liminal interrelations of relations and differences between the two historical presents of Klossowski's and our epoch.

For both 1947 and 2021 descend from the late eighteenth century, the context of which Klossowski philosophizes, namely the era of the French Revolution and the imprisonment of the notorious libertine, perhaps the most notorious, the Marquis de Sade. We recognize the countless studies of Sade in twentieth-century continental philosophical thought, particularly French thought, namely those by Bataille, Blanchot, Lacan, Deleuze, Foucault, and others who do not immediately come to mind (Gutting 2013). We will be mindful of this infinitely rich corpus. But our focus will be on Klossowski's text, particularly with a slow reading of the translated Preface of the 1991 English Edition of the 1967 Second Edition.

One should not neglect the full text of Klossowski and all its miraculous chapters and sections, particularly, "The Philosopher-Villain" and "Sade and Revolution," "Outline of Sade's System," and "Under the Mask of Atheism" (Klossowski 1991). Yes, we must not lose sight of those elegant and philosophically profound reflections. However, given the depths of sensitivity, creativity, nuance, and path-breaking intuitive bursts that Klossowski generates in the Preface, we cannot do justice to his entire text on Sade, at least in this one article. Therefore, we will restrict ourselves to his short Preface, and in that only a few of its initial propositions. It would appear that beginning a slow reading of the beginning of the Preface alone, and not in its entirety, preoccupies us given the monstrosity, terror, solemnity, unfathomable vertigo and bewilderment of trying to live moment to moment, day to day, and week to week of our historical present at the start of 2021. There is much to say about the Preface, alone, which is where we will stay and remain throughout this essay.

Klossowski reflects in his Second Edition of 1967 of what he should have continued to pursue and develop at the time of the original publication of 1947. In a humble honesty, he says he should have 'persevered in the original design' of "Outline of Sade's System" (Klossowski 1991, 5). For he would have, then, tried to develop a 'more rigorous examination of Sade's relationship to reason' (Klossowski 1991, 5). He then develops these six major areas of investigation or what he terms 'observations' (Klossowski 1991, 5). We will quote in full all six, and then proceed to deconstruct them while filling in the 'lacunae' of his earlier analyses through our own philosophical speculations and wanderings. We speak of an impossible simultaneity and therefore the relative cut within simultaneity of these tasks of reading slowly and writing independently. We have to deconstructively read each of the six and in interrelation to one another through this immersion of the Preface to Klossowski's scintillatingly original work. Quoting the English translation directly:

The Question as to Why We Have to Live Out the Agony of Our Epoch...

- (1) Rational atheism is the heir to monotheist norms, upholding a unitary economy of the soul, along with the possession and identity of a responsible ego.
- (2) If the sovereignty of man is the principle and the goal of rational atheism, Sade, liquidating the norms of reason, pursues the disintegration of man.
- (3) In the absence of any conceptual formulation other than that of the rational materialism of his age (as the "Outline" already notes), Sade made of atheism the "religion" of integral monstrosity.
- (4) This "religion" involves an asceticism, that of the apathetic reiteration of acts, which confirms the insufficiency of atheism.
- (5) Through this asceticism, Sade's atheism reintroduces a divine character in monstrosity, divine in the sense that its "real presence" is actualized only through rites, that is reiterated acts.
- (6), Thus it turns out that it is not atheism that conditions or liberates Sadean monstrosity; rather, this monstrosity leads Sade to derationalize atheism as soon as he tries to rationalize his own monstrosity by way of atheism (Klossowski 1991, 5-6).

With these bold propositions, one can see that it is not wise to run ahead and continue to comment on what follows in the Preface. There is enough here to lose oneself in an abyssal set of philosophical reflections as we respond to the torment we embody or inhabit in our historical present. The historical present pulsates an exteriority (hence never internal to a subject or mind experiencing anything at all), which then threatens to engulf the present by drawing it into a singularity of infinite density without possessing a point in time or space. This bizarre temporalization, not of beings, conceals itself. The mysterious entwinement of being and time, not being in time, the time of being, being as time or time as being fails to reveal itself as it failed to do so for the early and late Heidegger (1927; 1962). One does not need to invoke immediate intuitions, memories, and conceptual configurations of what 'atheism' and 'asceticism' can possibly mean at the outset of our inquiry into time, the epoch and epochal shifts, and the historical present. And this is way before we even attempt an understanding of these remarkably rich registers in Klossowski's text, namely 'divine character in monstrosity,' 'integral monstrosity,' 'derationalization of atheism,' and how this is bound up with a 'real presence' (Klossowski 1991, 5-6). No, we cannot just offer philosophical expansions in a vacuum. Then again, fidelity of getting the right interpretation of the text is not our goal either.

Rather, our uncanny thesis is that Klossowski is speaking to a realm other than atheism. We explore the aporetic, multi-dimensional (therefore neither one nor binary) kind of perversion internal to Christian divinity: one that does not pit atheism against religion, let alone early Gnosticism that was expunged from formal Christian doctrine as heresy and therefore inadmissible in what will follow as mainstream, dogmatic Christian revelation across denominations. Rather, it is about a new creation, an outgrowth and therefore violent attempt to supplant Christianity with something else that can take its place. Therefore, this is not a

work of religious studies, whether motivated by faith or by a pure, secular, atheistic and scientific study of religion. If this were biblical theory, history, and criticism, then one might be tempted to trace the history of heresy and gnosis in the first two centuries CE before: that is before the time when the New Testament canon was crystallized and Christianity was institutionalized as the state of religion of the ancient Roman Empire. But this is not the task at hand, even though Klossowski concludes his Preface with an allusion to heretics and gnosis of the early Church, namely Marcion ditheism and the Carpocratian orgiastic cult (Klossowski 1991, 7). He also admits this is where his 1947 original text falters, at least as he reflects back from the 1967 Second Edition (Klossowski 1991, 7).

Nor is our work a matter of excavating a primordial archē, or opaque preconditions before Christian revelation emerged and was institutionalized as the gentile identity of Western civilization, let alone attempts since Nietzsche to destroy that identity and postulate what could be other to it or post-Christian. Nietzsche was arguably the first to diagnosis the problem of nihilism in his stunning originality, although he had just as profound precursors in Schelling and Kierkegaard who began to sense its cancerous infiltration of nascent Western modernity. His own epoch and origin was split apart between a live birth taking place and a post-natal infanticide given his confrontation with everything that must be transvaluated. All the while and through it all, he succumbed in exhaustion with a new and bold creative attempt to think beyond all that has been thought, but in fragmentary form resulting in the twilight of his own cognitive eclipse, namely the decade long silence before his death in his state of complete mental paralysis. Nevertheless, between early Christianity and Nietzsche, lies the problem of Sade that Klossowski bravely encounters.

Indeed, we must go inside his text, deep inside without knowing from what end we will emerge, if we do at all. This all speaks of an adventure, but also a potential death trap. In an uncanny sense, we know that we are living to pass through the epoch and will risk everything to come out on the other side, but also distend the great possibility of death and a point of no return. Therefore, this is not about beginning and ending, beginning and not ending, not beginning and not ending. Furthermore, all of this has to do, somehow, with the monstrous threshold of the Western secular democratic conception of 'state' and 'people' and where they stand, namely on the brink of desuetude.

Rather than theological exposition or atheistic rebuttal to any revelatory theology, we will attempt an overcoming of any simplistic sense of atheism without reverting back to the original doctrines of a canon, theology, law of a specific religion – in this case Christianity and its various traditions, such as Orthodoxy, Catholicism, and Protestantism. Atheism is not the secularization of religion like Löwith's long-standing suggestion (1949) on modernity as the secularization of the eschatological. Nor will we take up, at least at the moment, Blumenberg's momentous response to all secularization theses by his own novel theory of epochal shifts and 're-occupation' of old premodern questions and

incessant self-assertions at answers without arriving at foundations (1966). We will not speculate of what makes the atheism inherent in rationality or the rationality of atheism operate in its full depths and complexity as countless other philosophers of history have (Taylor 2007). This is not a project in the philosophy of history, the philosophy of religion, and the temporalization of the historiography of the philosophy of history, the latter of which tries to understand temporality, epochal shifts, origins, ends, continuities, discontinuities, thresholds, and ruptures in, as, and of historical time (Blumenberg 1985; Koselleck 1985). How the historiographical imagination temporalizes the history of the philosophy of history without conducting either philosophy or empirical history has to be foreclosed.

Our hypothesis is that this internal critique beyond atheism has to do with the 'monstrosity' of our historical present regarding how a new concept other than religion is being born out of a neo-fascist sovereignty that is hijacking our democratic state. This question of asceticism, incessant material production, Weberian legacies on the critique of the Protestant Ethic notwithstanding, is linked to a new theory: that is, an attempt to understand the neo-fascist psychopathology in the heart of dissolving sovereign democratic state and political power configurations. These strange formations are protruding out of the heart of our secular, constitutional, legal, pluralistic democracy, and more so in the age of anonymous, invisible, technological fabrications of real persons and events. These protrusions, and actually contusions too, both reveal the implosion and supersession of normative conceptions of democracy that have descended to us from the late eighteenth century, namely rule by the people's consent, and the separation of powers between the legislative, executive, and judicial branches. Sacrosanct among them is the notion of free, fair, and transparent elections and peaceful transfer of power when the results are accepted. Yet, this basic idea of democracy is eluding us, revealing an unthinkable moment. Nevertheless, premodern conceptions of constitutional monarchs and monarchical systems are not simply resurrecting themselves out of the death of democratic delimitations of formal tyranny. A non-dialectical movement of interrelated patterns constitute a grander, non-Euclidean event that is engulfing us now and threatens to erase us from historical time. The three axes of physical, measurable time – past, present, and future- cannot help us. Nor do we want to spatialize time in any geometric or symbolic forms. And the notion of a simple, visible, tangible, physical event called the 'historical present' does not present itself as such and such thing, event, action, person, or movement.

Rather, it is when society births its own monstrosity, which then ensnares and erases its birth (and therefore no simple definition of death or end to what is birthing), that we know we are being terrorized by something we cannot see or define. We confront the uncanniness of the question of time, but never as a phenomenon or idea. It is our question, not Klossowski's or Sade's, to understand what constitutes the 'divine' in a new way, an inhuman divine that refuses to speak

to human beings until the conception of human being is completely vanquished. This is not a resumption of twentieth-century totalitarianism as Arendt astutely diagnosed (1951) since something entirely other is emerging in our historical present.

We have to attune our philosophical creativity to what is most preternatural in relation to the Other and as the Other that we are and are becoming. Paradoxically, and in a manner that can never be easily resolved, this points to no recognition of prior selves and identities and their previous epochal identities which are other to whatever present one imagines is being unfolded 'today.' So we must inhabit a relation and non-relation, neither-nor of what is self and Other, the self and other of their relation and non-relation. Our ultimate goal is to explore this philosophically and speculatively, and, therefore, launch from a reading of Klossowski on Sade, an act of self-disappearance. We depart from both their late eighteenth and mid-twentieth century contexts without arriving at a full conceptualization of our time and the thinking of the epoch as our time. As undeniably influential previous influences are, say Foucault and Derrida who loom large, we feel a haunting distance and schism cutting within the surface that we cannot see: that something is still missing in their analyses, whose remnants and traces we do not wish to trace ourselves. Therefore, the theme of 'departure' lies at the heart of what we are trying to assert as a form of extreme philosophical transcendence that does leave the world but burrows into its darkest and deepest singularity. We cannot and do not want to return to those times of the twentieth century (or past centuries) even though our historical present, and whatever sense of 'epoch' it carries within it, is rather unbearable and suffocating. This is what we are trying to articulate in these initial, hesitant remarks. It is like a motion-event that counter-intuitively sucks one into a void where no motion would exist; yet the whole event has its own kind of motility and dynamic manifestation.

We are trying to attest as a form of temporal self-strangulation, the aporetic knot of the historical present cannibalizing itself. The excess of this process becomes the nourishment of a thinking that is not afraid to tackle what is most horrifying, menacing, and threatening to any sense of a 'unified economy of soul' to use the previous quoted phrase (Klossowski 1991, 5). The future offers a proxy of itself, not its 'real' and 'authentic self: where the past should 'historize' in sensing what we will have been without being able to survive our own death if we are in fact no longer in time (Heidegger 1962). But we do not exist after our time either looking back on it like a ghost. Nor are we stretched between a beginning and end without ever knowing where we are. For even its pull, the pull of the future, is nowhere to be found. The future, as the no longer now, is not simply that which has yet to appear, but an internal resistance and refusal of birth, emergence, and recognition of the brand new. Therefore by emerging out of something entirely Other, our historical present does not want to be named or certified by some other higher authority. That itself becomes a question not of identity or non-

identity but the mystery that is the question of time, not its answer. So we speak of an 'un-answerability' that is somehow asserting itself as some strange new content, or, more specifically something other to any kind of phenomenal or noumenal content.

Let us return to Klossowski's propositions before our content becomes completely engulfed by mind-stifling paradoxes about time, historical time, epochal shifts, and the philosophy of history. This will be difficult to keep open so that many interweaving dimensions begin to flicker in tiny sparks as we traverse Klossowski's eerie propositions on Sade. We want to start the fire by rubbing the twigs against each other, and yet all we see is the hint of a beginning, not an actual one. We want to unpack them further while imagining the theory of neo-fascistic implosion of our democracies in our twenty-first century historical present and the strangest question of all: that is the problem of an alternative 'real presence' as the ghost of Christianity's invention but not for the Eucharistic purposes of today's Christian revelation. This is to say we do not want to repeat and resuscitate faith from the original events attested in the *New Testament Gospels* through the Western and Eastern patristics of the ages down to the current doctrinal and ecumenical statements by the religious authorities of the day, say the Vatican and the Pope.

Rather, we are going into the monstrosity to learn what has yet to be said about the meaning of the being of the human and what could be other to that, which the West has never succeeded in articulating. But hauntingly, this means we must inhabit Christian revelation from another side, deep within its horizon, but never articulated within its text; it is an illusory exteriority and the promise of a passage to another side, yet somehow encrypted as a secret locked away in what was, what is, and what is to come. The Johannine revelation does not terrify us with its monstrous content and rich symbolism, but asks for another text to respond to it and not repeat its bizarre late first century or second century CE contents. In principle, such a text, therefore, cannot be included in a fixed New Testament canon; nor is it part of the remnants that exist as non-canonical texts like the Gnostic Gospels (Pagels 1989). This other text is what will attempt in our reading of Klossowski on Sade.

Main Text

In the Gospel of Mark, which is unique, cryptic, and mysteriously terse compared to the other two elongated Synoptics, Matthew and Luke, which draw from Mark's content, let alone the rich theophany of logos that is the Gospel of John, we have one of the basic puzzles of which there are many in the *New Testament* (Lane 1974). Any quick intuition that may descend on us in a human way must be bracketed phenomenologically; nor should we move to circular logical fallacies that spin us around. The question of Jesus's departure is linked to a question of forsakenness. This relation is likened to an event as dissemination without a source. And this uncanny relation, which refuses to present read-made answers,

prefigures a completion to God's life on earth before announcing the breakthrough moment that founds all of Christian proclamation and kerygma, namely a resurrection – 'ἀνάστασις' in the Greek New Testament – and some kind of spiritual bodily – 'σῶμα πνευματικόν' in 1 Corinthians 15:44 – appearance after death to human witnesses. Humans, who do nothing but be born, live and die, cannot possibly understand what it means to witness death, let alone what is other to it as a miraculous reappearance. One cannot think of seeing a friend die one day, and them showing up at your door the next day looking like the way they were before the day they died. In Mark 15:34, which Matthew repeats in 27:46, it says:

Jesus cried out in a loud voice, "*Eloi, Eloi, lema sabachthani?*" which is translated, "My God, my God, why have you forsaken me?" (The New American Bible 1987, 1087)

We will keep this great, yet still to be deciphered Aramaic question from Jesus on the Cross, in the background so to speak. It forms a type of counterpoint as we work our way through Klossowski's propositions on Sade's attempt to go beyond atheism without returning to religious morality. To repeat the hypothesis, this will allow us to penetrate the depths as to why the living agony of our epochal historical present fails to present answers about the utter unpredictability of human history. Scattered through is not the point of finitude in historical time, but the mystery of the finitude of historical time itself, that it – historical time – is never present, but what is before and after is also crossed out. The question is not to divine some end goal towards which history is progressing, but how history itself faces its own historicity, its mortality and finitude, as some other logic of time pulses and throbs with a new possibility for thinking. This would be a name other than 'History,' and hence our historical present. We grow weary of using the term 'historical present' if we do not want to define it in advance to the reader.

Passing through the event on the Cross is not for nothing; it is phenomenologically bracketed to isolate the event, an event of appropriation, and not a reaffirmation of Christian faith contents understood in dogmatic and literal terms. Jesus's agonizing suspension, where no answer is given to the question right at the threshold of his death, is likened to the Sadean question of how a divine monstrosity is concealed within suffering. Both are stretched events from which a mystery begins to reveal a deeper enigma concealed within it – the alterity of God's voice to Himself, which means other than any relation between man and God, let alone the embodiment of God as man, namely Jesus. This is about non-relation, differentiation within the relation, not of an event but the event of difference itself that never appears. This can be contrasted with the rather simplistic distinction of *either* a Godless reality – the atheistic witness that God does not exist when one is abandoned – *or* the Eucharistic 'real presence' in which the sacrifice of Jesus, who really died, and his perpetual memorialization is anticipated in an earlier, Last Supper scene in the Gospels's account of his life.

In that last meal with the disciples, Jesus does not speak of being forsaken at death but betrayed by someone who will catalyze the narrative events that

ultimately leads to his death. We are not going to unpack these theological mysteries (Lane 1974). As stated before, this is not a logical question or paradox to be solved: that is, how God can forsake Himself if Jesus is the Son of God and co-equal to his Father (John 10:30) as part of an eternal Logos consisting of three Persons of Father, Son, and Holy Spirit with his unique two natures of – human and divine – which neither change or divide the other, nor separate nor mix with the other (Chalcedon, 451 CE). It is not an eternity that can end; nor is it an eternity without origin and end, which human reason cannot conceptualize on the basis of its own experience rooted in mortality and finitude. Rather, it is an affront that a perfect harmony of love results in dereliction. Jesus cannot be like a dying man who keeps his faith to the end that he will be saved, only to be deserted right at the moment of death where all hope evaporates. And the death of all humans means no human comes back to attest to the death that occurred, what it means, and where they are ‘now’ – in a realm beyond death, in a realm beyond events in pictorial linear time: this akin to the great problem Heidegger explored in trying to derive the ‘inauthentic’ linear time from something even more primordial and therefore never present as a ‘now’ (Heidegger 1962, 373). Then again, the Son of God is not some mythological non-human or God-human hybrid incapable of dying even though he takes human form from time to time, i.e. the legacy of Marcion heresy. There is no role for the ancient Greek myths here either like Zeus and Hercules.

Rather, the forsakenness will drive us into the heart of a singular question of dying and death that points to the transcendental horizon of what dying and death mean when it is stretched between these two notions: a.) a singular conception of completion without content/answer or even ethical orientation as to the right and virtuous way to comport oneself to death as a stretched relation to a limit and b.) the suspension of what an answer could look like right before the event of death is realized as a passing through in the heart of divine agony. We will keep this polarity in mind in trying to expand on Klossowski’s un-developed enterprise; or if you will what was forsaken in the 1947 project as attested by the Second Edition of 1968 where Klossowski confesses a lack of perseverance in the pursuit of a system. Something ended in the 1947 project that the 1968 Preface tries to attest in justifying the contents and chapters of its own later edition.

To stretch out Jesus through all these theoretical possibilities carries a series of uncanny resemblances with Klossowski’s ponderings of the true meaning and intentions of Sade beyond the simple notion of his unrestrained immorality, depravity, and ultimately inhuman nature, even evil. Let us return to Klossowski’s six propositions starting from the first:

- (1) Rational atheism is the heir to monotheist norms, upholding a unitary economy of the soul, along with the possession and identity of a responsible ego (Klossowski 1991, 5).

To even attempt a type of rationality to Sadean logic betrays the notion that his punishable otherness, his in-humanity, does not and should not deserve a type of

new moral valorization: one that we must conceive apart from any Judeo-Christian logics of 'though shall not kill' and in the New Testament Christian registers – 'love the enemy' and 'love the neighbor as thyself.' If that were the case, then it must be one that cannot appeal to our humanity in its basic instincts for goodness and hope; but this does not mean its apposite term is some glorified or mythic sense of evil and the demonic, i.e. a type of generic Satanic will apperceived in biblical-faith-based terms. We will assume that this is not Klossowski's intent, which we will argue carries a certain positivity even in a quest that is doomed to fail. It is not about saving Sade from his own reputation imparted to us after his time passed, as of one of the earliest of modern peoples whose viscerally vile and cruel nature cannot be denied. Some could say we hate the Sade because he created us, and yet we did not initiate the event that led to us becoming what we are. This weird creator was a banished created from the original Biblical Creator, an erasure which then becomes the origin of our modern sensibility. Yet even these reflections are a bit reductive; they presuppose naïve notions of time and becoming, and therefore inadmissible for the task at hand. We do not want to parrot what others have intuited about the Sadean epochal threshold (Foucault 1978, 148-149).

Our question is how to leverage the Sadean analysis offered by Klossowski with the question of departure and forsakenness on the Cross in *relation* to the suspension in answering the agony of our epoch as the death of democracy. But this term 'death' is just a placeholder for another term. We are speaking of the increasing neo-fascistic creation of a new type of monstrosity apart from the twentieth century. When we think of the last century, we see purely racist and colonialist ideas of totalitarianism and fascism; the latter threatened the internal development of true, peaceful, secular, constitutional Western democracies and their promises of individual equality, freedom, justice and rights in the previous ages of democratic revolutions of the eighteenth century and the decolonized, nationalist Global South democracies that hatched in the twentieth century. Sade's late eighteenth century is not simply an epochal threshold between the following: a.) the libertine, aristocratic, unrestrained, perverse creativity of what a modern psychological conception would construe today as 'psychotic' and b.) the modern democratic, social contractual constraints on a freely accepted, anonymous distribution of rights and duties to mitigate harm for any individual or group in a state of nature. It would seem that is what is criminal about democracy is that it miniaturizes the vast possibility of a soul that could be distended in a million different directions down to the individuated subject isolated to its boring, banal, and bounded point of invisible citizenry. It is tempting to expand on what this means for our empirical political present. However, Klossowski offers us the link between that eighteenth century context, which will not be elaborated further in our analysis, and the threshold of democratic modernity and its apostolic sanctioning of 'individual rights.' The individual right is not the right of the individual to be free of all individuals. Instead, it means to be free enough, at least

for the perverted, demonic and evil Sade, from the individual God to rationalize the ascetic will to enjoy the killing of another human being as the final act of transcendence; and this will to kill at the heart of Sadean logic, even as it tries to create a new a-morality against the morality of 'thou shalt not kill,' is what is being enacted today with regard to democracy's self-homicide. And this is what frightens us and stirs to action, namely to destroy this will to kill that was born with our modernity. For us, it serves as a buried possibility to theorize the transition from our historical present to another given the agony of living out this seemingly and indefinitely distended stretch of catastrophic time. Something other is emerging in, as, and in and of the place of our historical present of pessimism and turmoil, but that 'other' as something that remains unrecognized. That means something other than our epoch and its terminal point, or the endless deferral of an end, or embracing endlessness itself as if we could. All are impossibilities and what is other to them all remains concealed from human senses and experience; there is no spatial representation that the imagination can generate.

The triangle of our hypothesis starts to dissolve: Klossowski's Sade, forsakenness as departure from an alleged event of God's revelation nearly two thousand years ago, and the self-strangulation-temporalization of an epoch that is being lived it out in pure agony. It is as if all of Western linear historical time, beginning with the Judeo-Christian narrative and the ancient Greek and Roman pillars, is but a blip or instant that speaks to an-other side of what everything has truly meant but not from our standpoint as retrospective-looking historical present. Rather, the historical present has revealed that a Sadean logic has infiltrated the idea of democracy, a pushing of its limits beyond its metaphysical conception as something necessarily by and for the people. Today, democracy is willing to risk its own extinction, but not just in the passing cult of personality, a new type of anti-democratic sovereignty parading around as populist crystal of true democratic freedom – or the 'actual voice of the people' embodied in one clown-like tyrannical figure whose megalomania knows no bounds and has no moorings in reality. Because that self-denying concoction is being manufactured by a people from within a democracy, we have this disturbing feeling that our present is being held captive in a transcending yet specious social movement that justifies violence against the government and state. And the Hobbesian legacy of the social contract and theory of sovereignty suggest that the state, in this case a democracy and not a constitutional monarchy, has a duty to protect from harm as the ultimate form of well-being, peace, and non-violence. For those who live in them, democracies must be protected from harming themselves.

However, neo-fascism is the use of democracy against itself while a new ethos of self-righteous appropriation of national identity for one people and one leader threatens to dissolve all in its midst: what makes it 'neo' is the fact that it is happening in arguably the world's oldest modern democracy, namely the American one, whose constitutional constraints were created over nearly two and

half centuries ago as opposed to the twentieth century fascist movements in Europe. Germany, of course, is the most diabolical example, which had no democratic tradition in the nineteenth century, let alone the eighteenth century. All of these statements seem obvious, and perhaps its apocalyptic proportions seem to be exaggerated. But predicting our political future in real-world matters is not our concern. The real question is their deep underlying presuppositions, the dark forces inhabiting the conditions that make the historical present possible while erasing any possibility of how to understand what is happening in all directions and in accelerated fashion, almost day by day, and in some instances, moment by moment. This requires a return to the Klossowskian theoretical expansion of Sade and Sadean experience and its bizarre temporalization. Klossowski reveals to us a series of interlocking, aporetic knots, which are difficult to disentangle while eluding any linear unfolding of a clear argument and analysis.

The Haunting in Klossowski's Text

Starting with the first of Klossowski's six propositions, these are the terms of our analysis quoting his actual words and phrases:

Rationalization of atheism

Monotheistic Norms

Unitary Economy of the Soul

Responsible Ego

Sovereignty of Man as Goal of Rational Atheism

Disintegration of man

Rational Materialism

Atheism as Religion of Integral Monstrosity

Religion – Asceticism and Apathetic Reiteration of Acts

Insufficiency of atheism

Asceticism – Divine character in monstrosity

Divine as Real Presence – Actualized through rites as reiterated acts

Monstrosity derationalizes atheism while Sade tries to rationalize the monstrosity by way of atheism (Klossowski 1991, 5-6)

For Klossowski, one could assume that trying to rationalize atheism, to give it a self-justification is futile: because at the surface, it appears to be a type of liberation or freedom that reveals ironically not something gained but in fact that something is actually lost. And this subtraction is not simply this common notion that has been advanced perhaps since the eighteenth century Enlightenment and certainly during the nineteenth century: the act of killing God, denying the existence of the conception of God, or deciding to live apart from or in the absence

of any conception of God, is an act from which we have yet to recover. Those are the classical registers of atheism, and many in the West, perhaps more so in the younger generations, usually cling to them because they are easy to grasp. Moreover, one may assume the epochal shift is one of premodern, pre-Enlightenment monotheistic religious 'norms' as Klossowski says (Klossowski 1991, 5), whereby the distinction is not yet between a religious or atheistic self, to how the 'unitary economy of the soul' and 'responsible ego' (Klossowski 1991, 5) normalize in the modern secular context. For modernity is supposed to contrast discontinuously with what was simply the product of the Creator-God in the medieval, feudal, pre-modern Christian context. However, this is not the shift Klossowski is concerned with in his analysis of the perplexing figure of the Sade. Again, we are not, at least at the outset, investigating debates about secularism, atheism, religion and the origins of modernity (Blumenberg 1985).

The question is this: how does the 'unitary economy of the soul' and 'responsible ego' congeal out of the passage between a formal epochal manifestation of the soul seeking to join the transcendental, invisible, 'living' God with the successor epoch when that ascent to reconciliation is no longer desired or possible. What is supposed to grow out of this transition to atheism is the belief in the 'sovereignty of man' (Klossowski 1991, 5), the individual as free (unbound) and whose freedom is constitutionally defended. One belief is replaced with another: the movement is from a.) The non-sovereign nature of man's debt owed for the original sin he committed and therefore must be expiated for salvation to b.) The acceptance of man as free unit of existence where even the feeling of absolution becomes non-existent even when the origin of such freedom remains obscure. Yet this new state is actually undetermined. And therefore there can never really be a true state of atheism, and not because God exists as a prohibition or commandment against His own defiance and denial. The impossibility and possibility of God to exist in so far as this is understood always as man's relation to that impossibility or possibility is never reducible to human reason alone; then again, it is not a question of validation or hope in something faith can prove. What Klossowski sees in the Sadean moment is not a reconstitution of a new being – be it God or man – but actually a 'disintegration of man' (Klossowski 1991, 5).

Let us unpack these first set of reflections within our hypothetical structure of Klossowski on Sade, departure and forsakenness at the Cross, and our attempt to theorize the neo-fascistic implosion of normative democracy. What the death of democracy, or the belief in individual equality and liberty protected by the state, can possibly mean, after the so-called triumph of democracy over twentieth century totalitarianism and fascism, remains an open question for now. How are all three interrelated and not just as a question of succeeding epochs: the historical Jesus of the first century to the eighteenth-century Sade to our twenty-first century enigma that refuses total comprehension and understanding but exacerbates agony in living it out as an indefinite stretch of impossibility. It seems we will never become or overcome the stretch because we cannot even live in the

now or the moment or even identify with them. The ecstatic stretch, as Heidegger showed, is never reduced to a present now, let alone a present recollection of a past present or a present anticipation or expectation of a future present (Heidegger 1962). Perhaps we have never lived at all, and all that is and has been is just one long nightmare from which we cannot awake; so even the notion of facticity and the basic state for us is to be-in-the-world (Heidegger 1962) is also under question. We may not even exist at all as this historical present, which raises the stakes of how to deconstruct the limits of our own thinking. If there is a linkage of rethinking atheism and attesting to the pain of living out the historical present, then perhaps we should return to the forsakenness while incorporating what Klossowski says in these initial moments.

The Cross is not just a symbol, which has a central place in one particular world religion, namely Christianity. It is also a phenomenological bridge whose simple intuitions must be suspended: the total completion of an obedient will to live and die in all the pain and ignominy of a persecuted human being requires the utter annihilation of the total person, both body and soul, and its last second clinging to what could be beyond as one approaches the limit of non-being. To go forever into the abandonment is not proof of faith or acceptance of what is to come – namely a promised resurrection and therefore death means nothing, and in fact never occurs. Or death is just the ruse of its real possibility or the ultimate joker who can convince people of a reality that actually does not exist. Whoever said he we have to name cessation of physical life as this thing called ‘death’?

For Klossowski, this ‘inheritance’ of ‘monotheist norms’ (Klossowski 1991, 5) makes the project of an atheistic self-responsibility linked to the paradoxical goal of total sovereignty and individuality – the human as free from everything – while also a ‘disintegration’ ensues. This entropic disorder is the necessary rebuke for anything that can be identified as the moral value of obedience of will to death for another, Jesus’s God (*Eloi*), to the necessary self-implosion of the democratic sovereign self that would rather kill the social body in the name of perseverance. The Sadean moment is not simply the eighteenth century figure who was imprisoned after democracy was born. It harbors the secret of what sadism even means. This requires rethinking departure and forsakenness in general, but in non-Christological terms and the consequent destruction of political sovereignty. It is our task to define what these non-Christological terms are given what we have inherited down the ages in terms of Christian theology and its various traditions-Orthodoxy, Catholicism, and Protestantism. Our intuition is that this onto-theological construction of Western metaphysics is somehow bound up with a material apocalyptic: that is when even a democracy fails to defend itself from its own collapse into neo-fascistic, life-denying transcendence against the banal form of peaceful democratic transfer of power. The psychopathology of this whole movement awaits a more comprehensive metaphysical elaboration.

A cracking and splitting starts to appear between two poles, what we can term pre-democratic religious sadism and post-democratic and post-atheistic

generalized social sadism. It is the invisible space of difference between them that cracks or fissures, and therefore what appears are two counter-directions for analysis at the same time. But the point of origin of the crack remains hidden; the trace of that origin cannot be identified. The crux of the matter is an epochal threshold, whose temporalization is rather indiscernible, and not our empirical historical present, our current political crisis and the rise of right-wing authoritarianism and hyper-nationalism (U.S., Brazil, India), climate change, pandemics, the normalization of mass, violent social movements, and massively stifling inequality (socially, culturally, economic, politically, racially, sexually, etc.).

Rather, we need to understand how the non-dialectical relations and differences between atheism, monotheism, and democratic secularism instantiates a non-psychoanalytic yet sensualized ecstasy of religious pain and violence committed to an individual body, beginning with Jesus, to the sadistic generalization of purposeful and distributed pain and violence committed to today's democratic social body. A macro-suicide of democracy itself points to a 'beyond' state of its death that cannot be conceptualized in either pre-democratic (constitutional monarchical or tyrannical forms) or today's Western democratic norms of republicanism and parliamentarianism, say an election or constitutional crisis. Executive power and its abuses are not simply the matter, which is omnipresent as today's outstanding legal and constitutional question when a democracy is under threat. Democracy in many ways is the highest peak of illusion that monotheism demands when is committed to the notion of being One. Perhaps, democracy has always been an impossible idea, a necessary illusion. Atheism fulfills its conception in democracy, which fails to see where it's true birth lies. Inversely, one can say that democracy's own being towards its birth is never a possibility; only something that teases those who live within it, and yet are marginalized and oppressed under the veneer of equality and liberty. In fact, there is no real and universal truth as to why democracy had to be born at all, regardless of all those philosophers of history who proclaim teleologically that it is the 'end of history' (Fukuyama 1989). Let us remain in the crucible of this matter for some time before moving on to the remaining propositions on the 'religion of integral monstrosity,' 'religion as asceticism and apathy in constant repetition,' 'the divine character in monstrosity,' and finally what can only be explored in anti-Christian terms, this profound question of presence. Those terms are critical to reveal another kind of 'real presence' (Klossowski 1991, 6), a non-Eucharistic body, that is looming over us in this dark age of violence and disillusionment from both dogmatic religion and self-asserting atheism as valued in its own right. Atheism deceives itself as the successor to all dogmatism, and the utter impossibility to continue in this indefinite stretch of epochal agony. At the end of the day, atheism is meaningless. History may be meaningless for some, but when history consumes all meaning without a trace, then we have another matter altogether.

The cracking between the poles requires us to look at the cracks in each pole while trying to see the non-spatialized, intertwining-chiasm of how all points

intersect in interrelational patterns of movement in the arduous ask of thinking. The hardest thought is not some kind of sadistic enjoyment of Jesus's alleged masochism at that moment of abandonment on the Cross. Going towards death and running ahead to a transcendence that was preordained but not typified in a post-death resurrection appearance is likened to the infinite intensification of the will to be obedient to the will to death. Yet this will to will as death indefinitely stretched is a real possibility, but not for the sake of one's self-annihilation (suicide). Jesus did not commit literal, symbolic, or political suicide by being who he was, is, and to come. Rather, it is a commitment to the Voice speaking to itself of the highest act of sacrifice; nobody can enter into this secret messianic-consciousness as the infinite approach to death is actually culminated in the articulation of something that has already left, namely the protector waiting at death's gate. The receiver is not there, but that means where Jesus's cry was going is something so transcendent that it defies any simplistic notion of where God 'is.' One cannot even interpret his last question by playing with the notion of 'forsakeness' as something twofold – as in why is he being given up, as in offered as a sacrifice in addition to being left behind and alone in pure anguish. The two together form an aporia for which one side cannot speak for the other. The asymptotic approach of an infinite obedience as infinite act of love conceals another kind of finitude that defies human reason's philosophical and mathematical definitions of the infinite and the finite. This, paradoxically, is how the death of God can be *completed* so that God can be vindicated of any alleged lie that the death was just a trick or not real. God is not the joker par excellence. God did promise to come back by giving someone up completely, where no return is possible: the paradox is that this God and someone else are one and the same Logos (John 10:30).

The Cross splits apart between these vectors: a.) the infinite longing to die stretched out in agony because of the original obedience or the equality of Jesus the real Man who fails pain like us and the ineffable, omnipotent, omnipresent present monotheistic God or the Preexistent Logos that can never be approached by man; b.) the returning and encroaching moment of its transcendent temporalization precisely in the paradoxical act of withdrawal and abandonment; and c.) the increasing dramatic excitation that one is about to die alone in the name of an Other that is carrying over and beyond the possibility of death itself. But this Other, who deserted only but itself, crosses out the aloneness, for it is not a more sinister version of itself: like a copy watching itself in such pain but also in a sadistic excitement, just waiting in laughter at the humiliated body, to see what will actually happen at the event of death when one cries out for death to arrive. Rather, this condition of the highest religiosity prepares the way but also threshold between repeated, retrospective memorialization of the God-man who died that one Friday morning and the instantiation of the 'heir' to monotheistic theophany: namely Sade's 'rational atheism' inscribed in the very birth of democracy, the 'principle and goal' of which is the 'sovereignty of man.'

(Klossowski 1991, 5). As strange as this linkage is, we must keep in mind that we speak of the mystery of the epochal threshold, and not some empirical proof of a real historical connection between the Cross and Sade. Most would say any conflation of the two is completely absurd and morally repugnant, and for religious authorities, the highest form of heresy.

The path from Jesus's abandonment on the Cross to Sade's rational atheism flickers in an instant, and has nothing to do with a seventeen hundred year separation between two real historical beings. For the other pole, the post-democratic and post-atheistic historical present requires a set of reflections on what it means when democracy dies, and the horizon of Being beyond religion and atheism starts to appear. This is where we turn our reflections now, the splitting apart of relations and differences within this impossible question: if religion to atheism creates the democratic sovereignty of individual man, then what is the crossing of atheism to a realm beyond itself in the dissolution of the sovereign notion of democratic existence, whose metaphysical foreboding frightens us to the core. Taking away real equality and liberty from one who is legally and conceptually free is like taking away a security blanket from a child. What happens when a democracy starts to swallow itself up? The question could have been ludicrous decades ago, but today all norms have been thrown out the window.

This is where we can challenge simple notions of asceticism as the repetition of acts emptied of content and meaning, for example when a political election is not real or does not mean anything and whose reality can be questioned, just like the God who abandoned Himself at the moment of his own death. Repeating false elections that do not mean anything is not like the repetition of Jesus's hopeless and absconded deliverer from death, not the death and resurrection memorialized in repetition in Christian faith and practice, i.e. Eucharistic connectedness of past and future. Rather, we speak of that which is other to democracy, and not a simple repetition of the passage from religion to atheism. The passage from atheism to another is what we are considering as the post-Sadean psychopathology of a neo-fascism, one which rationalizes the self-destruction of the democratic social body. By trying to understand, we are trying to vanquish it. Would it be naïve in putting forward this intuition while claiming, perhaps disingenuously, that we are not able to represent our historical present whose shape and form refuses to present itself, refuses to answer us in the agony of our delayed completion? When dying becomes impossible is when democracy as the embodiment of the politics of finitude is under real threat.

The only way out of both poles crossing themselves out without suffocating within our own 'present' that is crossing itself out is a return to a speculative metaphysical theory. But not just any other metaphysical redux of the past, but rather, one appropriate and appropriating for our times and therefore unlike any other previous metaphysical tradition. The investigation must ask now what exactly Klossowski is speaking about regarding a type of new invention that emerges out of the Sadean historical threshold of experience. It is a threshold

because what is before and after do not appear in any simple apparition. Rather, we must go deep inside this black hole so to speak and try to decipher what exactly is being articulated in his remaining propositions.

Now we can turn to the abundant richness of propositions 3 to 6 to round out the last section of our investigation. The key terms for proposition 3 are 'rational materialism' and 'atheism as the religion of integral monstrosity.' As we have seen before, any common or intuitive senses of the distinction between atheism and religion evaporate, as does secularism as a substitute for atheism in contrast with religion. A philosopher of religion might try to probe what all these terms mean today, for example why the secular age does not replace religion but proliferates its senses and manifestations in new ways (Taylor 2007). But in Klossowski's strange and new ponderings, the traditional three moment Hegelian dialectic collapses and numerous other possibilities begin to proliferate. The linkages with current notions of ethics and communitarian justice are not so clear because, for us, a strident and self-conscious deconstruction of the emerging neo-fascism – and its underlying sadistic structure – require our unrelenting philosophical and metaphysical attention. That will require speculative philosophical wanderings because it is the groundswell of intuitions rooted in an anxiety that has more to do with fundamental questions of metaphysics rather than our current enslavement to the incomprehensibility of our empirical historical present. Let us venture some of those possibilities in this last section of our investigation.

For Klossowski, Sade, as impossible as it may seem to those who are 'normal' and do not or do not desire to do what Sade advocates, there is some rational connection, even in the absence of a 'conceptual formulation,' that links a 'rational materialism' with atheism as the "religion" of 'integral monstrosity' (Klossowski 1991, 5-6). There is no simple declaration of the obsolescence of religion and therefore the disavowal of a dogmatic concept of Godhead, which would anchor one down to prohibitions, taboos, and commandments that one can never transgress. Instead, something else is occurring. The material age one can say is the extension of Christian delay on the Cross as the embodiment of death takes shape as forsakenness; if one is not received after death, then one remains and this is transmuted into an internal spirituality of self-reproducing death that will eviscerate all matter. Atheism is not the absence of the Godhead and therefore the dialectical opposite of dogmatic religions, say Judaism and Christianity and their distinct revelatory structures. Rather, for Klossowski, it is in the Sadean threshold that we have the mystery of a transmutation: atheism *becomes* a "religion" of 'integral monstrosity' (Klossowski 1991, 5). What is this term 'religion' as it is being used here? What is the meaning of 'integral monstrosity'?

Before we analyze how Klossowski answers these weighty questions with the germinal formulations of propositions 4 through 6, let us take a step back. We will have to return to questions of 'asceticism,' 'apathy,' 'reiteration of acts,' and 'rites' (Klossowski 1991, 5-6) in the Sadean event of destruction. A type of

perverse positivity imagines a type of religion of the thanatological will to death that Klossowski says, at least here in this Preface, which Sade is initiating. The question is how to interweave this back to the problem of the epochal threshold. This is long before one even dares to confront propositions 4-6 on 'divine character in monstrosity,' the "real presence," and what we interpret as the non-circular twist in proposition 6: the last proposition is on how 'Sadean monstrosity is liberated' by 'derationalizing atheism' paradoxically by trying to 'rationalize the monstrosity' (Klossowski 1995, 6). We hypothesize that a phenomenologically complex robust and embodied notion of radically singular, individuated, non-relational, and non-transferable atheism is buried here; by that we do not mean an external frame that individuals enter into like they would a communal religion, but what is Other to any theocratic contract of this worldly-beings. One can say it is an atheism of one who is none other, and therefore other to atheism in group thought. Our interpretive assumption is that all of this points to a macabre yet mystical content waiting to be articulated. In our hypothesis, there is some buried linkage between forsakenness on the Cross (before Christian proclamation of resurrection is revealed in the Appearance to the disciples) and our historical present's self-consuming democratic sovereignty, which is the problem of an epochal threshold.

Turning to the propositions, let us philosophize about the layers of content buried in them. The 'religion of integral monstrosity' suspends any simple notion of departing from sacred commandments on not killing, which itself can invite the sanction of the death penalty. To punish the body and witness death is the passage between prohibitions to a heightened sense of freedom as if one were not reversing the Fall that gave birth to death and sin. But, rather, we have a rewriting of a pre-Fall state where pleasure is articulated in pure principle of freedom as evil. This is not the serpent suggesting that an apple should be bitten by the first two divinely created ancestors of all humanity. Rather, for Sade, the transcendence of a restricted conception of the human does not lead to a trans-human monster or inhuman, but a deeper level of interiority bordering on the mystical. The repetition of acts that creates excitation in the punishment of the body points to the liminal moment when one is approaching one's own death: the stretching event as pure expulsion of one's own self-consciousness of body and soul to a body becoming monstrous. But instead of consuming the monster's body as an anti-sacrament, the apex called death is transferred to the violated other, the one who is brutally subjugated, as the only way to rationalize a type of 'monstrosity.' This is not circular but an ecstatic event of interrelations that requires greater probing. Atheism is the content of this monstrosity. And the form takes on a 'religion' unlike any other existing world religions, which would negate atheism and promote compassion and justice for those who suffer.

Contrast that Sadean monster-event with the forsakenness on the Cross. For God to empty Himself completely in the event of death forces us humans to relate to an unanswered question when Jesus cries out about being jilted; the long

stretch of his approach to death shows another type of completion taking place. And this will to obedience to the Father's command to accept death founds Christianity's tropes of mercy, love, justice, and compassion for all who suffer. Believers no longer suffer in vain because suffering is not without hope given Jesus's death for humanity's sins and his triumphant resurrection. The Sadean threshold, which Klossowski miraculously unfolds in his unsurpassed analysis, has something to do with a confrontation with what would be other to the Cross. For us, this has some erratic linkage, although it appears to be a leap in logic, to the historical present on the self-suffocation of democracy whose logical extreme is maximum sovereignty of individual man blown up to all of society, not God or an individual emperor. Sade, in a way, shows a precondition of democracy as a monstrosity of infinitized individual sovereignty in the banal and abstract constitutional concept of a 'citizen' – when such a privileged right is discarded. One is free to do things allegedly in a democracy in contrast to a totalitarian dictatorship, but not whatever they want. Sade's monstrosity is also restricted because it cannot escape the threshold between Christian sacrifice and secular, constitutional democratic modernity. For us, however, not Klossowski, we are concerned with the rise of neo-fascism today, when all is sacrificed for an irrational drive to controvert all norms, procedures, and techniques of self-restraint that would otherwise guarantee a democracy from imploding. The first step is the fabrication of a threat to democracy that must be preserved; and in the perverse logic of saving democracy from an illusory threat, democracy itself must be destroyed. At that threshold moment a new totalitarianism begins to be conceived. Some other form as perpetual terror and self-destruction attempts to replace the concept of democracy.

Christianity cannot complete its own conception outside of the New Testament canonical text culminating in the supernatural tale of good at war with evil. In response to this inconsumable exteriority that Christianity cannot see, Sade becomes the material substitute of what would be Other to that longing, anticipation, and temporalization of the relation to God in a second act of reconciliation yielding a new creation. That is, after the first Reconciliation – or 'when the time was fulfilled, God sent his son' (Galatians 4:4) – but not to live but to die and reconcile all humanity back to God, which had previously fallen into sin. For Sadean monstrosity and its 'derationalizing of atheism' (Klossowski 1991, 6), it is possible that the human can approach its own divinity if it reaches a point of pure apathy and detachment in the repetition of acts that becomes the Sadean experience of pleasure when orchestrating and finally entering into the death of the other body. The time is unfulfilled and what is sent or delivered is a false sense of freedom whose pathetic reality is nothing but matter, blood, and remnants of flesh of the other. Such a 'real presence' is futility and forsakenness extolled to a pure and eternal abstraction, and hence a 'divine character in the monstrosity' (Klossowski 1991, 6). The real monster is that what born out of individual sovereignty. This is not a bounded object but the presupposition and purpose of

all democracy shape shifts; it becomes the tyrannical will to terrorize in the absence of dogmatic religious and moral constraints. What emerges is an embodied atheism that tries to articulate self-consciousness of its own content in the form of ecstatic, frenzied Sadean commitment to kill to satisfy pleasure. It turns out that killing God is not the most unthinkable event, but historical man killing himself.

Conclusion

For our epochal threshold, we can take this inverted relation between the Cross and Sade and try to situate it without collapsing its poles in to each other. The point is to bring to expression the haunting and anxiety of how History stands at the threshold of pure extinction: what should have culminated in the sustained, self-reproducing peace and eternal kingdom of a normal, stable, functional democracy, where no blood is shed, becomes its opposite. History is the memorialization of what has become extinct, but the idea of the extinction of History itself is not simply a negation of negation. It is a crossing to an-other side, a side other than History, which the monster of the neo-fascist historical present is trying to birth. Unlike Sade, this is not a matter of the individual unit, the basic atomic mass called the citizen that creates their own democracy for one, where only their rules apply and all other human beings are subject to a non-monarchical and non-state actor, namely the sadist who operates in secret and alone. The sadist transcends the highest justification which the state accords only to itself since the beginning of the social contract: the right to kill. However, with the historical present, this is now raised to the level of the entire democratic social body, when society as one large Human is not the preserved Leviathan of Hobbes in its pristine state, but, rather, a living megalith and a new idol. Its purpose is to destroy everything, including itself, and thus becomes the object of a national patriotic sacrifice that has no logic. To ask citizens to destroy the very thing that makes their freedom possible is like an inverted Sade, or rather Sade generalized to the idea of a new polis, the psychopathology of the neo-fascism of the historical present. To do the unthinkable and take pleasure in bloody revolution in the heart of a democracy with no end game in mind – this is our depraved state. But the reasons for why this malaise is the case – the time of the epoch – elude everyone in the present. That is the ultimate extension of, and not discontinuity from, the Sadean logic that failed to replace the forsakenness on the Cross.

Bibliography

- Blumemberg, Hans. 1985. *The Legitimacy of the Modern Age*. Cambridge: MIT Press.
- Foucault, Michel. 1978. *History of Sexuality*. Volume I. New York: Random House.
- Gutting, Gary. 2013. *Thinking the Impossible: French Philosophy Since 1960*. Oxford: Oxford University Press.

Rajesh Sampath

- Heidegger, Martin. 1962. *Being and Time*. New York: Harper and Row.
- Klossowski, Pierre. 1991. *Sade My Neighbor*. Evanston: Northwestern University Press.
- Koselleck, Reinhart. 1985. *Futures Past: On the Semantics of Historical Time*. Cambridge: MIT Press.
- Lane, William M. 1974. *The New International Commentary on the New Testament: The Gospel of Mark*. Grand Rapids, MI: William B. Erdmanns Publishing Company.
- Löwith, Karl. 1949. *Meaning in History: Theological Implications of the Philosophy of History*. Chicago: University of Chicago Press.
- New American Bible. 1987. Canada: World Bible Publishers.
- Pagels, Elaine. 1989. *The Gnostic Gospels*. New York: Vintage Press.
- Taylor, Charles. 2007. *A Secular Age*. Cambridge: Harvard University Press.

Information about Authors

Arnold Cusmariu has published in metaphysics, epistemology and more recently aesthetics. PDF files of his publications are available at www.academia.edu. Training in analytic philosophy (Ph.D., Brown University) proved invaluable when Dr. Cusmariu began making sculpture, enabling him to identify and then replace paradigm-level assumptions driving traditional figurative as well as abstract artworks. Stone is his preferred medium now, primarily alabaster and soapstone for their beautiful grain and color. A series with the musical title *Counterpoint*, inspired by Plato's metaphor of "interweaving forms," is currently at #32. Near completion is a logic book for children suitable for self-study, home schooling and classroom instruction. Contact: bravo323@gmail.com.

Michael F. Duggan holds a Ph.D. from Georgetown University in History and Philosophy. He has worked at the Supreme Court of the United States since 1991 and was the Supreme Court Fellow for 2011-2012. He has taught at Georgetown University and in New York University's Washington D.C. Program, and has guest lectured at Howard University Law School and at the American University. Contact: mikeduggan9@gmail.com.

Samson Liberman is lecturer of the Department of Social Philosophy at Kazan Federal University. He received his PhD in philosophy with a thesis titled social philosophy in a "post-post" situation: a methodological crisis and possible ways to resolve it (in Russian). His current teaching and research interests are social philosophy, philosophy of history, ontology of social, media philosophy, methodology of social research. Contact: samsonliberman@gmail.com.

Daniel Rönne is a researcher at Stockholm University, Department of Philosophy. Much of his research focuses on modal logic, including epistemic and doxastic logic, deontic logic and temporal logic. Some recent publications include: "The Moral Law and The Good in Temporal Modal Deontic Logic with Propositional Quantifiers," *Australasian Journal of Logic* (2020), "Boulesic logic, Deontic Logic and the Structure of a Perfectly Rational Will," *Organon F* (2020), "Boulesic-Doxastic Logic," *Australasian Journal of Logic* (2019), "Semantic Tableau Versions of Some Normal Modal Systems with Propositional Quantifiers," *Organon F* (2019), "Doxastic Logic: A New Approach," *Journal of Applied Non-Classical Logics* (2018), "Temporal Alethic Dyadic Deontic Logic and the Contrary-to-Duty Obligation Paradox," *Logic and Logical Philosophy* (2018), "Quantified Temporal Alethic Deontic Logic," *Logic and Logical Philosophy* (2015) and "Temporal

Symposion, 8, 1 (2021): 139-140

Symposion

Alethic-Deontic Logic and Semantic Tableaux," *Journal of Applied Logic* (2012).
Contact: daniel.ronnedal@philosophy.su.se.

Rajesh Sampath is currently Associate Professor of the Philosophy of Justice, Rights, and Social Change at Brandeis University. He completed his PhD at the University of California, Irvine in the humanities where he studied under the French philosopher Jacques Derrida, the founder of deconstruction. His areas of specialization center on the philosophy of history, historical time and epochal shifts. His most recent publications have focused on modern continental European philosophy, particularly, Hegel, Heidegger, Derrida and comparative philosophies of religion. In terms of applying continental philosophy to the social sciences, his research focuses on minority rights, theories of the state, sovereignty, and democracy. Contact: rsampath@brandeis.edu.

About the Journal

Symposion was published for the first time in 2003, as *Symposion – Revistă de științe socio-umane* (*Symposion – A Journal of Humanities*), with the purpose of providing a supportive space for academic communication, dialog, and debate, both intra and interdisciplinary, for philosophical humanities and social and political sciences. *Symposion: Theoretical and Applied Inquiries in Philosophy and Social Sciences* shares essentially the same purpose. Its main aim is to promote and provide open access to peer-reviewed, high quality contributions (articles, discussion notes, review essays or book reviews) in philosophy, other humanities disciplines, and social and political sciences connected with philosophy.

The old series published mainly Romanian papers. The new *Symposion* is an international journal, welcoming contributions from around the world written in English and French.

Although devoted especially to social philosophy and related disciplines (such as normative political theory, ethics, social epistemology, philosophy of culture, philosophy of technology, philosophy of economics, philosophy of education, and philosophy of law), the journal is open for original and innovative contributions in all philosophical fields and on all philosophy related topics from other humanities and social and political sciences. *Symposion* is also available for scholars developing interdisciplinary research, as long as it is philosophy related and/or it can open new approaches, pathways, or perspectives for (theoretical or applied) philosophical problems and philosophical thinking. Proposals for special issues devoted to a particular topic, theory or thinker are expressly invited.

The journal promotes all methods and traditions of philosophical analysis and inquiry (from ideal to critical or other types of non-ideal theory, from philosophical hermeneutics to logical and mathematical investigations of philosophical problems, from conceptual analysis to experimental philosophy, and from analytic to Continental philosophy). We also welcome papers on feminist philosophical (and philosophy related) topics, approaches or methods of inquiry.

From 2017, *Symposion* is published on a biannual basis, appearing at the end of May and November. It is published and financed by the “Gheorghe Zane” Institute for Economic and Social Research of The Romanian Academy, Iasi Branch. The publication is free of any fees or charges.

Author Guidelines

1. Accepted Submissions

The journal accepts for publication articles, discussion notes, review essays and book reviews.

Please submit your manuscripts and your proposals for special issues electronically at: symposion.journal@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.

There are no submission fees or page charges for our journal.

2. Publication Ethics

The journal accepts for publication papers submitted exclusively to *Symposion* and not published, in whole or substantial part, elsewhere. The submitted papers should be the author's own work. All (and only) persons who have a reasonable claim to authorship must be named as co-authors.

The papers suspected of plagiarism, self-plagiarism, redundant publications, unwarranted ('honorary') authorship, unwarranted citations, omitting relevant citations, citing sources that were not read, participation in citation groups (and/or other forms of scholarly misconduct) or the papers containing racist and sexist (or any other kind of offensive, abusive, defamatory, obscene or fraudulent) opinions will be rejected. The authors will be informed about the reasons of the rejection. The editors of *Symposion* reserve the right to take any other legitimate sanctions against the authors proven of scholarly misconduct (such as refusing all future submissions belonging to these authors).

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 and review essays must be no longer than 6000 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

Symposion

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. Authors should use the author-date system for text citations and Chicago style format for reference lists, as it is presented in *Chicago Manual of Style*. For details, please visit <http://library.williams.edu/citing/styles/chicago2.php>. 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 and French. 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 postal and email 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 last page of the paper.

8. Review Process

Symposion publishes standard submissions and invited papers. With the exception of invited contributions, all articles which pass the editorial review, will be subject to a strict double anonymous-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 submissions will be sent to at least two reviewers recognized as experts in their topics. The editors will take the necessary measures to assure that no conflict of interest is involved in the review process.

The review process is intended to be as quick as possible and to take no more than three months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors. Processing of papers in French may take longer.

The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Articles accepted will be published, as far as possible, in the order in which they are received and will appear in the journal in the alphabetical order of their authors.

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.

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.

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 working 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: *Symposion Journal*, 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. Copyright & Publishing Rights

The journal holds copyright and publishing rights under the terms listed by the CC BY-NC License. Authors have the right to use, reuse and build upon their papers for non-commercial purposes. They do not need to ask permission to re-publish

Symposion

their 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.

No manuscripts will be returned to their authors. The journal does not pay royalties.

14. Electronic Archives

The journal is archived on the Romanian Academy, Iasi Branch web site. The electronic archives of *Symposion* are also freely available on Philosophy Documentation Center, PhilPapers, Academia.edu, and CiteFactor.