SORITES

σωρίτης

An International Electronic Magazine of Analytical Philosophy
Indexed and Abstracted in ***THE PHILOSOPHER'S INDEX***
ISSN 1135-1349
Legal Deposit Registration: M 14867-1995

Editor: Lorenzo Peña
(Spanish Institute for Advanced Studies)
Associate Editor: Txetxu Ausín
(European University, Madrid)

Board of Editorial Consultants:
Jean-Yves Béziau, Enrique Alonso, Guillermo Hurtado, Manuel Liz,
Raymundo Morado, Raúl Orayen

Regular-Mail Address:
Prof. Lorenzo Peña
CSIC [Spanish Institute for Advanced Studies]
Department of Theoretical Philosophy
Pinar 25
E–28006 Madrid
Spain
Fax +3491 564 52 52
Voice Tph +3491 411 70 60, ext 18

INTERNET ACCESS:
<http://sorites.net/>
<http://www.ifs.csic.es/sorites/>
<http://www.filosoficas.unam.mx/~sorites>
Editorial e-mail inbox (esp. for submissions): <sorites@sorites.net>
Inquiries and subscription-requests: <sorites@sorites.org>

Issue #14 — October 2002

SORITES (ΣΩΡΙΤΗΣ)
ISSN 1135-1349
Issue #14 — October 2002
Copyright © by SORITES and the authors

MAIN INTERNET ACCESS:

<http://sorites.net/>
<http://sorites.org/>
<http://www.ifs.csic.es/sorites/>
<http://www.filosoficas.unam.mx/~sorites>
<sorites@sorites.net> (Editorial e-mail inbox, esp. for submissions)
<sorites@sorites.org> (Inquiries and subscription-requests)

SORITES

ISSN 1135-1349

ROLL OF REFEREES

Rainer Born . . . . . Johannes-Kepler Universitaet Linz (Austria)
Amedeo Conte . . . . . University of Pavia (Italy)
Newton C.A. da Costa . . . . . University of São Paulo (Brazil)
Marcelo Dascal . . . . . University of Tel Aviv (Israel)
Dorothy Edgington . . . . . Birbeck College (London, UK)
Graeme Forbes . . . . . Tulane University (New Orleans, Louisiana, USA)
Manuel García-Carpintero . . . . . University of Barcelona (Spain)
Laurence Goldstein . . . . . University of Hong Kong (Hong Kong)
Jorge Gracia . . . . . State University of New York, Buffalo (USA)
Nicholas Griffin . . . . . McMaster University (Hamilton, Ontario, Canada)
Rudolf Haller . . . . . Karl-Franzens-Universitaet Graz (Austria)
Terence Horgan . . . . . University of Memphis (Tennessee, USA)
Victoria Iturralde . . . . . Univ. of the Basque Country (San Sebastian, Spain)
Tomis E. Kapitan . . . . . Northern Illinois University (USA)
Manuel Liz . . . . . University of La Laguna (Canary Islands, Spain)
Peter Menzies . . . . . Australian National University (Canberra, Australia)
Carlos Moya . . . . . University of Valencia (Spain)
Kevin Mulligan . . . . . University of Geneva (Switzerland)
Jesús Padilla-Gálvez . . . . . Johannes-Kepler Universitaet Linz (Austria)
Philip Pettit . . . . . Australian National University (Canberra, Australia)
Graham Priest . . . . . University of Queensland (Brisbane, Australia)
Eduardo Rabossi . . . . . University of Buenos Aires (Argentina)
David-Hillel Ruben . . . . . School of Oriental and African Studies, University of London
Mark Sainsbury . . . . . King's College (London, UK)
Daniel Schulthess . . . . . University of Neuchâtel (Switzerland)
Peter Simons . . . . . University of Leeds (Leeds, UK)
Ernest Sosa . . . . . Brown University (Providence, Rhode Island, USA)
Friedrich Stadler . . . . . Institut «Wien Kreis» (Vienna, Austria)

SORITES

ISSN 1135-1349

Issue #14 — October 2002

TABLE OF CONTENTS

— Abstracts of the Papers 003

— «On the Analysis of Conditionals» by Simon Salzedo 007

— «Water, Phlogiston, Brains, and Vats» by Jussi Haukioja 016

— «Robots and *If...then*» by Ronald A. Cordero 021

— «A Dilemma for Robust Alethic Relativism» by William Ferraiolo . . 036

— «Can a Localist and Descriptive Epistemological Naturalism Avoid Dogmatic Foundations?» by Armando Cíntora 042

— «Cartesianism and the Private Language Argument» by Brian Garrett 057

— «A Trio on Truth» by Herbert Hrachovec 063

— «Futility and the Meaning of Life Debate» by Brooke Alan Trisel . . 070

— «On The Fourfold Root Of Philosophical Skepticism» by Mark Walker 085

— «Fodor's Epistemic Intuitions of Analyticity» by Wayne Wright 110

— «Wittgenstein: Transcendental Idealist?» by John M. Weyls 117

— Copyright Notice and Legal Disclaimer 122

— Release Notice 124

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 3-6
Abstracts of the Papers
Copyright © by SORITES and the authors

ABSTRACTS OF THE PAPERS

On the Analysis of Conditionals

Simon Salzedo

<Salzedo@brickcourt.co.uk>

Conditionals carry a conversational implicature that if their antecedant is unassertible, then their consequent is unassertible. The recognition of this implicature allows a single conversational analysis of counterfactual and other conditionals based upon the truth conditions of material implication.

Water, Phlogiston, Brains, and Vats

Jussi Haukioja

<jussi.haukioja@iki.fi>

Ted Warfield has presented a new version of the Putnamian argument for the conclusion that we are not brains in a vat. This version is intended to avoid reliance on some questionable background assumptions which other versions have made. It seems that Warfield's argument fails, for reasons pointed out by Anthony Brueckner. However, in this paper I present a new version of the argument — my version relies on assumptions no more objectionable than Warfield's, yet it is immune to Brueckner's objection.

Robots and *If…then*

Ronald A. Cordero

<cordero@uwosh.edu>

How shall we have robots handle conditional statements? In this paper I argue that we absolutely cannot let them use several of the presently accepted rules of inference involving conditional statements if we want to avoid odd, preposterous, or even disastrous results. I discuss several kinds of problems that could be encountered and suggest alterations to certain rules of inference to prevent such problems from arising.

A Dilemma for Robust Alethic Relativism

William Ferraiolo

<bferraiolo@sjdccd.cc.ca.us>

Robust alethic relativism is the thesis that no truth bearer is objectively true or false. According to the robust alethic relativist, the most we can ever say of any truth bearer (statement, belief, proposition, etc.) is that it is true or false *relative to* some conceptual framework, worldview, or other parameter (i.e. that it is «*true-for-X*»). In this paper, I will argue that robust alethic relativism is either self-refuting, or an entirely trivial and uninteresting thesis that cannot coherently serve as a theory of truth. I hope to show that Socrates understood this difficulty for the full-blooded relativist, and that his attack on alethic relativism is more effective than some have recognized.

Can a Localist and Descriptive Epistemological Naturalism Avoid Dogmatic Foundations?

Armando Cíntora

<cintora@prodigy.net.mx>

It is argued that epistemological naturalism is the result of a holist thesis plus a high valuation of empirical science. Epistemological naturalism criticizes the sceptic for entertaining unjustified global doubts and naturalism tries to avoid scepticism by taking for granted as non problematic our background scientific knowledge and by recommending only a localist or piecemealist mending of our corpus of knowledge, these corrections will be motivated by limited and justified questions.

It is argued that the epistemological naturalist:

i) Cannot justify without vicious circularity the most basic methods of science nor epistemological naturalism's localist recommendation.

ii) That if epistemological naturalism intends to be a description of genuine scientific methods then naturalism tacitly takes for granted, i.e., without justification, some epistemic norms.

iii) That natural science itself (evolutionary biology) produces traditional sceptic doubts, and therefore epistemological naturalism cannot avoid scepticism.

iv) That naturalism can neither avoid sceptic doubts by substituting an argumentative theory of justification with a reliabilist theory.

Cartesianism and the private language argument

Brian Garrett

<Brian.Garrett@anu.edu.au>

In this paper, I argue that neither the #257 argument nor the #258 argument in Wittgenstein's Philosophical Investigations undermines the coherence of the Cartesian Model, according to which a sensation word, such as 'headache' or 'tickle', gets its meaning in virtue of an act of 'inner' association or ostensive definition. In addition, I argue against the standard assumption that the diarist's language of #258 is logically private.

A Trio on Truth

Herbert Hrachovec

<Herbert.Hrachovec@univie.ac.at>

Truth is an embattled concept; many different positions have been put forward. One widely influential contribution has been Donald Davidson's theory. Although it has been derived from Alred Tarski's formal account of truth it has been claimed to offer a pragmatical solution to the problem by e.g. Richard Rorty. This dialogue explores the attraction Davidson's theory offers to philosophers of realist as well as relativist persuasion. There seems to be a core position useful to any of those philosophical schools: Truth occurs at the interface of two languages or two usages of a language. Some consequences of such a point of view are discursively explored.

Futility and the Meaning of Life Debate

Brooke Alan Trisel

<triselba@cs.com>

Are all human endeavors futile, as futilitarians contend? What does it mean when someone claims that «life is futile»? Although meaninglessness has been explored in great detail, the concept of futility, as used in the context of the debate about whether there is a «meaning of life,» has remained largely unexplored. Futility is a combination of the concepts of ordinary causation, failure, and repetition and is the opposite of effectiveness. Just as it would not make sense to claim that «life is effective,» it does not make sense when someone claims that «life is futile.» Life could be objectively futile only if there was an objective purpose of life, which there is no evidence thereof, and we were somehow failing to achieve this purpose. Striving to achieve a particular goal can be subjectively futile for an individual, but whether or not it is futile largely depends on how high an individual has set his or her expectations.

On The Fourfold Root Of Philosophical Skepticism

Mark Walker

<mark@markalanwalker.com>

Philosophical skepticism challenges us to demonstrate that knowledge is possible. Most often this challenge is made by questioning whether the attempts at justifying our epistemic claims are sufficient. In effect, then, the philosophical skeptic should be seen as arguing that knowledge is impossible because one of the necessary conditions for knowledge (justification) does not obtain. Some work in analytic epistemology suggests that knowledge has three additional necessary conditions, namely, that it must be the case that knowledge claims are believed, true, and that some additional concept obtains which rules out «Gettier-type» counter-examples. It is argued that if we accept that knowledge has three additional necessary conditions (in addition to the justification component) then this opens up the possibility for three additional types of philosophical skepticism. Skepticism based on the idea that our knowledge claims lack truth I term 'alethic skepticism'; skepticism based on the idea that the

belief condition does not obtain I term 'noetic skepticism'; and finally, I term 'gettier skepticism' the view that our knowledge claims do not rule-out Gettier-type counter-examples.

Fodor's Epistemic Intuitions of Analyticity

Wayne Wright

<wrightwt@comcast.net>

This paper argues that Jerry Fodor fails to adequately motivate his informational semantics because he does not exclude molecularism, a principal rival to his account of concepts. Supporting my position are Fodor's inability to explain away the strong intuitions often held on behalf of analyticity and his not offering a convincing argument for his claim that there is no way of making a principled analytic/synthetic distinction. Since he wishes to defend necessity and a prioricity, both of which are condemned by the Quinean anti-analyticity arguments, while denying analyticity, Fodor must provide an anti-analyticity argument of his own if his informational semantics is to be accepted. The result is that we have no reason to abandon the claim that there are meaning-constitutive interconceptual connections in favor of Fodor's atomistic informational semantics.

Wittgenstein: Transcendental Idealist?

John M. Weyls

<Weyls@juno.com>

In Jonathan Lear's and Barry Stroud's essay «The Disappearing We,» Lear presents Wittgenstein as transcendental idealist and parallels him with Kant. Stroud, while willing to grant some degree of Kantianness to Wittgenstein, is unwilling to press the parallel as far as Lear does. I will argue that both Lear's account of Wittgenstein as Kantian, and Stroud's objections as to the extent to which the parallel can be taken, are fraught with difficulties. I will attempt to show that the difficulties center on what I take to be Wittgenstein's paradoxical relationship with synthetic a priori judgments. If, like Kant, Wittgenstein holds them to undergird the sciences, then, contrary to what he maintains, he is not entitled to hold that concepts different from the ones we are used to are intelligible. On the other hand, if Wittgenstein rejects them and, consequently, their foundational status, he is committed to either one of two views, both of which he seems to reject — that mathematical statements are revisable in light of empirical facts, or that they are mere tautologies.

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 7-15
On the Analysis of Conditionals
Copyright © by SORITES and Simon Salzedo

ON THE ANALYSIS OF CONDITIONALS

Simon Salzedo
<Salzedo@brickcourt.co.uk>

I Introduction

By «conditionals,» I mean English assertions of the form «If…then…» and some assertions which can be easily rearranged into such a form. The analysis of conditionals is important beyond linguistics because conditionals are inextricable from any attempt to understand the force of the laws of science or nature which are supposed to predict what will happen in cases which are not determined by observation. Interesting accounts of concepts like causation and knowledge have also been given in terms of conditionals.[1]

There is some analogy between the conditional and the material implication (MI) which is defined by a two-valued, binary truth table to be false in the case of a true antecedent and a false consequent and to be true otherwise. Yet the assertibility of a conditional is not always coextensive with the truth of the equivalent MI. In particular, neither the falsity of an antecedent nor the truth of a consequent are sufficient to warrant the assertion of a conditional in which they feature, yet either guarantees the truth of such an MI. Such worries have lead many to despair of saying just what the link is between conditionals and material implications.

It has also been claimed that there are two types of conditionals each of which requires a different analysis. For example, it is said that

1) If Booth did not kill Lincoln somebody else did.

and 2) If Booth had not killed Lincoln somebody else would have.

are fundamentally different which can be judged from the fact that someone who asserted 1) could easily deny 2). Therefore, it is argued, we need analyses which will allow us to call 1) true while calling 2) false. Thus many have despaired of understanding conditionals as one type of assertion.

My aim in this essay is to overcome the dual despairs just described by defending the following claims:

1. (i) The assertibility conditions of all conditionals are explained by what Grice has called conversational implicature.[2]
   (ii) Combinations of times, tenses and moods indicate implicatures about the assertibility of the antecedent.
   (iii) These considerations show that conditionals can not simply be put into two classes, exemplified by 1) and 2) above but lie on a continuum between them.
2. The truth conditions of all conditionals are those of MI.
3. The objections which are often considered decisive against the Gricean approach to analysing conditionals can be answered.
4. The plausibility of the most popular analyses of conditionals is both explained by and improved upon by considering conversational implicatures.

---

[1] See Mackie [1] and Nozick [2].

[2] See Grice [3].

In defending these claims, I will speak of antecedents, consequents, conditionals and equivalent MIs despite misgivings which have been expressed about such terms.[3] My justification for using these terms is twofold. First, I do not think that there is any difficulty in recognising what is meant by them. Secondly, if a unified approach to conditionals using these terms is successful, then their worth will be proved. This second may be circular in its form, but the circularity is not vicious because analyses of this sort are useful or not useful rather than true or false.

II Assertibility Conditions

Following Grice, I hold that there is a 'Cooperative Principle' in conversation, that each contribution should be appropriately informative, true, backed by evidence, relevant, clear and perhaps other things. Where the application of this principle to what is said in some utterance entails some unsaid proposition, that unsaid proposition is a conversational implicature of the utterance. An important feature of conversational implicatures is that they may be cancelled by the speaker or by their context. Cancellable implicatures can also be attached to particular forms of utterance, which may be termed generalised conversational implicatures.

An utterance which breaks the Cooperative Principle, or which has false implicatures, is not assertible without some indication of its peculiarity. Most true sentences are not assertible most of the time because they are not relevant or not the most informative available. In any given situation, though, there will be many possible utterances which would not break any part of the Cooperative Principle and which are assertible.

Where all that the speaker wishes to convey is some assertion, a conditional with that assertion as the consequent or its negation as the antecedent will not normally be assertible. This is for the simple reason that neither the truth of the consequent nor the falsity of the antecedent follow from the conditional by normal truth-functional rules. Even though such a conditional is true, it is generally not assertible in this context because it is not appropriately informative. An exception to this generalisation demonstrates how it works: «If this is justice then I am a banana» succeeds in conveying the unassertibility of its antecedent because its consequent is known by all parties to the conversation to be not merely false, but chosen for the obviousness of its falsity.

Conditionals suggest some connection between antecedent and consequent. More specifically, I want to propose that «If A then C» carries a generalised conversational implicature of «If not A then not C». As an example, consider how «the cradle will rock anyway» can act as a qualification (which cancels the implicature) when conjoined with «if the wind blows the cradle will rock» or as an objection to the same sentence when not qualified.

Conditionals like 1) are asserted when it is appropriate within the conversation to discuss the consequences of the antecedent. Either the antecedent is believed to be true or supposed to be so for the purposes of a reductio. Such conditionals accordingly implicate that their antecedents are assertible. Consider the contrast between «If Booth killed Lincoln, and I think he did,…» and «If Booth killed Lincoln, though I am not saying that he did,…». The conjunctions in these and many similar sentences are not comfortably interchanged because «though» or «but» are only appropriate for indicating the cancellation of an implicature or another change of course. This implicature is easily cancelled. For example, a heavy stress on the «if» in the above sentences may be enough to remove it, allowing «*If* Booth killed Lincoln, and I am not saying that he did,…»

Conditionals like 2), by contrast, implicate that their antecedents are not assertible. 2) is assertible in situations where there is an assumption in the conversation that Booth did kill

---

[3] See Dudman [4] and [5].

Lincoln. This has been disputed by, for example, Appiah (see Appiah [7] p164), who suggests that «Had the butler done it, there would have been blood in the pantry» can be part of an argument that the butler did it. But even here there is a presumption of the butler's innocence, the implicature of which is cancelled by a later stage of the argument. In isolation, the sentence suggests that the butler did not do it and that there was no blood in the pantry. What is missing from Appiah's argument is the conversational context for his example.

In deciding whether a conditional like 1) or one like 2) is more appropriate, what is important is not the speaker's belief regarding the antecedent's truth, but his belief about its assertibility in the given context. Often, the two will coincide; but where the speaker is seeking to overthrow a presumption of his listeners, he may adopt that presumption to present an argument by reductio. This is the most likely context for Appiah's example.

More generally, when a conditional «If A then C» is asserted, some presumption about A's assertibility has been adopted. The tense or mood words in the conditional indicate which presumption is in force, that is, whether A is or is not assertible in that conversational context. Thus, «The butler did not do it. If the butler had done it, there would have been blood in the pantry.» is a more natural utterance than «The butler did it. If the butler had done it, there would have been blood in the pantry.» The latter utterance demands the appendage «and there *was* blood in the pantry» to cancel the implicature of its second sentence and thus resolve that sentence's conflict with its first. The former utterance stands alone without confusion.

The implicature regarding the assertibility of «Booth killed Lincoln» is, I have argued, the most important difference between conditionals like 1) and those like 2). The question of taxonomy should therefore be approached by considering that implicature. Such an approach, restricted to the generalised implicatures of the different forms of conditional, reveals a continuum of tense and mood combinations. Examples of the most discussed such combinations might be:

3) If the wind blew then the bough broke.
4) If the wind blew yesterday then the bough will break tomorrow.
5) If the wind is blowing then the bough will soon break.
6) If the wind blows tomorrow then the bough will break.
7) If the wind blows then the bough breaks.
8) If the wind had blown then the bough would have broken.

In the same way as the implicature of 1) of the assertibility of its antecedent was demonstrated by considering the most comfortable conjunctions for conjoining to the conditional its antecedent or its negation, we can show that 3),4) and 5) all implicate the same thing. Though they do so decreasingly strongly.

6) is more or less neutral, as future events can be discussed without any presumption about whether they will actually occur. Since the first part of 7) is the same as that of 6), the test we are using reveals the same neutrality. This is reasonable because conversations about general rules normally presume that the antecedent is true in some but not all instances which are being considered in the conversation.

If 6) lies between 3) and 8) in the sense that I have suggested then the debate about with which of the two groups it should be placed can be seen to be misguided. The debatability of the point is, however, not surprising.

In this taxonomy, the implicature is decisive, and the generalised implicature of an isolated sentence may be overridden by elements of the context. And the neutral 6) may be used when an assumption about the antecedent's assertibility is determined by its context. The tense and mood words are not themselves definitive of the type of conditional; they merely indicate it in the absence of other signals.

III Truth Conditions

For formal purposes, the conditional is translated into an MI with a defined truth table; the truth of the antecedent and consequent are assessed and the truth value of the MI falls out from the truth table.

In non-formal settings a hearer who questions an assertion questions its assertibility. This may be on any of several grounds two of which are that the speaker does not or should not believe it to be true. That is, the speaker is lying or is mistaken. But while falsity leads to non-assertibility, truth does not lead to assertibility; put another way, truth is a necessary but not sufficient condition of assertibility.

A different ground for questioning an assertion is that it is misleading. An assertion may be misleading if it makes conversational implicatures which are false. In this way, 2) is often misleading when asserted by a speaker who believes that Booth did not kill Lincoln because it implicates, in the absence of other information, that Booth did kill Lincoln.

When we do wish to assess the truth of a conditional in an informal context, something very similar to the formal case takes place so long as the antecedent is presumed to be true. Thus, if Booth did not in fact kill Lincoln, then both 1) and 2) share the truth value of «somebody else killed Lincoln».

If, however, Booth did kill Lincoln, then it is less obvious what would determine the truth of 1) and 2). We can imagine two people who disagree about the truth of 1) or of 2). Then suppose that the falsity of «Booth did not kill Lincoln» was established to the satisfaction of both of them. Now their debate over 1) or 2) would take on a different character. It is about opinions and no longer about facts. They agree that Booth killed Lincoln and so they agree that somebody else did not. Any remaining argument is not so much over the truth of 1) or 2), but over the evidence for those sentences. The fact that such a debate is possible is, therefore, not an objection to the claim that all conditionals with false antecedents are true. In the same way, we can stipulate that all conditionals with true consequents are true without thereby closing off the chance of understanding why only some of them are assertible. The line drawn here between opinion and fact may seem arbitrary. But there are distinctions to be made between misleading, misjudging and lying which alternative schemes would struggle to accommodate.

IV Objections and Cases

As noted above (Section I), it is often claimed that conditionals like 1) and 2) are different in some fundamental other than an implicature about their antecedent.[4] But the differences between 1) and 2) are due to the different attitude towards «Booth did not kill Lincoln» of which the implicature is an expression.

With 1), the conversation is on the basis that in the actual world Booth did not kill Lincoln. We can then consider the fact as to whether or not someone else killed Lincoln. This fact about the past determines the truth of the conditional. In 2), there is a presumption that Booth did kill Lincoln. We must then consider the question of how things would be different for other people killing Lincoln if they had been different for Booth doing so. There is no fact about the past or present (or even future) available to help us here and we do not look for one. Where the antecedent is presumed false, we almost invariably lose interest in the truth of the conditional. We swap truth for other standards of assessment and the reason for the change is our assumption of the unassertibility of the antecedent.

There is a further question which may be raised here about whether conditionals like 2) have truth conditions at all. An approach which allows some other elements than truth in assertibility is essential both to explain the non-assertibility of statements whose truth no one would deny (many theorems of mathematics or formal logic, for example, are assertible in

---

[4] See, for example, Lewis [6] p3, and Appiah [7] pp 164-5.

only a tiny number of situations) and to explain the wrongness of misleading someone without telling a direct lie. Yet we cannot ignore the fact that truth-conditional falsity always bars assertibility. «Booth did not kill Lincoln, and neither did anyone else» is a clear denial of 2) just because it entails the falsity of 2). This suggests ascribing to all conditionals the truth conditions of MI which can be seen as necessary but not sufficient for their assertibility. That the implicature of a false antecedent carried by some conditionals is enough to direct the hearer to considerations other than truth does not mean that there is no use for truth conditions. First, if the antecedent of a conditional like 2) turns out to be true, then the truth of the conditional is decided in the usual, truth functional way. Secondly, in formalising we are always interested in truth, and it would be churlish to refuse to allow this notion of truth to decide the matter when there is no conflicting interest.

Adams[5] points out that there are truth-functionally valid inferences of which the premises may be assertible while the conclusion, a conditional, is not. According to Adams, his examples are not convincingly explained by conversational implicature. Based on the account given above, Adams' examples admit of the following explanations.

(i) Adams takes two sentences, A and B:
A. It will not rain in Berkeley next year.
B. It will rain in Berkeley next year.
He says that we would not infer «If A then B» from not A or from B. In this case, A and B entail each other's negation. There is on my version of the conversational theory a particular objection to the assertion of «If A then not A» which is that it has an implicature of «If not A then not not A» which entails A which is false. This is on top of the question of why we should want to make such an inference. Outside a book of logic puzzles it is hard to imagine being interested in the question of whether to accept that conditional or its negation beyond deciding between A and not A.

(ii) Adams says that for the following A and B we might assert «If B then not A» but would not infer the contraposition, «If A then not B»:
A. There will be a terrific cloudburst tomorrow.
B. It will rain tomorrow.
There is here an additional premise, «If A then B,» which is implicit in any discussion of cloudbursts. If we do assert «If B then not A» we mean to convey «not A» and this is done by the normal truth-functional rules according to which «not A» follows from the two premises. Given «not A», we would not assert «If A then not B» despite its literal truth because the presumption of not A leads us to consider other things than truth. We have no interest in making purely logical deductions about the consequences of false antecedents. In these cases we move on to consider other evidence than facts and other standards of assessment than truth.

(iii) Adams says that we do not reason from «A or B» to «If not A then B» with the following A,B:
A. It will rain in Berkeley next year.
B. It will snow in Berkeley next year.
The implicature of connection which comes with the conditional is one reason why not. «If not A then B» implicates «If A then not B» and so we have «A or B but not both» which is most likely to be unassertible in this case. A related reason is the implicit premise «A» which turns «If not A then B» into a conditional the truth of which derives from the falsity of its antecedent and so is not a good standard of assessment.

(iv) Adams has two examples of hypothetical syllogism. In the first these sentences are used:

---

[5] See Adams [8] Ch1.

A. Smith will die before the election.
B. Jones will win the election.
C. Smith will retire after the election.

Adams notes that we do not move from the premises «If A then B» and «If B then C» to «If A then C». The reason is simple: the two premises are not assertible together in any one context. Once the question of Smith dying before the election is raised, we would not assert «If B then C» until we had decided that «not A» was assertible. The high probability of each of the premises is not an issue in deciding whether to assert them. In conversation, the important considerations for judging conditionals for which the facts are not (yet) available are the ones which are raised in the conversation or are taken as read by the parties.

In his other example, Adams has

A. Jones will study.
B. Jones will pass.
C. Jones will graduate.

Here we can move from «If A then B» and «If B then C» to «If A then C». The reason is not Adams', which is that the first premise is implicit in the second. It is that there is no conflict here. The truth functional rules do apply when there are no inconvenient implicatures or antecedents presumed to be false.

(v) Finally, with A,B,C as for the Smith and Jones election example, Adams says that we cannot reason from «If (A or B) then C» to «If A then C», noting that the premise is, although highly probable, intuitively absurd. This absurdity is what Adams claims cannot be explained by implicature.

But on the conversational theory, high probability is not even a prima facie reason for assertion. «If (A or B) then C», when «If A then not C» is assertible, is odd because it is uninformative with no commensurate gain in clarity or brevity. What is meant is «If B then C» and if we do not introduce probability, then no urge to assert «If (A or B) then C» emerges.

An objection to such explanations is that we do not always assert the stronger, or more informative, of particular options. Jackson, for example, says that «If the sun goes out of existence in ten minutes time, then the earth will be plunged into darkness in about eighteen minutes time» is assertible despite being only marginally more probable than the negation of its antecedent which would be stronger[6]. But we can consider the consequences of antecedents which we take to be false regardless of how probable their falsity is. And in any case, with the implicature of connection between consequent and antecedent, a conditional is not always less informative, even though it may be logically weaker, than the negation of its antecedent. The cases where a conditional is unassertible due to being uninformative are those where the implicatures which go with the conditional are not appropriate, perhaps because the negation of the antecedent is all that is meant. For Jackson's example, there will be some occasions on which the conditional will be assertible, and others on which the negation of the antecedent will be. What decides between these types of occasion is the content of what is being discussed. Assertions should be relevant as well as informative.

Appiah ([7] Ch8) also adduces an example against the principle of asserting what is more informative rather than what is less. He suggests that I can say «John or Mary will arrive soon» while believing that it is far more likely to be John. The circumstances of this utterance are that either one of the couple is required for a meeting. While the assertion is logically weaker than «John will arrive soon», it is more relevant because the interest is in the arrival

---

6 Jackson [9] pp 566-7.

of «John or Mary» as an entity. In another context in which John and Mary have different roles in the meeting, Appiah's sentence would be misleading due to being uninformative, because it would then be relevant to consider whether or not John will arrive soon.

Other conditionals which have been thought problematic for a unified conversational account are those like «Your essay is a bit short this week, if you don't mind me saying so» or «If you are hungry, there are biscuits on the table.» These illustrate well that English conditionals deal in assertibility and not just truth. If you do mind or are not hungry, then I would not assert the respective consequents of those conditionals. I do not withdraw my belief in or assent to the consequents, only my assertion of them.

V Other Accounts

The account I have given of conditionals may be criticised for vagueness. Other accounts like those of Adams or Lewis give more specific criteria for the assertibility of conditionals in terms of probability and spheres of possible worlds respectively.

Adams suggests that inferences should be assessed by the principle that it should be impossible for the premises to be probable and the conclusion to be improbable. The probability of a conditional, «if A then C», is held to be the probability of C given A, p(C/A), which is p(A & C)/p(A). Where this probability is high, broadly speaking, the conditional is assertible. This account is only meant to apply to conditionals like 1).

The attraction of this account is due to its attempt to make clearer the nature of the connection between antecedent and consequent which is indicated by the use of a conditional. It makes the connection into something within the grasp of logicians. However, it takes it out of the grasp of the users of English conditionals. Competent speakers do assess assertions for relevance, informativeness and support by appropriate evidence (even though they may be unable to describe these practices of assessment in much detail), but hardly ever for high probability. The reason why they do not poses another problem for a probabilistic scheme. In so far as the account is restricted to conditionals about past events, all probabilities involved are 1 or 0, and the criteria do not diverge from truth-conditional criteria. In so far as we are considering future events or general laws or counterfactual events, it is hard to see how the relevant probabilities could be calculated short of providing a complete description of the universe at the relevant time along with a probabilistic theory of changes to it through time. It might be neater if our practice of assessing evidence was based on measuring or intuiting probability, but for most purposes it is not plausible to say that it is.

Lewis suggests that conditionals like 2) are true just when the relevant MI holds at all of the nearest possible worlds to the actual world. This is a plausible working out of the circumstances in which these conditionals are assertible because it retains the vagueness of the conversational view. It seems attractive because it provides a formal framework for their manipulation which avoids the perception of awkwardness involved in counting all conditionals like 2) as true so long as their antecedents are false.

The central difficulty with possible worlds is that it is no easy matter to specify which worlds are closer. The nearest world in which Booth did not kill Lincoln is that in which what is changed from the actual world is just that which the parties to the conversation recognise as needing to be changed. We do not have to build in our minds a fully working, consistent model of a possible world. We just notice those aspects which interest us for the moment. Attributes of the possible worlds are therefore inescapably context relative. Closeness of worlds can be felt within a conversation, but cannot be objectively measured by logicians outside it. For this reason, possible worlds provide only a framework for analysis, and not much content.

Our ability to use conditionals with impossible antecedents is an additional problem for Lewis's approach. We can use sentences like «If Godel's Incompleteness Theorems had been false, then formalism would have been more attractive». This use is not explicable within

Lewis's framework, which must be limited to conditionals with antecedents presumed to be false, but which are possible.

There is also a danger of a regress in analysing conditionals by possible worlds. Full descriptions of worlds could not be given without using conditionals like 2) which bring in a new infinity of possible worlds and so more conditionals and so on.

VI Conclusion

Assertibility is best viewed as being a property of an utterance by a particular person in the context of a particular conversation. It should be ascribed to conditionals which do not breach the Cooperative Principle and which do not carry uncancelled implicatures which are unassertible. Truth is one of the requirements of the cooperative principle and is ascribed to each conditional as it would be for its equivalent MI.

The conditionals which are relevantly like 2) are those the antecedents of which are assumed in the conversation to be false. Because of this assumption, there is a special problem in saying what kind of evidence is appropriate for supporting such conditionals as there is no fact which could decide them. The solution to this problem would be a description of our practice of assessing such statements for support, coherence or plausibility. This practice does not reduce to a formal analysis and it is therefore not possible to use a description of it to replace the MI notion of truth. The attempt is also misguided because assertibility depends on other factors as well as truth.

I have not given the full content of an analysis of conditionals here. What I have tried to do is to suggest that the Gricean conversational framework is adequate to contain such an analysis and that the temptation towards more precise, but restricted, analyses should be resisted. All conditionals, like other assertions, must adhere to the Cooperative Principle and this is the basis for their assertibility. The unusual implicature of the conditional form is of a connection between the antecedent and consequent. I have suggested that this connection should be specified as being that «If A then C» implicates «If not A then not C».

References

[1] JL Mackie, *The Cement of the Universe*, (Oxford: Oxford University Press, 1974)
[2] Robert Nozick, *Philosophical Explanations*, (Oxford: Oxford University Press, 1981)
[3] HP Grice, 'Logic and Conversation', *Syntax and Semantics, Volume 3*, (London: Academic Press, 1975) 41-58
[4] VH Dudman, 'Indicative and subjunctive', *Analysis* 48 (1988) 113-22
[5] VH Dudman, 'Grammar, Semantics and Conditionals', *Analysis* 50 (1990) 214-224
[6] David Lewis, *Counterfactuals*, (Oxford: Basil Blackwell, 1973)
[7] Anthony Appiah, *Assertion and Conditionals*, (Cambridge: Cambridge University Press, 1985)
[8] Ernest Adams, *The Logic of Conditionals*, (Dordrecht: Reidel, 1975)
[9] Frank Jackson, 'On Assertion and Indicative Conditionals', *Philosophical Review* 88 (1979) 565-89

Simon Salzedo
<Salzedo@brickcourt.co.uk>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 16-20
Water, Phlogiston, Brains, and Vats
Copyright © by SORITES and Jussi Haukioja

WATER, PHLOGISTON, BRAINS, AND VATS

Jussi Haukioja

<jussi.haukioja@iki.fi>

Ted A. Warfield (1998) has presented a new version of the Putnamian[1] argument for the conclusion that we are not brains in a vat (henceforth BIVs).[2] Warfield's version relies on widely held views about content externalism and self-knowledge, but he intends his version to avoid reliance on other, more questionable background assumptions which other versions have made. Anthony Brueckner (2001) claims, however, that Warfield's version fails. In this paper I will argue that, even if Brueckner's reply succeeds, it leaves room for a new version of the argument. The new version relies on assumptions no more objectionable than Warfield's, yet it is immune to Brueckner's objection.

Warfield's version of the argument goes as follows (Warfield, 1998, p. 129):

(A) I think that water is wet. [Self-knowledge]

(B) No brain in a vat in an otherwise empty world can think that water is wet. [Externalism about content]

(C) So, I am not a brain in a vat in an otherwise empty world. [A, B]

This kind of a version of the BIV argument was, as far as I know, first suggested by Thomas Tymoczko (1989).

Warfield's version avoids reliance on two questionable assumptions. Firstly, we need not assume the implausibly strong externalist thesis that «one can think about Xs only if one has had causal contact with instances of X.» (Warfield, 1998, p. 130) Here Warfield's version departs from what seems to be Putnam's own most recent formulation (Putnam, 1992). It is easy to see that this externalist thesis is implausible — otherwise we could prove that we have had causal contact with unicorns or phlogiston, on the basis of the apparent fact that we can have thoughts about them. Since we can have thoughts about entities with which we have not had causal contact, (B) does not follow from the mere fact that there is no water in the BIV world. Warfield (1998, p. 131) presents two ways in which one might think about water without having interacted with water. First, one might theoretically construct the concept of water from the concepts of hydrogen and oxygen. Second, one might defer to others in one's

---

[1] Putnam, 1981, Chapter 1.

[2] The sceptical scenario under consideration here is one according to which all sentient beings are permanently brains in a vat, hooked to a computer which just happens to produce in them sensations of exactly the kind we experience. To simplify the discussion, we will further assume that there is nothing else in the world, and that there is no designer of this setup — it just happened to pop into existence. Like Warfield and Brueckner, I will restrict my attention to this case, and ignore other sceptical scenarios (such as cases of recent and/or temporary envatment).

language community. But Warfield claims that neither of these ways is available to the BIV — first, we are assuming that there is no hydrogen or oxygen in the BIV world; second, there are no experts on 'water' for the BIV to defer to. So (B) seems to follow without reliance on the implausible principle noted above.

Secondly, Warfield avoids taking a stand on the question of what, if anything, the BIV's words or concepts refer to. Many versions of Putnam's argument rely on the assumption that, for example, the term 'water' (or the corresponding concept), as used by the BIV, refers to the computer states responsible for its 'watery' experiences, or to phenomenal images.[3] Warfield's version only assumes that the BIV's water-thoughts are *not* about *water*. And this much, Warfield claims, follows from the externalist considerations above.

Brueckner objects that, in assuming (B), we are assuming that we are not in the same predicament with 'water' as we were with 'phlogiston':

> In defending [B] without reliance upon any problematic existence assumptions, I must countenance the possibility that my term 'water' is on a par with my 'phlogiston'. But if there is no water and 'water' thus fails to refer to any existing natural kind, then nothing in content externalist theory allows me to argue that a brain in a vat is barred from thinking that water is wet — barred from thinking exactly what I am now thinking via my sentence 'Water is wet'. (Brueckner, 2001, pp. 111-112)

Hence, according to Brueckner, Warfield is covertly assuming that water exists.

One might disagree with Brueckner about whether the BIV and I *would* think the same thought via 'Water is wet', were 'water' to fail to refer. In other words, one might try to respond to Brueckner by arguing that even if *both* our 'water' *and* the BIV's 'water' fail to refer, it does not self-evidently follow that our water-thoughts and the BIV's water-thoughts have identical contents, and count as the same thought.

However, I will not attempt such an argument here. Rather, I will suggest a new argument which, like Warfield's, relies on neither of the two assumptions noted above, but is immune to Brueckner's objection against Warfield. My version does, however, rely on the following equivalence principle:

(EP) If there are Ps, my concept of P refers to Ps.

(EP), it seems to me, can be known to be true *a priori* by any thinker. The first person formulation is crucial: I can know that, if there are Ps, *my* concept of P refers to them. But, of course, the externalism supported by Twin Earth thought experiments entails that *someone else's* qualitatively identical concept may in fact *not* refer to Ps, if her environment is different.

Warfield (1998, p. 134) considers a principle closely related to (EP), as a possible premise in anti-BIV arguments, but concludes that such an argument will be «difficult to provide» (*ibid.*). However, here is such an argument:

---

[3] As it happens, I find it plausible to claim that the BIV term 'water' *does* refer to states of the computer, and that an anti-BIV argument could be based on this assumption (see Tymoczko, 1990 for what I take to be the most promising argument of this kind). Warfield (1998, p. 131) objects to this line of reasoning: «I am not willing to argue that 'water' in my language refers to water while 'water' in the language of a [BIV] does not because […] this would require me to offer a priori reasons for thinking that my 'water' thoughts are referential.» This, however, misses the point. All that this form of the argument would require is that the BIV's 'water' refer to computer states, and that *our* 'water' *not* refer to computer states.

(1) I have the concept VAT[4]. [Self-Knowledge]

(2) Either there are vats or there are no vats. [Law of Excluded Middle]

(3) If there are vats, my VAT refers to vats. [EP]

(4) If there are vats, the BIV's VAT does not refer to vats. [Externalism]

(5) If there are vats, I am not a BIV. [1, 3, 4]

(6) If there are no vats, there are no BIVs. [Definition of 'BIV']

(7) If there are no BIVs, I am not a BIV. [Universal Instantiation][5]

(8) If there are no vats, I am not a BIV. [6, 7]

(9) I am not a BIV. [2, 5, 8]

(4) seems to follow from the same line of reasoning as Warfield gave for (B). The BIVs, we are assuming, have not had causal interaction with vats. Were they to be able to refer to vats, the reference would have to arise in some other, indirect way. Clearly deference to other members is out of the question. Could the BIVs VAT come to refer to vats if they constructed it from other concepts? For example, could they not refer to the vat in which they are with the definition «the container in which I am»? Maybe they could, maybe they could not. But note that even if they could, and their VAT were associated with this definition, that would mean that *their* VAT would not be qualitatively identical to *our* VAT — for our VAT is not associated with such a definition. Our VAT refers to the kind of thing we *have* had causal interactions with (assuming, as we do in the antecedent of premiss 4, that vats exist).[6]

In this argument, it is crucial that we are discussing vats and not, for instance, water — there is no suitable analogue of (6) for water. For a version which employs WATER to work, we would need to show that the BIV's WATER denotes *something* — most plausibly, the states of the computer to which the BIV is connected. But this would be to diverge from Warfield's original project of not relying on such assumptions. By stating the argument in terms of VAT instead of WATER, we avoid Brueckner's objection. *If* our VAT fails to refer, that can only be because there are no vats, and hence no BIVs; if it *does* refer, the externalist considerations above show that the BIV's VAT cannot refer to vats. Either way, we can know that we are not brains in a vat (of the Putnamian kind).

Finally, it might be objected, have we not just described a counter-example to the equivalence principle (EP)? Isn't the BIV scenario one in which there are vats, but VAT does not refer to vats? This objection would miss the importance of the first-person formulation of (EP). The principle holds of *our* VAT in the BIV scenario, because the concept refers to all and only the vats we are assuming to exist. On the other hand, the BIVs would also be correct in believing (EP) — that is, they would be *correct* in thinking 'If there are vats, my concept of vats refers to vats', because *none* of the occurrences of 'vat' in that sentence would refer to the vat in which they are.

---

[4] I use small capitals to refer to concepts.

[5] To be precise, use of Universal Instantiation here requires the further premise 'I exist'.

[6] Our BRAIN may be different in this respect, and perhaps the same argument could not be run on BRAIN. One might claim that one's BRAIN is just the concept of «that organ in which my thought processes occur», although I have doubts about whether the BIV's concept ORGAN would count as the same concept as ours.

My argument does not establish that our VAT in fact *does* refer, or that *any* of our concepts refer. But, if sound, it does prove that our thoughts about BIVs of the kind discussed here are not *about us*. Of course, one might also take this to be an argument against the compatibility of self-knowledge and content externalism and claim that, if the conjunction of these theses enables us to deduce that we are not brains in a vat, the conjunction must be false.[7] However, here I echo Warfield's (1998, pp. 137-140) sentiments. The argument given here is not a *general* refutation of sceptical scenarios, but a highly restricted one. I grant that it may be surprising that such an argument can be given. But if one feels that the premisses have considerable independent plausibility (as many do), the possibility of an argument of the kind given here should not lead one to abandon them.[8]

References

Boghossian, P.A. (1997): 'What the Externalist Can Know a Priori', *Proceedings of the Aristotelian Society* 97, 161-175.

Brueckner, A. (2001): 'A Priori Knowledge of the World Is Not Easily Available', *Philosophical Studies* 104, 109-114.

McKinsey, M. (1991): 'Anti-Individualism and Privileged Access', *Analysis* 51, 9-16.

Putnam, H. (1981): *Reason, Truth and History*, Cambridge: Cambridge University Press.

Putnam, H. (1992): 'Replies', *Philosophical Topics* 20: *The Philosophy of Hilary Putnam* (ed. C.S. Hill), 347-408.

Tymoczko, T. (1989): 'In Defense of Putnam's Brains', *Philosophical Studies* 57, 281-297.

Tymoczko, T. (1990): 'Brains Don't Lie: They Don't Even Make Many Mistakes', in M.D. Roth and G. Ross (eds.): *Doubting: Contemporary Perspectives on Scepticism*, Dordrecht, Kluwer, 195-213.

Warfield, T.A. (1998): 'A Priori Knowledge of the World: Knowing the World by Knowing Our Minds', *Philosophical Studies* 92, 127-147.

Jussi Haukioja
Department of Philosophy. University of Turku
FI-20014 Turku
<jussi.haukioja@iki.fi>

---

[7] This would be to treat the argument as an instance of the so-called 'McKinsey recipe' (McKinsey, 1991; Boghossian, 1997). One should note, however, that the purported conclusion of these arguments is that one could know *a priori* that, for example, water exists. I agree that such facts cannot be known *a priori*. But the conclusion of the anti-BIV argument is considerably weaker.

[8] This paper was written while I was visiting the Research School of Social Sciences (RSSS) at the Australian National University. I wish to thank the RSSS, the Alfred Kordelin Foundation, and the Academy of Finland for making the visit possible. I am also grateful to the following people for comments: Lisa Bortolotti, Philippe Chuard, Nic Damnjanovic, and Stewart Saunders.

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 21-35
Robots and if … then
Copyright © by SORITES and Ronald A. Cordero

ROBOTS AND *IF … THEN*

Ronald A. Cordero

<cordero@uwosh.edu>

0. Prefatory Note

For the sake of simplicity in this paper I am going to use expressions of cognitive process and propositional attitude, such as «reason,» «know,» «suspect,» «remember,» and «think,» without elaboration in referring to robots. In doing so, I do not mean to be taking a position on the question of whether or not robots can actually reason, think, remember, suspect, or know in the same way humans can. I simply wish to avoid the complexity of saying, for example, that a robot «believes$_R$,» where «believes$_R$» is defined as «is in the robotic state which corresponds to the state of mind of a human who believes.» I do think the question of whether or not entities like robots or non-human animals can have propositional attitudes and engage in cognitive processes is an interesting one, but I do not think that it has to be answered before issues regarding the use of «if…then» by robots can be resolved, and I do not propose to try to answer it here.

1. Introduction

We are going to have robots that reason. The scientific, commercial , and military motives for building them and letting them work with some degree of autonomy are going to prove irresistible. But this means that if we want to avoid awkward or even disastrous consequences, we must program these robots to make only valid, humanly intelligible inferences. And this may not be a simple thing to do where «if…then» is concerned. Conditional («if…then») statements are extremely important in human communication, but their analysis has long been a source of disagreement among logicians. In the present paper, I argue that unless we wish to court disaster, we cannot let our robots use the most widely accepted rules of inference concerning conditional statements and I propose alternative rules that I believe will stave off disaster.

There is at present a long list of commonly accepted «rules of inference» for what is commonly called «statement logic.» These «rules» are in essence analytically true statements of entailment relations between statements that contain expressions such as «not,» «or,» «and,» «if…then,» and «if and only if.» Because these analytically true statements are statements of fundamental truths of logic, I shall refer to them as axioms. Some statement-logic axioms express one-way entailment relations, saying that statements of one sort entail statements of another sort. Others express two-way entailment relations, saying that statements of a first sort entail and are entailed by — and thus are equivalent to — statements of a second sort. If, as is usually done, pairs of similar statement-logic axioms are counted as a single axiom, the total number of commonly accepted axioms is eighteen. If each axiom were to be counted separately, the total number would be twenty-four. Fewer than half of these axioms involve conditional statements. Using a single-headed double-barred arrow (⇒) to indicate one-way

entailment and a double-headed double-barred arrow (⇔) to indicate two-way entailment, those which do can be represented as follows:

Modus Ponens	$((p \supset q) \wedge p) \Rightarrow q$
Modus Tollens	$((p \supset q) \wedge \sim q) \Rightarrow \sim p$
Hypothetical Syllogism	$((p \supset q) \wedge (q \supset r)) \Rightarrow (p \supset r)$
Constructive Dilemma	$(((p \supset q) \wedge (r \supset s)) \wedge (p \vee r)) \Rightarrow (q \vee s)$
Transposition	$(p \supset q) \Leftrightarrow (\sim q \supset \sim p)$
Material Implication	$(p \supset q) \Leftrightarrow (\sim p \vee q)$
Material Equivalence	$(p \equiv q) \Leftrightarrow ((p \supset q) \wedge (q \supset p))$
	$(p \equiv q) \Leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$[1]
Exportation	$((p \wedge q) \supset r) \Leftrightarrow (p \supset (q \supset r))$

If our robots are not to make serious mistakes in reasoning with conditional statements, some of these axioms will have to be given to them in altered forms. In the following sections I shall illustrate the sort of problems that will arise if robots are allowed to use all these axioms without alterations — and shall suggest alternative axioms that can preclude the problems. In a sense, this will involve putting certain axioms «off limits» to robots, establishing what might be called a list of «forbidden inferences.»

2. Material Implication

Among the classic statement-logic axioms, one real potential trouble maker is the one commonly known as «Material Implication.» In its traditional form it encapsulates an analysis of conditional statements that has been widely used by logicians since early in the twentieth century — and which in fact goes back to Philo of Megara in the fourth century BC.[2] According to this analysis, when we say «If a, then b,» we are basically saying that either a is false or b is true. A conditional statement on this analysis, that is, is equivalent to the disjunction of the consequent with the negation of the antecedent. This core meaning has been held to be common to the different types of «if…then» statements,[3] and it is this meaning that the horseshoe (⊃) has been used to symbolize. In rendering «If a, then b,» as «$a \supset b$» , we are taking the statement to mean $\sim a \vee b$.

Is there any reason why robots should not be allowed to use the Material Implication axiom in its traditional form? The answer is definitely «Yes»: if they are allowed to do so, their reasoning simply cannot be trusted. And there is not just one way in which the axiom can cause problems: it can generate unacceptable inferences in a variety of ways.

For one thing, application of Material Implication in conjunction with the axiom called «Addition» ($a \Rightarrow (a \vee b)$) can generate the infamous «paradoxes of material implication» —

---

[1] This second Equivalence axiom effectively involves conditionalization, since in the light of the first Equivalence axiom it equates to $((p \supset q) \wedge (q \supset p)) \Leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$.

[2] See David H. Sanford, *If P, then Q: Conditionals and the Foundations of Reasoning* (London: Routledge, 1989) 14-26.

[3] See, for example, Howard Kahane and Paul Tidman, *Logic and Philosophy: A Modern Introduction* (Belmont: Wadsworth, 1995) 25.

in which one finds that any true statement is materially implied by absolutely any other statement, and that any false statement materially implies any other statement whatsoever! Suppose, for example, that a robot, ℜ, knows that Ms. Gonzales is in Paris. (The statement P is true.) With Material Implication and Addition, ℜ could reason as follows…

1	She is in Paris.	P	Known
2	Either she's not in Germany or she's in Paris.	$\sim G \vee P$	1;Addit., Commutat.
3	So if she's in Germany, she's in Paris.	$G \supset P$	2; Material Implication

If ℜ realizes that G is false — since to be in Paris is not to be in Germany — ℜ might phrase 3 counterfactually — in terms of what *would* be the case *if* G were true: «If she were in Germany, she would be in Paris.» But that would hardly alleviate the paradox. And of course, ℜ could similarly infer that she would be in Paris if she were in Spain, India, or on the moon! Clearly something isn't right here.

When statements known to be false are involved, the results of using Material Implication together with Addition can be just as bizarre. A robot can easily reach the conclusion that contrary counterfactual conditional statements are both true. Suppose ℜ knows that a certain stock did not go up. (The statement U is false.) Material Implication and Addition would permit the following peculiar line of reasoning…

1	The stock did not go up.	$\sim U$	Known
2	Either it didn't go up or we made a profit.	$\sim U \vee P$	1; Addition
3	So if it had gone up, we would have made a profit.	$U \supset P$	2;Material Implication
4	Either it didn't go up or we didn't make a profit.	$\sim U \vee \sim P$	1; Addition
5	So if it had gone up, we would not have made a profit.	$U \supset \sim P$	4;Material Implication

Presumably, we would not want to have ℜ thinking both 3 and 5! We could ill afford to have our robots endorsing contrary assertions. Nor would we want them to reach a conclusion like either line 3 or line 5 on the mere «evidence» that the stock did not go up!

However troublesome — and worth avoiding — they may be, such paradoxical results are by no means the extent of the trouble that the use of Material Implication in its usual form can occasion.[4] Especially serious problems have been noted to arise when the axiom is applied to statements involving negated conditionals. A beautiful example was presented by Charles L. Stevenson in 1970:

> This is false: if God exists then the prayers of *evil* men will be answered. So we may conclude that God exists, and (as a bonus) we may conclude that the prayers of evil men will not be answered.[5]

Using the original version of the Material Implication axiom, we must regard the conditional $G \supset P$ as logically equivalent to $\sim G \vee P$ and it's negation as equivalent to $G \wedge \sim P$. It seems that we have either a new proof of the existence of God or a serious problem with Material Implication!

Evidently, reasoning such as this could make robots highly unreliable. Suppose we instruct ℜ, whom we have programmed to use Material Implication in its original form, to monitor various transmissions for remarks on certain topics and to verify intercepts on those

---

[4] For a brief discussion of problem cases, see Robert E. Rodes, Jr., and Howard Pospesel, *Premises and Conclusions: Symbolic Logic for Legal Analysis* (Upper Saddle River, NJ: Prentice Hall, 1997) 261-66.

[5] «If-iculties,» *Philosophy of Science* 37 (1970): 28.

topics by turning to a reliable source (RS) — perhaps a human operator. Imagine then that one day ℜ encounters the assertion «If Jones was involved, Smith was involved.» Following instructions, ℜ seeks confirmation from RS:

ℜ: «Is it true that Smith was involved if Jones was?»

RS: «No.»

The problem now is that ℜ, applying the standard axioms, can reason as follows:

1	It's not true that Smith was involved if Jones was.	~(J ⊃ S)	From RS
2	It's not the case that either Jones wasn't involved or Smith was.	~(~J ∨ S)	1; Material Implication
3	Jones was involved but Smith wasn't.	J∧~S	2;DeMorgan,Double Negat.
4	Jones was involved.	J	3; Simplification
5	Smith wasn't involved.	~S	3; Simplification

This leaves ℜ with two beliefs confirmed by a reliable source: (1) that Jones was involved and (2) that Smith was not. If later queried as to who was involved, ℜ will reply accordingly: Jones was, but Smith was not. And any decisions subsequently made by ℜ will be based on these conclusions. For instance, if instructed to distribute certain sensitive information exclusively to individuals not involved, ℜ will send that information to Smith but not to Jones.

But there is no reason to believe that RS in replying as it did meant to say that one of the two individuals was involved and the other was not. RS's reply may have reflected no more than the belief that Jones *could* have been involved without Smith being involved as well. As a consequence of applying Material Implication, ℜ has acquired beliefs that could well be false.

Will this problem with negated conditionals arise only when a robot is told that some conditional statement is false? Evidently not. There are various other ways in which a reasoning robot could encounter negated conditionals. This could happen, for example, if a robot thinking for itself applied the axiom Modus Tolens to a conditional statement with a conditional antecedent. Suppose ℜ knows that if the object breaks when Jane drops it, it isn't made of plastic. A problem could arise if ℜ learned that the object is in fact plastic and proceeded to reason as follows…

1	If it breaks if she drops it, it's not plastic.	(D⊃B)⊃ ~P	Known
2	But it is plastic.	P	Learned
3	So it's not the case that it will break if she drops it.	~(D ⊃ B)	1,2;DoubleNegation, MT
4	It's not true that either she won't drop it or it will break.	~(~D ∨ B)	3; Material Implication
5	She will drop it and it won't break.	D∧~B	4;DeMorgan,DoubleNegat.
6	So she is going to drop the object.	D	5; Simplification

ℜ is thus left thinking that Jane is going to drop the object. By using the original Material Implication axiom, ℜ has arrived at a belief that does not follow from the premises and may well be false. No careful human reasoner would infer the last line from the first two.

Use of the original Material Implication axiom on negated conditionals could also lead robots to have false beliefs about the beliefs of others. Suppose ℜ monitors the following conversation on the assumption that neither of the parties is lying:

Smith: They won't attack unless they are threatened.

Jones: I don't agree.

In this case, ℜ will understand Smith to believe that ~T ⊃ ~A and Jones to believe that ~(~T ⊃ ~A) , which by the original Material Implication axiom equates to ~TA. ℜ, that is, has Jones down as thinking that they will not be threatened but *will* attack. But Jones might not have that belief at all: he might only (and in fact would probably only) believe that the attack *could* occur without any threat being made.

Clearly, reasoning robots cannot be allowed to use the Material Implication axiom in its traditional form: the problems that can result are just too serious. But then the question arises as to what form of the axiom, if any, robots should be given. What I am going to suggest is that they be given precisely half of it. In its usual form, the axiom is a statement of two-way entailment: (p ⊃ q) ⇔ (~p ∨ q). And this is of course equivalent to the conjunction of two one-way entailment axioms: (p ⊃ q) ⇒ (~p ∨ q) and (~p ∨ q) ⇒ (p ⊃ q). Of these, I propose letting our robots have only the first, putting the second, as it were, on the list of «forbidden inferences.» I want to have them think, as does P. F. Strawson, that while conditional statements entail disjunctions, disjunctions do not entail conditional statements.[6] Obviously this involves taking the position (or at least giving our robots to understand) that a conditional statement means something significantly more than the disjunction of its consequent with the negation of its antecedent. To say «If p, then q,» is, on this position, to say more than just «Either p is false or q is true.» But, at least insofar as statement logic is concerned, our robots will get only the one-way entailment axiom (p ⊃ q) ⇒ (~p ∨ q). And to avoid confusion, in fact, I think it may be best not even to use the name «Material Implication» in referring to it. For want of a better name, I shall refer to it simply as «Deconditionalization.»

Furthermore, in the same interest of avoiding confusion, I think it may be a good idea not to continue the traditional use of the horseshoe symbol (⊃) in representing conditional statements. Use of this symbol goes back to Giuseppe Peano, who created it in 1891 by turning the letter «c» around. He would use it, he said, to represent «*b* is a consequence of *a*,» or «if *a* then *b*,» and in doing so he may well have meant to capture the full meaning of «if…then.»[7] But since that time the symbol, slightly elongated in form and dubbed the «horseshoe,» has become so thoroughly associated with the material implication interpretation of «if…then» that to use it while placing half of the traditional Material Implication axiom «off limits» could be badly confusing. According, I shall use the pound sign (#), rendering «If p, then q,» as (p # q), with the understanding that conditional statements so represented retain the full meaning of natural language statements — not just a part of that meaning, however significant that part might be.

Two points should be noted here with respect to Deconditionalization. First, I am continuing to put the right-hand side of the axiom in terms of a disjunction, though it could just as well be put as the negation of a conjunction: p # q ⇒ ~(p∧~q). And second, the axiom can be conveniently taken as having a second line:

Deconditionalization: (p # q) ⇒ (~p ∨ q)

(p∧~q) ⇒ ~(p # q)

The second line, which evidently results from transposing the first and applying De Morgan, may in certain cases allow robots to see the falsity of a conditional claim.

---

6 See Strawson, *Introduction to Logical Theory* (London: Methuen, 1952) 82-83.

7 «Principii di logica matematica,» Rivista di matematica, 1 (1891): 1-10.

Now the question is whether giving our robots Deconditionalization instead of Material Implication will remedy the sort of problems already noted. I believe in fact that it effectively eliminates them. First of all, the paradoxes of material implication are prevented from arising. With regard to the first example above, for instance, ℜ can infer that Ms. Gonzales is either not in Germany or else is in Paris — but cannot get from that to the conclusion that she is (or would be) in Paris if she is (or were) in Germany. Without the «off-limits» axiom $(\sim p \vee q) \Rightarrow (p \supset q)$, $\sim G \vee P$ does not entail $G \supset P$. Nor could ℜ, in the second example, get from the knowledge that the stock did not go up to the conclusion that we would have made (or would *not* have made) a profit if it had gone up. In the list of axioms ℜ is allowed to use, there is no longer one for going from a disjunction to a conditional statement.

But might not robots still encounter the paradoxes if they are allowed to use the logical technique known as conditional proof (CP) — assuming certain statements to be true and reaching conclusions about what follows if they are? In particular, could not ℜ reason as follows in the case of the first example?

1.	She's in Paris.	P	Known
2.	She's not in Germany or else she's in Paris.	$\sim G \vee P$	1; Addition, Commutation
3.	She's in Germany.	G	Assumed for CP.
4.	She's in Paris.	P	2,3; DoubNeg, Disj.Syll.
5.	So if she's in Germany, she's in Paris.	G # P	3-4; CP

In other words, could not ℜ reach the same paradoxical conclusion even without Material Implication by using conditional proof? The answer, I submit, is that ℜ could not. The proof above would not work because ℜ could not assume G on line 3. To do so would be, in effect, to add a premise inconsistent with the premise on line 1. If Ms. Gonzales is in Paris, she is not in Germany. And our robots cannot be allowed to assume things known to be false. ℜ would be instructed to reason only from consistent premises and so could not accept both «She is in Paris,» and «She is in Germany,» as premises. Since in this case ℜ knows the former to be true, ℜ could not assume the latter.

Yet isn't it ever possible to assume something false (suppose it to be true) for the sake of seeing what *would* follow? Could not ℜ reason about what would follow if Gonzales *were* in Germany, saying, as it were, «We know she's not in Germany, but what if she were?» That could indeed be done, but no paradox would arise — because in supposing for the sake of argument that she is in Germany, ℜ would have to abandon the original premise that she is in Paris. «Very well. We'll suppose she's in Germany and not in Paris.» And without P, the argument that generated the paradox would not go through.

Similarly, CP could not be used to revive the paradox from the second example either:

1.	The stock isn't going up.	~U	Known
2.	Either it isn't going up or we'll make a profit.	$\sim U \vee P$	1; Addition
3.	It's going to go up.	U	Assumed for CP.
4.	We will make a profit.	P	2,3; DoubNeg, Disj.Syll.
5.	So if it goes (were to go) up, we'll make (would make) a profit.	U # P	3-4; CP

Here it is obvious that ℜ cannot make the assumption in 3: it directly contradicts the premise on line 1.

The point is really quite general: robots will have to be instructed not to accept any statement as a fact or an assumption unless it is consistent with all statements already accepted

as facts or assumptions. Before committing any statement to memory as true, reasoning robots will have to check for the consistency of that statement with all other statements already held to be true. We cannot expect robots to reach acceptable conclusions if we permit them to tolerate inconsistency.

It may be worthwhile to note at this point that with Deconditionalization replacing Material Implication there seems to be no need to tell our robots anything at all about «possible worlds» in conjunction with conditional statements. The elaborate analyses introduced by David Lewis and Robert Stalnaker[8] will not have to be taken into consideration in order to have robots use «if…then» correctly — even in counterfactual cases.

And what now of the problems that can arise from the use of Material Implication in conjunction with negation? Our robots will not encounter them, I think, so long as they stick to using Deconditionalization instead of Material Implication. The reason is simple: without the two-way entailment, robots will not be able to get from ~(p # q) to p∧~q. They will not find «This is false: if God exists then the prayers of *evil* men will be answered,» equivalent to «God exists, and the prayers of evil men will not be answered.» By the axioms permitted them, ~(G # P) is not equivalent to G∧~P. Nor is «It's not true that Smith was involved if Jones was,» ~(J # S), equivalent to «Jones was involved, but Smith wasn't,» J∧~S.

A question of course arises as to just what we should tell our robots the negation of a conditional statement is equivalent to — if not the conjunction of the antecedent with the negation of the consequent. The answer, I believe, is that we cannot tell them that the negation of a conditional is equivalent to anything that can be expressed in statement logic. «If a then b,» means something significantly more than «Either a is false or b is true,» so to deny that «If a then b,» is true is to do something more than assert that a is true and b is not. And the something more cannot be put in terms of simple statements and logical operators. In statement logic, it appears, our robots cannot have *any* axiom of the form «~(p # q) ⇔ …».

But will they at least have an axiom of the form «~(p # q) ⇒ …» so that they can infer *something* from the negation of a conditional statement? I am afraid the answer is that we cannot give them any such one-way entailment axiom in statement logic either. As I have said, in asserting «If a then b,» one is asserting more than «Either a is false or b is true.» The latter is, to be sure, part of what one means — as is explicitly acknowledged by Deconditionalization ((p # q) ⇒ (~p ∨ q)). But ~a ∨ b is only part of what is asserted when one asserts (a # b). So when (a # b) is denied, it is not necessarily the case that ~a ∨ b is being denied. By analogy, in asserting that Jones is a bachelor, one is asserting both that Jones is a man and that Jones is unmarried — so that, if I say «Jones is not a bachelor,» I am not necessarily saying that Jones is married. (I may be denying that Jones is a man.)

What then are robots to do when they encounter the negation of a conditional statement? Humans, after all, are not stopped cold by such an encounter! If we want our robots to reason like humans (who are reasoning correctly, that is), we cannot leave them stymied when conditionals are negated. We have to tell them how to proceed in such cases.

Consider first the sort of case in which the truth of a conditional is denied by someone with whom a robot is communicating. Here the robot can simply ask for an explanation,

---

[8] See Lewis, «Counterfactuals and Comparative Possibility,» and Stalnaker, «A Theory of Conditionals,» both in William L. Harper, Robert Stalnaker, and Glenn Pearce, eds., *Ifs* (Dordrecht: D. Reidel, 1981) 41-85.

inquiring — as it were — why the other party (OP) believes the conditional in question to be false. In practice, this will involve asking whether the other party believes something that would entail ~(a # b). In most cases, if OP believes a conditional statement is false it will be because OP believes (1) that the contrary of the conditional is true or (2) that it is possible for the antecedent of the conditional to be true while the consequent is false. So in most cases our robot will be able to proceed by formulating questions about just these two possibilities.

If, however, we want to represent these questions symbolically, we will have to go beyond statement logic, since while the contrary of (a # b) can be represented in statement logic as (a # ~b), a symbolic representation of the assertion that the antecedent can be true without the consequent being true requires a symbol for possibility. We need to represent the statement «It is possible that a∧~b». Moreover, we cannot use the ordinary possibility operator of modal logic (◊), since this symbol is usually used to represent *logical* possibility. If we wrote «◊(a∧~b),» we would be representing «It is logically possible that a∧~b.» But the possibility that our robot has to ask about when a conditional statement is denied is not limited to logical possibility. We need, accordingly, a symbol for possibility in general. For ease of understanding, I will use a black diamond (♦) for this purpose. The possibility it represents may be logical possibility — or may be possibility of some other sort:

1) They may not get here by five.
   ♦[~(They will get here by five.)]
   ♦(~G)
2) She may have read the report.
   ♦(She read the report.)
   ♦(R)

Similarly, if our robot has to speak of certainty in discussing conditionals, a black square (■) will be used to indicate certainty in general, as opposed to logical certainly in particular:

1) They will get here for sure by five.
   ■(They will get here by five.)
   ■(G)
2) She definitely read the report.
   ■(She read the report.)
   ■(R)

Naturally, we will also have to provide our robots with a general Possibility-Certainty axiom to the effect that ~♦(p) ⇔ ■(~p). Thus «It's not possible that they will get here by five,» ~♦(G) , will be interpreted by robots as equivalent to «It is certain that they will not get here by five,» ■(~G).

In asking about beliefs that would entail the negation of a conditional, our robots are going to be relying on certain other axioms that we ought to make explicit at this point and which, in fact, are not hard to state in statement logic augmented by general possibility and certainty operators. A first additional axiom, which can be called Conditional Contrariety, merely notes the fact that if a conditional statement is true, its contrary must be false:

Conditional Contrariety (p # q) ⇒ ~(p # ~q)

Loosely put, if the truth of a first statement (the antecedent) means that a second statement (the consequent) is true, it is not the case that the truth of the first statement means that the second is false. If it is the case that you will win if you enter the contest, it is not the case that you will not win if you enter. If it is true that she would go to Paris if she did not

have to go to Chicago, it is false that she would not go to Paris if she did not have to go to Chicago. Our robots will have to be able to make inferences such as these.

A second additional axiom for the inference of the negation of a conditional records the analytic truth that a conditional statement is false if it is possible for its antecedent and the negation of its consequent both to be true: $\blacklozenge(p \wedge \sim q) \Rightarrow \sim(p \# q)$. Interestingly, this axiom embodies an important part of the interpretation of «if…then» proposed by the Stoic Chrysippus, who headed the Stoic School in Athens after Zeno and Cleanthes in the third century BC. What Chrysippus apparently held was that a conditional proposition is true when (and only when) it is impossible for its antecedent to be true and its consequence false.[9] This means he would endorse both the axiom just stated and another that I am not willing to let our robots have: $\sim\blacklozenge(p \wedge \sim q) \Rightarrow (p \# q)$. (Giving them that would only, I fear, lead them into more paradoxes.) Still, out of historical deference, I think we can refer to the axiom I *am* proposing to give robots by his name:

Chrysippus $\blacklozenge(p \wedge \sim q) \Rightarrow \sim(p \# q)$

$(p \# q) \Rightarrow \sim\blacklozenge(p \wedge \sim q)$

The second line, of course, is merely the first line transposed.

It may be worth while noting that with Chrysippus, our robots will have an alternate «route» from $\sim p \wedge q$ to $\sim(p \# q)$ — in addition, that is, to that employing the second line of Deconditionalization. In all probability we will have to give our robots a modal axiom concerning the relation between what is and what is possible (in the general sense of possibility):

Actuality-Possibility $p \Rightarrow \blacklozenge p$

$\sim\blacklozenge p \Rightarrow \sim p$

(Here again, the second line of the axiom is merely the first line transposed.) But by applying this axiom, robots can reason from $\sim p \wedge q$ to $\sim(p \# q)$ without using Deconditionalization:

1	Smith was at the meeting and Jones wasn't.	$S \wedge \sim J$	Discovered
2	So it was possible for Smith to be at the meeting without Jones being there.	$\blacklozenge(S \sim J)$	1; Actuality-Possibility
3	So it's not the case that Jones was there if Smith was.	$\sim(S \# J)$	2; Chrysippus

Now we can return to the question of how to have robots respond when the other party, with whom they are communicating, negates a conditional statement. In the abstract, then, a robot could respond as follows:

OP: $\sim(A \# B)$

$\Re$: Do you think that $(A \# \sim B)$ or that $(A \wedge \sim B)$ or that $\blacklozenge(A \wedge \sim B)$?

Here, $\Re$ is inquiring about beliefs which OP might have that would *entail* $\sim(A \# B)$. Now consider the following concrete example:

---

[9] Josiah B. Gould, *The Philosophy of Chrysippus*, (Albany: State University of New York Press, 1970) 80.

O P	It's not the case that if they drink that beverage they will get sick.	~(D # S)
ℜ	Do you mean that they can drink it without getting sick — or that if they drink it they won't get sick?	♦(D∧~S)? (D # ~S)?
OP	I mean it's possible for them to drink it and not get sick.	♦(D∧~S)

Notice that ℜ does not even have to ask whether OP believes that D∧~S. OP clearly does not have to believe that they are going to drink the beverage and not get sick in order to think it is false that D # S. And if as a matter of fact OP does happen to be convinced that the parties in question are going to drink and not get sick (D∧~S), OP can be assumed to be using the Actuality-Possibility axiom too, and thus to believe (or at least to *constructively* believe) that ♦(D∧~S). So the two alternatives suggested by ℜ are sufficient. The truth of either would entail (and thus explain the belief of OP) that ~(D # S). After OP's responds, ℜ will know what OP is thinking.

There will also be cases in which the party who denies the truth of a conditional does so because of the belief that its contrary is true:

ℜ:	If we take this road, will we get to the camp?	(R # C)?
O P	No.	~(R # C)
ℜ	Do you mean that we can take this road and not get to camp — or that we won't get to camp if we take it?	♦(R∧~C)? (R # ~C)?
OP	I mean we won't get to camp if we take this road.	R # ~C

Here again there is no point in ℜ inquiring whether OP thinks that R∧~C. OP may not know whether the road will be taken or not, but OP is definitely of the opinion that taking it would prevent getting to camp. By asking such questions when a conditional is negated by some other party, a robot can at least avoid mistaken conclusions as to what that party believes.

Is there nothing else that the party denying a conditional might mean? If fact, in rare cases, when inquiring why OP denies that (a#b), ℜ may find that OP thinks neither that a#~b nor that ♦(a∧~b). There is a third possibility, though it is not one that will be encountered very often. Moreover, as it happens, it is not something that can be couched in terms of statement logic — or in statement logic augmented by a modal operator. Suppose, for example, someone says, «If Jim has his lucky rabbit's foot with him when he plays, he'll win,» L # W, to which someone else replies, «That's not so,» ~(L # W). Suppose then that ℜ tries the usual approach, asking questions concerning beliefs that would entail the contradictory of the conditional:

O P	That's not so.	~(L # W)
ℜ:	Do you mean it's possible that he'll have his lucky rabbit's foot with him when he plays and still not win — or do you mean that if he has his lucky rabbit's foot with him when he plays, he won't win?	♦(L∧~W)? (L # ~W)?
OP	Neither: I mean that his having his lucky rabbit's foot with him when he plays would have no effect at all on whether or not he wins.	

OP may be of the opinion that Jim is going to win whether or not he carries the rabbit's foot. (OP may know, for example, that the contest has been fixed, or that the opposition is just no match for Jim.) Or OP may have no idea of how the contest is going to turn out — but may be certain that the outcome is not going to be affected by the absence or presence of a

rabbit's foot. In this case, OP denies that L # W , because OP believes that the truth of L has no bearing on the truth or falsity of W. This, however, is not something that can be expressed in statement logic with or without augmentation by a possibility operator. So in this case ℜ will have to attribute neither the belief that ♦(L∧~W) nor the belief that L # ~W to OP on the basis of the latter's assertion that ~(L # W). ℜ could, of course, proceed to ask, «Do you think the truth of L is unrelated to the truth of W?» and note the answer as indicative of OP's belief.

And what about cases in which a robot encounters the negation of a conditional while following a line of reasoning itself? What if ℜ infers a negated conditional as in the example about the object that might be dropped?

1.	If it breaks if she drops it, it's not plastic.	(D # B) # ~P	Known
2.	But it is plastic.	P	Discovered
3.	So it's not the case that it will break if she drops it.	~(D # B)	1,2; DoubleNegat, MT

Without Material Implication, ℜ is in no danger of going from line 3 to D∧~B. The improper inference from lines 1 and 2 to the conclusion that she is going to drop the object does not go through. But is there nothing that ℜ can infer from line 3? I think the answer has to be «Nothing in statement logic or statement logic with modal operators.» With axioms other than statement-logic axioms, ℜ could get something like «D would not mean that B» or «The facts that would make D true would not also make B true.» Ultimately, of course, our robots will have to have such non-statement-logic axioms.

It should also be noted that robots using the material implication interpretation of negated conditionals and not having access to the axiom I am calling Chrysippus would be unable to make certain valid inferences. Suppose, for example, that ℜ hears from a reliable source that either Jones is lying or area two was contaminated if area one was. ℜ interprets this as J ∨ (O # T). Then ℜ learns from another reliable source that in fact it is possible that area one was contaminated but area two was not: ♦(O∧~T).

1.	Either Jones is lying or else area two was contaminated if area one was.	J ∨ (O # T)	Known
	It may be that area one was contaminated but area two wasn't.		
2.		♦(O∧~T)	Learned

The problem here is that ℜ cannot infer from these premises that Jones is lying. In order to get J from premise 1 by Disjunctive Syllogism on the material implication interpretation of negated conditionals, ℜ would have to have O∧~T , not just ♦(O∧~T). ℜ would have to know, that is, that area one had actually been contaminated while area two had not. But the mere possibility that area one could have been contaminated without area two being contaminated would certainly be enough to eliminate the second disjunct in 1 and prove Jones a liar. The Chrysippus axiom would, of course, do the trick, by permitting an inference from line 2 to ~(O # T).

Or suppose that in a similar case ℜ knows that if both Smith and Jones were at the meeting, then Roberts was too. And then ℜ learns that Jones might have been at the meeting without Roberts being there as well. So ℜ starts to reason…

1.	If Smith and Jones were at the meeting, then Roberts was too.	(S∧J) # R	Given
2.	It's possible that Jones was at the meeting but Roberts wasn't.	♦(J∧~R)	Learned
3.	If Smith was at the meeting, then Roberts was too if Jones was.	S # (J # R)	1; Exportation

On the material implication interpretation of negated conditionals, ℜ would need J∧~R to get ~S from 3 by Modus Tollens. But ℜ has only ♦(J∧~R), and so could go no further — without Chrysippus. With Chrysippus, however, ℜ can reach the humanly obvious conclusion:

4.	It's not true that Roberts was at the meeting if Jones was.	~(J # R)	2; Chrysippus
5.	Smith wasn't at the meeting.	~S	3,4; Modus Tollens

3. Material Equivalence

As the reader may have suspected, replacement of the two-way-entailment axiom Material Implication by the one-way-entailment axiom Deconditionalization will necessitate analogous alteration of one of the pair of two-way-entailment axioms known as Material Equivalence:

a) $(p \equiv q) \Leftrightarrow ((p \supset q) \wedge (q \supset p))$

b) $(p \equiv q) \Leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$

First of all, to indicate that the full meaning of «if…then» is involved, the horseshoe will be replaced by the pound sign. But then what about the triple-barred equal sign? It represents «if and only if» but is intimately connected, through this axiom, to the material implication interpretation of biconditionalization. For clarity, I propose replacing it with a double pound sign, ##, understood to represent the full ordinary meaning of «if and only if»:

a) $(p \,\#\#\, q) \Leftrightarrow ((p \# q) \wedge (q \# p))$

b) $(p \,\#\#\, q) \Leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$

Aside from these changes, the first line of the axiom does not need to be altered. It simply notes that saying «p if and only if q,» is equivalent to saying «If p then q, and if q then p.» The second line of the axiom, however, says that the assertion that p if and only if q is *equivalent* to the assertion that either both statements are true or both are false. But this would permit ℜ to make inferences such as the following…

1.	Amanda is in Brasil.	A	Known
2.	Juan is in Argentina.	J	Known
3.	Amanda is in Brasil and Juan is in Argentina.	A∧J	1,2; Conjunction
4.	Either Amanda is in Brasil and Juan is in Argentina or else Amanda isn't in Brasil and Juan isn't in Argentina.	(A∧J) ∨ (~A∧~J)	3; Addition
5.	So Amanda is in Brasil if and only if Juan is in Argentina.	A ## J	4; Material Equivalence

Presumably, we do not want ℜ reaching a conclusion like 5 from premises like 1 and 2. Knowledge that two statements are both true should not lead to the conclusion that one of them is true if and only if the other one is! But in the light of what has already been said about Material Implication, the solution is obvious: we simply reduce the second line of Material Equivalence to a statement of one-way entailment:

a) $(p \,\#\#\, q) \Leftrightarrow ((p \# q) \wedge (q \# p))$

b) $(p \,\#\#\, q) \Rightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$

This will effectively preclude the move from line 4 to line 5 in the current example. Should we then continue to refer to the axioms (a) and (b) as Material Equivalence? Perhaps — because of the way in which we are rejecting the material implication interpretation of «if-then» and «if and only if» — we should not. For simplicity, I shall refer to this pair of axioms as «Biconditionalization.»

4. Transposition

Troublesome as it may be, the Philonian (material-implication) interpretation of «if…then» is not the only source of potential difficulties for robots reasoning with conditional statements. Robots could also run into trouble when applying the traditional form of the axiom known as Transposition, $(p \supset q) \Leftrightarrow (\sim q \supset \sim p)$. In particular, problems could arise if a robot applied the axiom to conditionals having subcontraries as antecedent and consequent. When the antecedent and consequent of a conditional statement could both be true but could not both be false, Transposition can lead to trouble. For example, a robot knowledgeable about European geography could easily reason as follows…

1.	If he's not in Spain, he's in France.	~S # F	Learned from a reliable source
2.	If he's in France, he's not in Germany.	F # ~G	Known
3.	If he's not in Spain, he's not in Germany.	~S # ~G	1,2; Hypothetical Syllogism
4.	So if he's in Germany, he's in Spain.	G # S	3; Transposition

We cannot have our robots reasoning to impossible conclusions like that on line 4 — so something obviously has to be done. But what? Perhaps the best we can do is to block applications of Transposition like the one on line 4 by inserting a sort of «filter» into the axiom:

Transposition $\quad ((p \# q) \wedge \blacklozenge(\sim q \wedge \sim p)) \Rightarrow (\sim q \# \sim p)$

The effect of the conjunct with the general possibility operator will be to filter out cases in which antecedent and consequent are subcontraries. (For obvious reasons, conditionals in which antecedent and consequent are contraries will not be encountered as true premises.) Taking this approach will mean that Transposition cannot be written as a two-way entailment statement, but that should pose no particular problem. And the use of Transposition in this form would clearly block the step from line 3 to line 4 in the example just given.

Are there other problems with Transposition that could be prevented by giving robots this axiom in the altered form? It does appear that the proposed «filter» — could also block certain absurd results that robots might encounter when using Transposition if they were told to interpret «q, even if p» as p # q. Suppose we tell $\mathfrak{R}$ that even if the hosts of some reception served fruit, they did not serve cherries. Without the filter, $\mathfrak{R}$ could proceed to reason as follows:

1.	Even if they served fruit, they did not serve cherries.	F # ~C	Acquired from a reliable source
2.	So if they served cherries, they did not serve fruit.	C # ~F	1; Transposition and Double Negation

But the conclusion on line 2 is the sort of nonsense we cannot permit. Using the restricted version of Transposition would, however, prevent the inference to line 2, since it is not possible that the hosts served cherries but did not serve fruit.

Moreover, precisely the same sort of problem could arise with a counterfactual «even if» conditional. Suppose $\mathfrak{R}$ knows that Jones is not in France and learns from a reliable source

that even if he were in France, he would not be in Paris. It would be only too easy for ℜ to reason in the following manner:

1.	Jones is not in France	~F	Known
2.	Even if Jones were in France, he wouldn't be in Paris.	F # ~P	Learned from a reliable source
3.	If Jones were in Paris, he wouldn't be in France.	P # ~F	2; Transposit., and Double Negation

Here again, the absurd conclusion could be prevented by giving ℜ only the restricted version of Transposition. The incompatibility of P and ~F would keep Transposition from being applied to line 2.

But while the restricted version of Transposition can prevent the sort of problems indicated with «even if» statements, I do not think it is the best solution. I think in fact it would be preferable *not* to let our robots interpret «q, even if p» as p # q. The question is how we should have them interpret it. As a minimum, perhaps, we can tell them to take «q, even if p» to entail q. We can let them assume that someone who asserts an «even if» statement is asserting the truth of the part that is *not* preceded by «even if». (Because «even if» statements seem *not* to be ordinary conditional statements, it may be best not even to use the terms «antecedent» and «consequent» for their parts.) If a speaker says, «Even if they served fruit, they did not serve cherries,» we can have our robots think that the speaker has at least asserted «They did not serve cherries.» And if someone writes, «Even if Jones were in France, he wouldn't be in Paris,» we can have our robots take the writer to have endorsed the proposition that Jones is not in Paris.

But is there nothing more we can or should tell robots about the meaning of «q, even if p»? There *is* one thing we can add innocuously, although I am not sure that it will prove to be of great use. It is simply ~(p # ~q). We will then be telling our robots that «q even if p» at least entails q ∧ ~(p # ~q). However, as already noted, there is not much within statement logic that robots will be able to do with the negation of a conditional.

5. Conclusion

Here then are the «forbidden inferences» that we must declare off limits to our robots:

1) $(\sim p \vee q) \Rightarrow (p \supset q)$

2) $((p \wedge q) \vee (\sim p \wedge \sim q)) \Rightarrow (p \equiv q)$

And here are the alternate axioms that I have argued our reasoning robots will have to use:

Deconditionalization	$(p \# q) \Rightarrow (\sim p \vee q)$
	$(p \wedge \sim q) \Rightarrow \sim(p \# q)$
Conditional Contrariety	$(p \# q) \Rightarrow \sim(p \# \sim q)$
Chrysippus	$\blacklozenge(p \wedge \sim q) \Rightarrow \sim(p \# q)$
	$(p \# q) \Rightarrow \sim\blacklozenge(p \wedge \sim q)$
Biconditionalization	$(p \#\# q) \Rightarrow ((p \wedge q) \vee (\sim p \wedge \sim q))$
Transposition	$((p \# q) \wedge \blacklozenge(\sim q \wedge \sim p)) \Rightarrow (\sim q \# \sim p)$
Possibility-Certainty	$\sim\blacklozenge(p) \Leftrightarrow \blacksquare(\sim p)$

Actuality-Possibility $p \Rightarrow \blacklozenge p$

$\sim\blacklozenge p \Rightarrow \sim p$

Using these axioms, I submit, reasoning robots will be able to «get it right» when making inferences involving «if…then,» and we will consequently be able to trust their reasoning when conditional statements are involved.

Ronald A. Cordero
Department of Philosophy. The University of Wisconsin at Oshkosh
Oshkosh, Wisconsin, USA 54901
<cordero@uwosh.edu>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 36-41
A Dilemma for Robust Alethic Relativism
Copyright © by SORITES and William Ferraiolo

A DILEMMA FOR ROBUST ALETHIC RELATIVISM

William Ferraiolo

<bferraiolo@sjdccd.cc.ca.us>

The robust alethic relativist claims that no truth bearer is objectively true. The most we can ever say of any truth bearer (statement, belief, proposition, etc.) is that it is true *relative to* some conceptual framework, worldview, situation, or other parameter. The most we can ever say of any truth bearer, according to the robust alethic relativist, is that it is *true-for-X*. In this paper, I will argue that robust alethic relativism is entirely untenable, as it is either a self-refuting thesis, or it must degenerate into something that cannot coherently serve as a *theory of truth* (or anything else) at all. I hope to show that Socrates understood this difficulty for the thoroughgoing relativist, and that his objection against this thesis is more compelling than has been recognized.

Socrates and Protagoras

Socrates was, perhaps, the first to charge the doctrine of robust alethic relativism with self-refutation (i.e. in order that it be true, it must be false). His charge is offered in response to Protagoras' (alleged) assertion that «man is the measure of all things — alike of the being of things that are and of the not-being of things that are not» (*Theaetetus* 152a). This assertion was interpreted (by Plato, Socrates, and Aristotle) as an expression of Protagoras' commitment to a thoroughgoing relativism about truth. Plato's *Theaetetus* (152a-171d) shows his argument against Protagoras proceeding as follows:

1. You hold that what seems to me to be true *is* true *for me*, and that what seems to you to be true *is* true *for you*.
2. I believe that doctrine to be false (i.e. it seems false to me).
3. If the doctrine you have put forward is true, then it must be false (because it *seems* so to me and what seems so to me *is so*).
4. Hence, your doctrine cannot be true — it is self-refuting.

Of course, the standard charge against Socrates' objection to Protagoreanism is that it crucially ignores the relativization of truth *to the individual*. Protagoras (so the counterobjection goes) claims that relativism is true *for him*. He is also committed to the thesis that if Socrates does not believe it, then it is not true *for Socrates*. Thus far we have neither self-refutation nor any contradiction. Relativism is true *for Protagoras* and it is not true *for Socrates* (Trp & ~Trs). This does not amount to relativism being both true *and* not true in the same respect, but merely to its being true relative to one individual and not true relative to another. Socrates' objection, therefore, misses the mark because it illicitly assumes the objectivist's account of truth in generating the alleged paradox or self-refutation. In short, the objection begs the question against Protagoras.

Jack Meiland has noted Plato's (or Socrates') apparent failure to see the role of the relativizing locution «true for…» in the Protagorean doctrine:

Plato's own attempt, in the *Theaetetus* to show Protagorean relativism to be self-refuting appears to be radically defective due to Plato's dropping of the relativistic qualifier (the «for me» in «true for me») at crucial points. (Meiland, 1979: p. 54)

A number of other authors have expressed similar concerns about the Socratic charge of self-refutation (see, for example, Jordan, 1971; Swoyer, 1982). But can a theory *about the nature of truth* be, itself, true or false only relatively to some parameter or other? We must note that Socrates challenges Protagoras (albeit *in absentia*) to defend the viability of the Protagorean doctrine *itself*, and not just its application to some arbitrary truth-bearer. Socrates does not choose just any proposition to make the case that it must be false in order that it be true given relativism; he makes that case against Protagoras' articulation of relativism *itself*. The robust relativist's thesis is supposed to range over *all* truth bearers. Can Protagoras' rejection of *objective* truth be true only *relatively*? Can Socrates' claim that there are *objective* truths be only *relatively* true?

The Socratic Dilemma

Socrates' charge is essentially in the form of a dilemma, one horn of which branches into a subsidiary dilemma. The overarching dilemma is as follows:

D: Either the doctrine of alethic relativism is *objectively* true (in which case it is self-refuting) or it is only *relatively* true.

Socrates also saw that if Protagoras were to opt for the second disjunct (as, apparently, he must), he would get caught in the following subsidiary dilemma:

D_S: Either the concept of relative truth depends ineliminably upon the existence of some objective truth(s) (and so is, as before, self-refuting) or it collapses into a triviality about beliefs.

Before explicating this double dilemma, it is worth pointing out that, *pace* Meiland (*et. al*), Socrates appears to be fully aware that Protagoras means to assert only the relative truth, or truth-for-X, of that which is believed by X, or of that which *seems to be so* to X. In fact, Socrates carefully constructs his dilemma so that he will not be guilty of begging the question against Protagoras:

> No, he [Protagoras] will say, show a more generous spirit by attacking what I actually say, and prove, if you can, that we have not, each one of us, his peculiar perceptions, or that, granting them to be peculiar, it would not follow that what appears to each becomes — or is, if we may use the word «is» — for him alone to whom it appears … For I do indeed assert that the truth is as I have written. Each one of us is a measure of what is and of what is not, but there is all the difference in the world between one man and another just in the very fact that **what is and appears to one is different from what is and appears to the other**. (*Theaetetus*: 166c-d — emphasis added)

Socrates seems to have developed his objection with the proper understanding of Protagoreanism in mind.

If the robust alethic relativist is to defend his thesis, then he must do so in (broadly speaking) one of two ways. He must make the case either that the relativistic thesis is true *simpliciter* (i.e. objectively true), or that the thesis is true in the *relative* sense. Obviously, the first disjunct is unacceptable to the robust relativist as it involves the affirmation of a non-relative truth (of which, *ex hypothesi*, there are none). It must be that the robust relativist

means to defend the relative truth (the truth-for-X, replacing X with whatever relativistic parameter you will) of the relativistic thesis. But what exactly would the *relative* truth of the claim that there are no *objective* truths amount to? Is there any account of relative truth upon which a defense of robust alethic relativism may be founded?

The Second Branch of the Socratic Dilemma

The relativist must be operating with some account of truth-for-X that does not rest upon, or ineliminably involve any notion of objective truth (to which the robust alethic relativist is not entitled). Neither can the relativistic account of truth amount to some simple epistemic property of the parameter to which truth is relativized. If truth is relativized to individuals (as Protagoras apparently suggests), the truth-for-X of any given truth bearer must amount to something other than the mere fact that X *believes* the truth bearer in question to be true. If «S is true-for-X» amounts to no more than «X believes S,» then the relativistic notion of truth is entirely uninteresting. Any truth bearer's truth-for-X is a trivial entailment of its being believed by X. Given such an «account of truth,» there can be no dispute and no productive discourse about the superiority of relativism as opposed to any «competing» theory of truth. If whatever anyone believes thereby becomes true-for-her, then the issue cannot even be joined by theorists in opposing camps. Indeed, there can be no opposing camps. There can be no dispute about anything, nor is it possible for anyone to ever have a false belief. If you believe it, it's true-for-you (and *true-for-you* is as true as it gets). What then is Protagoras, or any robust alethic relativist, doing advocating this doctrine and rejecting other theories of truth? Aren't they all equally true-for-whomever-believes-them?

Socrates is aware of the difficulties that Protagoras faces in attempting to defend the relative truth of relativism if the doctrine's «relative truth» means no more than its being believed by some cognizer:

> … if, just as no one is to be a better judge of what another experiences, so no one is better entitled to consider whether what another thinks is true or false, and, as we have said more than once, every man is to have his own beliefs for himself alone and they are all right and true — then, my friend, where is the wisdom of Protagoras, to justify his setting up to teach others and to be handsomely paid for it, and where is our comparative ignorance or the need for us to go and sit at his feet, when each of us is himself the measure of his own wisdom … for to set about overhauling and testing one another's notions and opinions when those of each and every one are right, is a tedious and monstrous display of folly if the Truth of Protagoras is really truthful and not amusing herself with oracles delivered from the unapproachable shrine of his book (*Theaetetus*: 161d-162a)

Socrates is getting at a fundamental problem for the robust alethic relativist. If relative truth is not just, at root, the same thing as objective truth, then it had better not end up being tantamount to mere belief. What does it mean to say that one theory of truth (e.g. relativism) is, in any interesting sense, superior to another if the «truth» of the matter is simply determined for each disputant by virtue of that disputant's particular beliefs?

Furthermore (and more importantly), the relativist who defends such a doctrine is not offering a coherent *theory of truth* at all, but is merely articulating a triviality about beliefs. The relativist asserts:

R: Whatever X believes is true-for-X.

But if being «true-for-X» amounts to no more than being believed by X, or seeming so to X, then R (upon analysis) amounts to:

R': Whatever X believes is believed by X.

Obviously, R' is a trivial and uninteresting claim. It is hardly a *theory of truth* (or of anything else for that matter). This type of «relativism» no longer entails the assertion of any *relative truth* at all. This type of «relativism» is not *relativism*! The relativist, therefore, owes an account of true-for-X that is neither the trivial, uninteresting «thesis» R', nor involves the objectivist notion of truth *simpliciter*. Is there any such account to be given?

Impaled On The Final Horn

What, within the just-mentioned parameters, can «true-for-X» mean? What is there «between» objective truth and mere belief? Meiland mentions a solution that is fairly well representative of attempts to find a middle ground. Relativists, Meiland suggests, must add a *third term* to the truth relation. Standard «objective» truth is a two-term relation between statements (or other truth bearers) and the world. On the objectivist interpretation, a statement is true if and only if there is some objective fact to which the statement is appropriately related (by, for example, correspondence — however, exactly, that relation is to be understood). Put simply, objectivist truth is a word-world relation. Relative truth is, by contrast, a three-term relation involving statements, the world, and something like a conceptual framework or worldview as its relata. Meiland's explicit characterization of these competing conceptions of truth is as follows:

> (1) The concept of absolute truth seems to be a concept of a *two*-term relation between statements (or perhaps propositions) on the one hand and facts (or states of affairs) on the other. But the concept of relative truth, as used by some relativists, seems to be a concept of a *three*-term relation between statements, the world, and a third term which is either persons, world views, or historical and cultural situations.
>
> (2) The relation denoted by the expression 'absolute truth' is often said to be that of correspondence. The relativist can make use of the type of notion and say that «P is true relative to W» means something like «P corresponds to the facts from the point of view of W» (where W is a person, a set of leading principles, a world view, or a situation). (Meiland, 1977: p. 571)

What we have here is, essentially, a correspondence relation that obtains between, not a statement and the world, but rather a statement and some *world version* («the facts from the point of view of W»).

Harvey Siegel correctly points out that Meiland is still really just offering a two-term evaluation of truth, and then dismisses the account on the grounds that it must either co-opt an objectivist conception of truth or else it will fail to be anything more than truth-in-virtue-of-so-seeming (Siegel, 1986; pp. 234-240). In effect, Siegel's charge is that Meiland's account fails to prevent «relative truth» from collapsing into mere belief if truth-for-X is not just a pseudonym for truth *simpliciter*. If world versions are the truth-makers, and world versions are just constructs, unrestrained by *objective* truths (which, according to the robust alethic relativist, do not exist), then we are left with a more complex version of truth-in-virtue-of-so-seeming.

First, as Siegel points out, Meiland's truth relation holds between statements and the world-as-construed-by-X. True-for-X is just a *two*-term relation between words and world *versions*:

> On the relativist conception, the world is not distinguishable from the third relata (either persons, world views, or historical and cultural situations). What are related by the alleged three-term relation are statements and the-world-relative-to-W (where W is a person, a set of leading principles, a world view, or a situation — in short, where W is the third relata). On the relativist conception, the world cannot be conceived as independent of W; if it is so conceived, the relativist conception collapses into an absolutist one, for it is granted that there is a way the world is, independent of statements and W's. This is precisely

> what the relativist must deny, however. So Meiland's three-term relation collapses into a two-term relation, between statements and the-world-relative-to-W … (Siegel, 1986, pp.234-5)

This point, in itself, is not terribly problematic for Meiland. Whether his truth relation involves two or three terms is an ancillary matter. In setting up world versions as truth-makers, however, Meiland lands back in the same predicament from which he sought to extract himself with his «three-term» conception of relativistic truth.

In the absence of a body of objective, invariant truths, there can be no constraint upon the world versions concocted by cognizers. The versions created need not (in fact, *cannot*) conform to pre-existing objective truths (as, according to the robust alethic relativist, there is no such thing). World versions are constructed in an alethic vacuum without external guidance or constraint. A world version is, therefore, *as you like it*. If the-way-that-things-seem-to-X, an unrestrained construct, serves as truth-maker for statements, then any statement is true-for-X so long as it *seems so* to X (or «fits in» with X's world version). This will be the case with any relativistic conception of truth that is not in any way grounded in, or constrained by, objective truth. Whether a statement (or other truth-bearer) is true-for-X will inevitably be a function of whether it fits in with the world-as-it-seems-to-X. But that just means that the statement is true-for-X if it seems so to X. So, again, we are left with a triviality about how things seem that is hardly distinguishable from R'.

There can be no legitimate dispute between the relativist and the objectivist if the relativist «defends» relativism by claiming that relativism «fits in» with her worldview. In fact, such a claim, as we have seen, cannot serve as the expression of a *theory of truth* at all. The robust alethic relativist, it would seem, can neither embrace objective truth nor offer a coherent account of relative truth that does not ineliminably involve some body of objective truth. If the onus of providing a viable alternative to objective truth (truth *simpliciter*) is on the relativist, then the original Socratic charge of self-refutation stands until some such alternative is concocted. But if the relativist's conception of truth is not ultimately constrained by *objective* truth, it degenerates into the aforementioned triviality.

Conclusion

Protagoras did not offer an adequate response to Socrates' charge in the *Theaetetus* — he was dead, after all. It is not, moreover, clear that *any* robust alethic relativist *can* offer an adequate response. If truth makers are entirely unrestrained constructs, then they simply are however we believe them to be. Any set of truth-makers obtains *for me* so long as I believe it to. So, it is what I believe (or the way that things seem to me) that determines the «relative truth» of what I believe! The claim that I believe what I believe, or that things seem to me the way that they seem to me, is hardly a *theory of truth*. It is just an empty triviality. Socrates had Protagoras pegged, and I see little hope that the intellectual heirs to Protagoreanism will fare better than their patriarch in the horns of the Socratic dilemma.

References

Jordan, J. (1971). «Protagoras and Relativism: Criticisms Bad and Good». *Southwestern Journal of Philosophy 2*: pp. 7-29.

Meiland, J.W. (1977). «Concepts of Relative Truth». *The Monist* 60: pp. 568-582.

Meiland, J.W. (1979). «Is Protagorean Relativism Self-Refuting?» *Grazer Philosophische Studien* 9: pp. 51-68.

Plato. *The Collected Dialogues of Plato* (1961). Hamilton, E. and Cairns, H. (eds.) Princeton: Princeton University Press.

Seigel, H. (1986). «Relativism, Truth, and Incoherence». *Synthese* 68: pp. 295-312.

Swoyer, C. (1982). «True For». In *Relativism: Cognitive and Moral*. Krausz, M. and Meiland, J.W. (eds.). Notre Dame: University of Notre Dame Press. Pp. 84-108.

William Ferraiolo
San Joaquin Delta College
5151 Pacific Ave.
Stockton, CA 95207. U.S.A.
<bferraiolo@sjdccd.cc.ca.us>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 42-56
Can a Localist and Descriptive Epistemological Naturalism
Avoid Dogmatic Foundations?
Copyright © by SORITES and Armando Cíntora

CAN A LOCALIST AND DESCRIPTIVE EPISTEMOLOGICAL NATURALISM AVOID DOGMATIC FOUNDATIONS?

Armando Cíntora

<cintora@prodigy.net.mx>

…In my naturalism, I recognize no higher truth than that which science provides or seeks.

(W. V. Quine.)

I

Rationality demands that we justify our beliefs and methods, and then if rational we should justify the methods of science, including our most basic scientific methods, methods such as induction, deduction, and inference to the best explanation. Now, given the lack of past success in providing a non-viciously circular justification of these ultimate scientific methods, the sceptic will conclude that these most basic scientific methods are accepted as correct dogmatically, that is, by an act of faith. If not, the sceptic will ask to be proved wrong by demanding a non-viciously circular justification of these basic methods.

It could be argued, however, that the sceptic is unreasonably over ambitious in his demands, since the sceptic is looking for answers to ultimate questions. It could be argued that instead, one should stay content with limited or particular problems, because our science and the methods presupposed by it have generally been successful. Therefore, it could be claimed that it is unreasonable to entertain global doubts about our science and its methods, and that it is then unreasonable to ask for ultimate justifications for these methods. Otherwise, it could be argued that since our science and its methods have been overall effective in predicting, in giving us control over, some aspects of nature; that since our science has in general provided us with reliable knowledge, why doubt our science and its methods, why search for their global or ultimate justifications? Especially, since these sought justifications are likely to be unavailable.

A recommendation for local, for restricted questions, can be aptly illustrated by the well known metaphor of Neurath's, of a boat that is constantly repaired and improved while always navigating in the open sea; and never being taken to port for a complete overhaul. This boat can only be changed or repaired piece mealy and this is done only when required; the proposal is that we do the same with our system of knowledge, that we question and revise it only here and there and only if serious doubts were to demand it. The localist believes that local or particular justification is all we need to be rationally justified about our corpus of knowledge.

Localism is epistemologically optimist since it believes in potential unlimited improvement: it assumes that our methods, goals and beliefs can in principle proceed indefinitely with a continuous and gradualist process of betterment. Localism then assumes that there are no large-scale errors or gaps in our current corpus of knowledge. Localism assumes that in general, our background knowledge is correct, it has to assume this, because this is the prerequisite to go on with a reformist approach. This conservative assumption of

localism is the prerequisite for not doing a general overhaul; this optimistic assumption is the prerequisite for ignoring global questions. Thus,

> We can change it [our conceptual scheme] bit by bit, plank by plank, though meanwhile there is nothing to carry us along but the evolving conceptual scheme itself.
>
> (Quine, 1953b, p. 78.)

And in case our conceptual scheme were to be confronted with anomalous empirical evidence, then the localist conservatively recommends to accommodate the empirical anomalies with a minimum of alteration to our conceptual scheme.

> Our boat stays afloat because at each alteration we keep **the bulk** of it intact as a going concern.
>
> (Quine, 1960, p. 4.) (Emphasis added.)

Thus, we revise some of our particular beliefs while taking for granted the general validity of the bulk of our scientific procedures and results, while taking for granted our everyday common sense beliefs. These basic common sense presuppositions are akin to the ship's hull, they are what keep the boat afloat; thus, local questions (scientific and normative epistemic ones) are examined against background knowledge, a background that at least for the time being is considered as non-problematic and as consensual. For example, particular features of our methods of inquiry may be evaluated against a framework of accepted common sense beliefs, scientific theory, and some basic methods (such as induction and deduction.) This framework is not questioned: *if the ship keeps on navigating* we just tinker with it. The assumption is made that those of our beliefs and methods that as a matter of consensus have worked can be taken for granted; that they are presumptively true or correct. That is, it is assumed that they are innocent until proven guilty, it is thought that to question them would be an unnatural and unfounded doubt. Thus,

> We cannot begin with **complete** doubt … A person may, it is true, in the course of his studies, find reason to doubt what he began by believing; but in that case he doubts because he has **positive reason** for it, and not on account of the Cartesian maxim. Let us not pretend to doubt in philosophy what **we do not doubt in our hearts.**
>
> (Charles S. Peirce, pp. 228-9.) (Emphasis added.)

Peirce's idea appears to be that we should do not take seriously doubts that we cannot entertain psychologically (i.e., «in our hearts»), and Peirce believes that we should avoid global or complete doubt, he also claims that doubts must be justified. According to Peirce, when one has a genuine doubt, it is because one has specific reasons for doubting, and then one can examine these reasons to find whether they are good reasons for doubting. Pierce's proposal is to deal with real (i.e., local or restricted) and justified questions or problems, rather than with invented wider problems which we cannot entertain psychologically, as those wider doubts proposed by Descartes in his *First Meditation*.[1]

---

[1] For the pragmatist tradition, even amongst restricted questions, deliberation can be excessive, because too much deliberation interferes with a successful life; hence, some rashness is necessary for survival. Thus,

> There is no more miserable human being than one in whom nothing is habitual but indecision, and for whom the lighting of every cigar, the drinking of every cup, the time of rising and going to bed every day, and the beginning of every bit of work, are subjects of express volitional deliberation.
>
> (W. James, p. 122.)

Deliberation is for the pragmatist primarily a means of solving *particular* human problems and it is not carried for its own sake.

The pragmatist tries to avoid (or is it evade?) global sceptical questions by focusing exclusively on local or particular questions. The localist advices us to emulate the alleged attitude of Kuhnian 'normal' scientists who take the presuppositions of their paradigm for granted, and only doubt their paradigm if they have good reasons for doubting it, reasons which for normal scientists can be persistent and numerous important anomalies. The localist wants to reform philosophy so that it imitates normal science by proceeding in a piecemeal fashion, that is, by taking for granted those background assumptions that have the backing of experience, by taking for granted those assumptions that have the backing of scientific tradition.

If we call a 'normal world' a world that is consistent with our general common sense beliefs about how the world is, then the localist is saying that we should be interested in doing science and philosophy in 'normal worlds'. The localist is prescribing that we should not be interested in the difficulties of acquiring knowledge in some bizarre logically possible world inhabited by a malicious Cartesian demon, or in an outlandish world of brains in vats.

II

Quine, as result of his pragmatism, has defended a localist position, but he has added to localism his holist thesis, and as result of his holism, Quine has questioned the synthetic-analytic distinction,

> If (holism) is right … it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may. **Any statement can be held true come what may**, if we make drastic enough adjustments elsewhere in the system. Conversely, by the same token, **no statement is immune to revision**.
>
> (Quine, 1953, p. 43.) (Emphasis added.)

Hence, for Quine, any statement is revisable, this implies that for Quine philosophical beliefs are also modifiable by experience, and then the boundary between science and philosophy (and in particular, between science and epistemology) becomes blurred. On the other hand, it is thought that science proceeds by asking local or particular questions, and that when doing so, science takes for granted its background knowledge and methodology (except if good reasons to doubt some of its background presuppositions were to appear here or there.) Moreover, the pragmatist highly values science and its restricted investigations because of their practical results. As a result, Quine claims that a gradualist localism is the way to proceed in all our investigations, such as epistemological and scientific ones, and epistemology is to be appraised by the method(s) of science. And since science and philosophy are thought to form a continuum, sceptical challenges should arise *within* science, and we should use science to respond to them. Thus,

> … **skeptical doubts are scientific doubts** (…) Epistemology is **best** looked upon, then as an enterprise within natural science. Cartesian doubt **is not** the way to begin.
>
> (Quine, 1975, p. 68.) (Emphasis added.)

For Quine there is no 'first philosophy', that is, for Quine there is not a philosophy that is logically prior to any empirical knowledge. For Quine, there are not extra-scientific methods to assess from some place outside science, the epistemological merits of scientific theories, thus,

> … **Naturalism: abandonment of the goal of a first philosophy**. It sees natural science as an inquiry into reality, fallible and corrigible **but not answerable to any supra-scientific tribunal,** and **not in need of any justification beyond observation and the hypothetico-deductive method**… Naturalism does not

> repudiate epistemology, but assimilates it to empirical psychology (…) [The naturalist] tries to improve, clarify, and understand the system **from within**. He is the busy sailor adrift on Neurath's boat.
>
> (Quine, 1981, p. 72.) (Emphasis added.)

Epistemology becomes the study of science from within science, and in this way epistemology loses its special character, for this doctrine (from now on 'naturalism') the empirical sciences[2], their methods and results are what guide philosophy. Thus,

> Science itself, in a broad sense, **and not some ulterior philosophy**, is where judgment is **properly** passed, however fallibly, **on questions of truth and reality.**
>
> (Quine, 1992, p. 295.) (Emphasis added.)

Epistemological naturalism[3] considers human knowledge a natural phenomenon to be studied the same way as any other aspect of nature, epistemological naturalism does not answer the philosophical sceptic; rather it says that Cartesian scepticism is psychologically and scientifically implausible. The naturalist takes for granted what the sceptic questions.

Naturalism could be characterized as the rejection of transcendental argument, that is, of non-empirical argument, naturalism recommends replacing *a priori* philosophy with scientific theory; and it claims that epistemology is just the study of science from within science.

III Critical Evaluation of the Localist-Naturalist Thesis

If the localist-naturalist approach is used to justify our scientific methods it is circular

Quine tells us in a quote above that science is in no need for «any justification beyond observation and the hypothetico-deductive method», hence Quine values observation and the hypothetico-deductive method, which he considers as the correct research methods, since he also claims that science is «where judgment is properly passed… on questions of truth and reality.» The question now arises of how does Quine know that this is the *proper* method to judge «on questions of truth and reality».

On the other hand, since Quine claims that natural science «is not answerable to any supra-scientific tribunal» (since there is no room for an «ulterior philosophy»), then for Quine any justification for what he takes to be scientific method has to come from within science. And given that the sciences in turn are selected and validated by scientific method, the justification of scientific method has to come, in Quine's approach, ultimately from scientific method itself. Hence, we end by circularly justifying scientific method with scientific method, because there is no 'first philosophy', because epistemology is just an activity within natural science, because science is the *only* tribunal where questions of truth and reality are 'properly' settled.

If it were answered, for example, that the available evidence, say the evidence provided by the history of successful science, warrants belief in scientific method, then this argument would be circular. Because we use scientific method to select what is to be taken as bona fide

---

[2] Which empirical sciences are privileged as the archetypes to follow depends on the naturalist philosopher.

[3] In summary, Quine's naturalism appears to be the result both of his holism and of the high value he gives to science and to its localist research strategy. Quine's high valuation of science is shown, for example, in Quine's last quote above, where he claims: «science itself,…, is where judgement is properly passed.»

successful science, and to decide that the historical evidence so selected supports a belief in scientific method. We cannot validate in a non-circular way the methods of empirical science by appeal to some empirical science, Quine is aware of this fact, thus,

> If the epistemologist goal is validation of the grounds of empirical science, he defeats his purpose by using psychology or other empirical science in the validation.
>
> (Quine, 1969, pp. 75-76.)

Hence, for a Quinean, epistemology instead of seeking a quixotic justification for our most basic methods and presuppositions will search to describe, to explain, to understand, via empirical science, the origin of our beliefs and the conditions under which we take them to be justified. In particular, it will seek to do this for our scientific beliefs and methods. Thus,

> If we are out simply to understand the link between observation and science, we are well advised to use any available information, including that provided by the very science whose link with observation we are seeking to understand.
>
> (Quine, 1969, p. 76.)

Epistemology thus becomes part of natural science, in the sense that the only legitimate epistemological questions are questions answerable or resolved by scientists using the methods of the empirical sciences, and any other epistemological questions are seen as traditional idle philosophical queries.

In this manner, Quine is trapped in a web of belief since he tries to »improve, clarify, and understand the system from within,» he is a prisoner of one of many possible Neurath boats. He takes for granted, as the localist that he is, his scheme's background knowledge (in particular, his scheme's methodological assumptions), as well as the assumption of no large-scale errors or gaps in his scheme. Furthermore, when he claims that science is where, «judgment is properly passed,» Quine is making an unjustified normative claim.

It may be retorted that our demands of justification for what Quine takes to be scientific method means that we doubt this method, and that these doubts must be insincere, because to doubt the hypothetico-deductive method is impossible psychologically,[4] or as Peirce would have said, because we cannot doubt it «in our hearts.» The answer is that the psychological impossibility of these doubts is irrelevant, because the important question is whether these doubts are logically cogent. This was the point made by Hume about our almost irresistible inductive psychological propensities, propensities that nevertheless lack logical justification, so Peirce's advice is misdirected, since it doesn't distinguish the psychological context from the logical one. Moreover, Peirce's recommendation is itself unjustified, if not, why should one rest contented with only local or particular questions?

The pragmatist might answer: 'because local problems are solvable, while global, ultimate ones are insolvable', the pragmatist's injunction would then be: if you want to be means/ends rational, then deal only with solvable problems.[5] The pragmatist will insist that to ask for

---

[4] 'Creation scientists' who propose an alternative scientific methodology exemplify, though, that these doubts are not psychologically impossible.

[5] Instrumental or means/ends rationality allegedly advices that, *if one wants to be rational, and if one wants goal A, then one should look for the justified optimal means, amongst those available to us, to attain or continuously approximate the desirable and sought goal A.* Therefore, if A is an impossible and not continuously approximable goal, then there won't be any means available to attain or get close to A, and

justification all the way down to the 'bedrock' is unreasonable, that it is unreasonable because methodological bedrock non-viciously circular justifications cannot be provided. In other words, the pragmatist recommends: don't ask what cannot be provided, stick to fruitful local questions, such as those of science.

Two points now demand further analysis:

1) Does a localist-naturalist meta-methodology intend only to describe and explain how scientists proceed when revising their scientific theories and methods?

2) Does a naturalist theory of scientific method recommend that we follow a piecemeal procedure?

Does a localist-naturalist meta-methodology intend only to describe how scientists proceed when revising their scientific theories and methods?

If a naturalist scientific meta-methodology were merely to describe the methods and goals of some of the sciences, or if it were just to describe how scientists proceed during their scientific investigations, then this would not be enough to answer our traditional epistemological questions, since we also want to know if scientists *ought to* follow any of their described research strategies.

The normative aspect of methodology is illustrated by the fact that in the past methodologists have criticized some aspects even of the leading scientific theories of their time, they criticized them because these past dominant scientific theories failed according to these methodologists' canons. For example, Einstein *qua* methodologist thought that scientific theories should be deterministic even though quantum mechanics (the dominant theory in its field) is — at least *prima facie* — not deterministic.

Furthermore, if we were to try to get the 'ought' of normative epistemology from the 'is' of psychology (or some other empirical science) we would be suspect of committing the 'naturalistic fallacy' (a fallacy famously discussed for example by G. E. Moore.) Thus, a purely descriptive approach is not possible, because description is itself a cognitive activity with a normative dimension. For example, one selects for description those theories with characteristics that one considers as desirable in a scientific theory.[6] The theories chosen as archetypical scientific are deemed worthy of description, and from this description, the naturalist hopes to infer methodological standards. In sum, the naturalist requires — if rational — of some methodological standards to select his substratum of putative scientific theories, and then from a descriptive study of this substratum he infers methodological standards, the whole process is circular, and so the naturalist ends getting only those norms that he started with. Otherwise,

> In a naturalistic epistemology, theories are selected as scientific if they vindicate entrenched methodological assumptions; and we decide which methods to accept in accordance with a descriptive study of those selected theories[7].

---

then A would be means/ends irrational.

[6] These desirable theoretical characteristics could be known only tacitly.

[7] This is a modification of the following quote,

> [In a naturalistic epistemology] methods are accepted if they vindicate entrenched theoretical assumptions; and we decide which theories to accept in accordance with accepted methodological standards.

If not, why not describe the work of creation scientists and from an analysis of this description infer scientific method? Since this last option will be very likely considered unpalatable the naturalist would have to justify his selection of putative scientific theories, that is, he would have to justify the methodological canons that led him to his theoretical selection. Hence, Quine should tell us why what he considers as archetypes of the sciences are genuine sciences. Quine, however, claims that,

> Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. **For me normative epistemology is a branch of engineering.** It is the technology of truth-seeking, or, in a more cautiously epistemological term, prediction. **Like any technology, it makes free use of whatever scientific findings may suit its purpose.** It draws upon mathematics in computing standard deviation and probable error … It draws upon experimental psychology in exposing perceptual illusions, and upon cognitive psychology in scouting wishful thinking … **There is no question here of ultimate value,** as in morals**; it is a matter of efficacy for an ulterior end, truth or prediction.**
>
> (Quine, 1986, *Reply to M. White*, p. 665.) (Emphasis added.)

Again, how does Quine know that what he takes to be «scientific findings» are *bona fide* scientific results? How does he know that «truth or prediction» are valuable cognitive ends? Quine believes he knows this because he surely has applied, even if only tacitly, some methodological standards to decide this matter, and then the normative aspect of methodology creeps in when deciding which results to call scientific. Tacit norms also creep in when Quine decides that «truth or prediction» are valuable cognitive ends.[8]

Quine will probably argue that a descriptive study of the empirical sciences will show that these are their aims, but again since the sciences don't select themselves, how were the sciences selected? If the sciences were selected using some methods efficacious for the attainment of some cognitive ends, then the ends of the sciences were already there, in the methods and ends that helped to select them, thus we end discovering and describing the very same methods and ends that we prejudged are the methods and ends of science.

In sum, Quine's naturalism cannot be wholly descriptive, because a full-bloodied descriptivist naturalism would be incapable of getting started, since all description requires of some methodological standards, or norms, to recognize what is relevant and valuable of description. Or else, a descriptive naturalism requires some super-naturalistic cognitive methods and goals, it requires a vantage point outside science, it needs a moderate first philosophy.

This becomes especially clear, once one realises that even if the successful (say, in pragmatic terms, i.e., the empirically adequate) scientific theories were to somehow select themselves, a couple of questions would remain:

i) That of whether the methods presupposed by these pragmatically successful scientific theories are the *proper* scientific methods, and

ii) That of whether these pragmatically successful scientific theories constitute knowledge.

(Hookway, 1990, p. 223.)

[8] By the way, these are two very different cognitive aims, which of the two is the genuine scientific end? That it is important to decide whether the goal of science is truth or prediction, is shown by the on going debates between realists and instrumentalists, for example, L. Laudan claims that truth is an irrational cognitive goal.

Quine assumes an affirmative answer to these last questions, but by doing so, he is taking for granted, in spite of himself, a prior philosophy: pragmatism.[9]

Does a naturalist theory of scientific method recommend that we follow a piecemeal procedure?

If naturalism is a normative injunction in favour of a piecemeal procedure, that is, if it is an injunction for dealing with problems only when they arise, without questioning entrenched theoretical and methodological assumptions; if naturalism recommends a tinkering localism when dealing with philosophical and epistemological questions (because allegedly this is how science proceeds), then how does naturalism justify his prescription for localism?

The localist-naturalist might answer that such demands for justification are precisely the kind of questions that localism excludes, if so, localism appears as a self-serving prescription. For the sceptic, naturalism is an *ad hoc* strategy that avoids what it cannot answer, thus, the naturalist asks us to,

> … rest content with a policy of piecemeal tinkering whose legitimacy as a way of approaching truth cannot be established. If this is all that can be said, naturalistic epistemology appears to **acquiesce in skepticism rather than try to overcome it**.
>
> (Hookway, 1990, p. 223.) (Emphasis added.)

Again, why should we accept the naturalist's advice: only local questions! Consider that traditionally, it has been considered philosophically legitimate to ask whether scientific procedures as a whole are justified. The localist will likely retort as follows,

> Yes, inquiry s a risky and fragile process, we must to certain extent trust to luck. However, do we have some positive reasons for mistrust in our process of inquiry? Or else, is this justification required for the growth of knowledge? If it is not, should we care to have this justification?

The answer is that this justification is required if we want to know whether our process of knowledge acquisition is reliable. The naturalist could still retort that,

> … science is innocent unless proved guilty while our metascience is guilty unless proved innocent.
>
> (Carnap, quoted by Hookway, 1988, p. 198.)

But, why is the naturalist strategy more prudent? Is this evaluative judgement also going to be taken for granted? The naturalist will likely insist that,

> The ship *keeps navigating*, how it does it, we still don't fully understand, but this is no reason to undermine our hope in its going on and in eventually understanding it

---

[9] Thus, Quine claims that,

> We cannot detach ourselves from [our conceptual scheme] and compare it objectively with an unconceptualized reality. Hence it is meaningless, I suggest, to inquire into the absolute correctness of a conceptual scheme as a mirror of reality. Our standard for appraising basic changes of conceptual scheme **must be**, not a realistic standard of correspondence to reality, but a **pragmatic standard**. Concepts are language, and the purpose of concepts and of language is **efficacy in communication and in prediction**. Such is the ultimate duty of language, science and philosophy, **and it is in relation to that duty that a conceptual scheme has finally to be appraised.** (Quine, 1953, p. 79.) (Emphasis added.)

Hence, even though Quine maintains that we cannot say which conceptual scheme is objectively correct or true (and in this sense none are better), he claims that we can still compare conceptual schemes in terms of their shared aim for efficacy in prediction. In the case of scientific conceptual schemes, this comparison can also be made because Quine believes conceptual schemes also share some basic methodological strategies, such as the hypothetico-deductive method. In other words, Quine believes that all scientific conceptual schemes share some basic methodological and axiological presuppositions, but how does Quine justify this belief of his?

better. It is true that this hope lacks positive reasons to back it (except for a history of some **successes**), but at least it also lacks negative arguments against it, except for the absence of a non-viciously circular justification.

However, how do we know that the ship will keep on navigating? It could well sink at any moment, on the other hand, how does the naturalist know that what he considers a history of some scientific «successes» is really that, i.e., a history of *objective* scientific successes, and not a history of something else.

Otherwise, how does the localist know that our world is a 'normal' one? All we know is that so far our world seems to have been normal, from this to conclude that it has in fact been normal is taking for granted a prejudice. But even if our world has in fact been normal, will it keep being normal? The naturalist's belief in the normality of our world — and in the persistence of this putative normality — may be natural or spontaneous, but so are the sceptic's doubts, this as shown by the fact that these traditional sceptical questions keep on recurring.

For the Quinean our most basic cognitive methods are in no need of justification, what require justification are, instead, the 'unnatural' doubts of the sceptic: to doubt what has served us so well for so long requires a justification. The Quinean holds the following conditional principle P:

P: If it works, then don't justify it, because it doesn't need a justification.

However, how do we go in P from the antecedent to the conclusion? How is this principle going to be justified? Either this principle is an *a priori* prescription, or it can be justified empirically. Now, to justify it empirically we would require of the very same methods (such as the hypothetic-deductive method) that this principle claims don't need justification. The principle is then in the end saying that it itself doesn't need of an empirical justification, then P has the character of a stipulation, of an *a priori* prescription, a character which goes against the Quinean dislike for 'first philosophy'. Moreover, the Quinean appears to advise: Forbidden to ask questions which we cannot answer! Forbidden to question, what we consider obvious! But,

> … believing something to be obvious does not obviate the need to defend it, or at least the need to acknowledge that belief as an assumption … that one makes.
>
> (Worrall, 1999, p. 348.)

The need to justify the obvious becomes especially pertinent when one considers that according to an evolutionary perspective it could be biologically advantageous (energy and time wise) to find obvious what is strictly wrong, but close enough (survival wise) to the truth.

Thus, consider that biological evolution selected our cognitive system for optimal efficiency *vis a vis* promoting biological survival and reproduction in a prehistoric terrestrial environment of middle-sized objects, and that as our investigations take us into the micro and macro cosmos, farther and farther away from our original problem situation, our cognitive architecture could prove insufficient. In other words, it is doubtful that the cognitive capacities that proved adequate to hunt a mammoth will also be sufficient to explore Mars, to do philosophy, and to develop a unified field theory in physics. In this way,

> **A naturalized epistemology begins by setting aside the classical justificatory questions of the adequacy of our knowledge-gathering practices, but ends up providing the basis for a new suspicion that there are deep limits for our knowledge in all but the most implausibly homogeneous and manageable of possible worlds**. Indeed, it would be an odd accident if our subjective canons of scientific

> acceptability turned out to match in all respects the objective character of the universe. Why should our cognitive capacities be adequate for all domains? … We are… unlikely to have entirely correct and complete theories; **our innate cognitive biases may cause us to accept some falsehoods and reject some truths.**
>
> … There seem to be possible worlds that would be too complicated for us or a society of experts to represent feasibly… The breadth and depth of putatively possible knowledge may be intrinsically too great for a both manageable and complete world view…
>
> (Cherniak, pp. 127-9.) (Emphasis added.)

This conclusion becomes more plausible once one recalls that evolution selected those of our ancestors with cognitive capacities correct enough to promote their biological survival and reproduction, and that evolution did not necessarily select reliable truth producing and truth transmitting cognitive capacities even for dealing with the middle sized objects of our ancestor's primeval savannah[10]. Therefore, our contemporary biology undermines two of naturalism's key assumptions:

i) That piecemeal improvement can proceed indefinitely.

ii) That there are no large-scale errors in our conceptual scheme.

The naturalist criticized the sceptic for entertaining unjustified or idle doubts, and ironically we now discover that science itself provides justified sceptical doubts, doubts analogous to those of the old sceptic. Now, to reject these last doubts someone could speculate on,

> … a particular type of cosmology, one that ensures a preestablished harmony of man with the universe. It would be a peculiar coincidence in need of much explanation if, for every domain, every one of the interesting true theories, and all of them together, should just happen to be simple enough to be usable by, and intelligible to, us.
>
> (Cherniak, p. 129.)

And he could go on to make an assumption of *veracitas Dei* (as Descartes and Thomas Reid did) to underwrite his belief in a pre-established harmony of man's mind and the cosmos, to back his hope that our cognitive means are adequate for our cognitive ends. These conjectures, however, will likely be unsavoury to the naturalist, because of their speculative metaphysical character.

IV IS Justification only Argumentative?

The localist-naturalist can still argue that behind the sceptic's doubts examined so far lurks the assumption that justification is only argumentative i.e., the belief that a proposition is justified by inferring it -say, deductively or inductively — from some premisses, and only thus. Therefore, it follows that if there are logical limits to argumentation, then there will be also logical limits to justification. Our sceptic has confined justification only to inferential relations amongst propositions and our sceptic has required that the justified believer have a conscious reason[11] for thinking that his belief is true.

---

[10] This because, natural selection, being interested only in survival, it had to cut corners to save energy and time.

[11] Or at least, the justified believer should have his beliefs justified by reasons that can be made conscious -after adequate self examination or reflection- that is, the justifying reasons should be capable of becoming conscious.

The naturalist, on the other hand, *also* welcomes 'externalist' non-argumentative justifications, such as those provided by, say, some psychological unconscious processes. It is claimed, for example, that beliefs caused or generated by overall reliable truth generating psychological processes (or beliefs transmitted, from previously justified beliefs, by generally reliable belief transmitting processes), in an environment normal for the formation or transmission of such beliefs, are justified.

For this doctrine, *reliabilism*, beliefs would be justified even if the subject were unaware of the belief generating and transmitting processes or faculties going on in his mind, and because of this unawareness of the *justificans* the believer will in general have no reason for thinking that his beliefs are true or likely to be true, but will nonetheless be justified in accepting his beliefs. Examples of possible reliable 'source' processes are perception, memory, reasoning and intuition, while examples of possible reliable 'transmitting' or inferential processes are deduction and induction.

Reliabilism deals successfully with a scepticism concerning observational statements, since for reliabilism observational or basic statements can be justified if they are generated by some reliable non-inferential psychological processes, such as the processes of perception of a healthy subject in a standard situation. While in the argumentative conception of justification, only other statements can justify basic statements, a requirement which leads us into the familiar sceptical quandary: an infinite regress of justificatory statements, and to stop the regress of statements, circularity or dogmatism. Popper, for example, deals with this trilemma by concluding that a form of conventional dogmatism is unavoidable, that is, some basic statements have to be taken as true pro tem by a convention made by a scientific community. A conventional agreement, though, that could be revised and substituted by another conventional agreement[12], if serious criticism of the first conventional basic statement were to arise. Still, the basic statements at which we stop the regress have the character of dogmas in the sense that they are accepted as true — again, even if only temporarily- without an argumentative justification. Popper arrives at this doctrine of his, because he believes that statements can only be justified by other statements, and therefore he believes that psychological processes even if reliable can at most cause or motivate our decision to accept some basic statements, he would say that the reliabilist confuses justification with causation or motivation[13].

> Every test of a theory, whether resulting in its corroboration or falsification, must stop at some basic statement or other which we *decide to accept.*
>
> … The basic statements at which we stop, which we decide to accept as satisfactory, and as sufficiently tested, have admittedly the character of *dogmas*, but only in so far as we may desist from justifying them by further arguments (or by further tests).
>
> … Experiences can *motivate a decision*, perhaps decisively, and hence an acceptance or a rejection of a statement, but a basic statement cannot be *justified* by them — no more than thumping the table.

---

[12] Assuming the scientific community can reach an agreement about which basic statements are not problematic for the time being.

[13] Popper's position can perhaps be clarified *via* an ethical analogy, thus imagine a criminal who would causally explain his crime by showing that his action was the result of an emotional process (say, love or benevolence) that in general, and in standard situations leads, to good actions. Would we say that this causal explanation justifies as good his crime? Many of us would find an affirmative answer as counterintuitive. On the other hand, the adage says: *to know all, is to forgive all*, forgive perhaps, but not justify as good or right.

(Popper, sections 27-9.)

Hence, it seems that reliabilism can avoid an scepticism of basic statements, while the traditional argumentative or internalist doctrine of justification cannot avoid scepticism.

The reliabilist, however, has to deal with sceptical difficulties of his own once the sceptic asks for a justification of the beliefs in the reliability of the so-called reliable processes. Thus, the justification for the beliefs about the reliability of some processes will be provided by some other belief generating 'reliable' processes, and to stop a regress the reliabilist, like the argumentative internalist before him, will end with circularity or dogmatism concerning the reliability of some process. The reliabilist will stop the regress by arguing that our cognitive processes, such as our inductive processes, are reliably self-supporting,[14] or by arguing that various of our cognitive processes are supported by some more basic self-supporting cognitive processes, or by arguing that our cognitive processes mutually (i.e., circularly) support each other,

> An important component of a reliabilist theory of knowledge would surely be a list of reliable faculties: perception, memory, introspection, inference, and perhaps others. But how could one justify the addition of a faculty to the list except by use — direct or indirect- of that very faculty? **And is that not as viciously circular as declaring a source reliable by accepting its reports at face value and inferring that it issues truth? Such reasoning is unreliable and in any case unacceptable.** We may perhaps avoid vicious circularity by allowing a faculty to gain support from the use of other faculties. But these would need support of their own and how could they gain it except by each leaning on the others? Reliabilism is thus driven to seek **refuge in a wide enough circle, which it must regard as benign, perhaps in virtue of its wide diameter.**

(Sosa, E., p. 95.) (Emphasis added.)

However, both a viciously circular argument with a wide diameter and one with a small diameter are equally logically unacceptable, if there is any difference between these two circles it would be just a matter of psychological obviousness. The wide diameter circle may be regarded as «benign» (i.e., as a bona fide probative argument) only because its circular ness remains hidden, only because its fault is not apparent, but if so, this looks as a deceptive or hypocritical strategy, it looks like a simulation game.

For example, assume that one has belief *B* that our memory has been in general a reliable belief producing cognitive process. Now, if someone asks for a justification of *B*, we could justify it by saying that belief *B* is generated by our memory cognitive processes. That is, we would justify *B* by invoking our memory — i.e., circularly — and if in addition, we were to infer that our memory cognitive processes will probably continue being reliable, we would have to assume also that our inductive cognitive processes are reliable.

Moreover, the reliabilist assumes that a belief *B* is justified in case cognitive processes that are in general reliable produce *B* (or transmit *B* from other justified beliefs.) Now, if the reliabilist is in turn going to justify his theory of justification he will argue either:

i) That the reliabilist theory of epistemic justification is justified because possible overall reliable cognitive processes, such as reasoning plus imagination, generate the reliabilist's theory of justification. It is, however, problematic to argue that reasoning and imagination — once taken beyond our strongest intellectual intuitions- are by themselves in general reliable belief generating cognitive processes, given that it is almost a truism that reasoning and imagination have often lead us into absurd theories or beliefs.

---

[14] Cf., Papineau for a reliabilist rule circular justification of induction.

Or,

ii) The reliabilist will end up with an argumentative internalist justification of his theory of justification, an argumentative justification that will ultimately lead again into the sceptical trilemma, of infinite regress, circularity, or dogmatism. Hence, in the end, the reliabilist finds himself in the same sceptical muddle from which he tried to extricate himself.

References

Cherniak, C. 1986: *Minimal Rationality.* The MIT Press, USA.

Hookway, C. 1988: *Quine.* Polity Press, U. K.

Hookway, C. 1990: *Skepticism*. Routledge, London & New York.

James, W. 1890: *The Principles of Psychology.* Dover, New York, 1950.

Papineau, D. 1992: «Reliabilism, Induction and Scepticism», *The Philosophical Quatterly* Vol. 42, 166, pp. 1-20.

Peirce, C. S. 1868: «Some Consequences of Four Incapacities», In *Philosophical Writings of Peirce.* (Ed. Justus Buchler) Dover. New York, 1955.

Popper, K. 1959: *The Logic of Scientific Discovery.* Harper and Row and Basic Books.

Quine, W. V. O. 1953: «Two Dogmas of Empiricism» In *From a Logical Point of View*. Harvard University Press. Cambridge, Mass. Second Edition, 1980.

Quine, W. V. O. 1953b: «Identity, Ostension, and Hypostasis» In *From a Logical Point of View*. Harvard University Press.

Quine, W. V. O. 1960: *Word and Object.* MIT Press. Cambridge, Mass.

Quine, W. V. O. 1969: «Epistemology Naturalized» In *Ontological Relativity and Other Essays*. Columbia University Press. New York.

Quine, W. V. O. 1975: «The Nature of Natural Knowledge» In *Mind and Language*. (Edited by S. Guttenplan.) Clarendon Press. Oxford.

Quine, W. V. O. 1981: «Five Milestones of Empiricism» In *Theories and Things*. Harvard University Press. Cambridge, Mass.

Quine, W. V. O. 1986: «Reply to Morton White» In *The Philosophy of W. V. O. Quine*. (Edited by L. E. Hahn and P. A. Schilpp). Open Court. La Salle, Illinois.

Quine, W. V. O. 1992: *Pursuit of Truth.* Harvard University Press. Cambridge, Mass.

Sosa, E. 1991: «Nature unmirrored, epistemology naturalized» In *Knowledge in Perspective. Selected Essays in Epistemology.* Cambridge University Press. UK.

White, M. 1986: «Normative Ethics, Normative Epistemology, and Quine's Holism», (1982) In *The Philosophy of W. V. O. Quine* (Edited by L. E. Hahn and P. A. Schilpp). Open Court. La Salle, Illinois.

Worrall, J. 1999: «Two cheers for Naturalised Philosophy of Science — or: Why Naturalised Philosophy of Science is Not the Cat's Whiskers.» In *Science and Education* 8: 339-361.

Armando Cíntora
Department of Philosophy
Universidad Autónoma Metropolitana in Mexico City
<cintora@prodigy.net.mx>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 57-62
Cartesianism and the private language argument
Copyright © by SORITES and Brian Garrett

CARTESIANISM AND THE PRIVATE LANGUAGE ARGUMENT

Brian Garrett

<Brian.Garrett@anu.edu.au>

1.

One of the main reasons for interest in the «private language» debate is the assumption that the impossibility of such a language has significant implications for the philosophy of mind. In particular, its impossibility has been thought to undermine the Cartesian Model of how natural language sensation words get their meaning.[1] According to this model, a sensation word, such as 'headache' or 'tickle', gets its meaning in virtue of an act of 'inner' association or ostensive definition. Even though you and I use the same English word (e.g., 'headache') what I mean is defined with reference to my headaches, and what you mean is defined with reference to yours. I argue that this model is not undermined by the private language argument. Further, this model is normally thought to imply that the meanings of sensations words, so defined, are logically private, intelligible only to their user. I argue that this assumption is false.

2.

First some comments on Wittgenstein's private language argument(s). At #202 of the *Philosophical Investigations* Wittgenstein offered a condensed version of the private language argument explicitly presented at #258. At #243, the issue of private language is raised again. It is crucial to be clear about the sense of 'private' in this and other passages. Wittgenstein's concern is with whether there can be a logically or necessarily private language, a language necessarily intelligible to only one person.[2] The way Wittgenstein approaches the issue is to ask whether someone could invent a language in which words «refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language.» (#243)

One point needs to be noted about this passage (and here I am indebted to E. J. Craig's excellent «Meaning and Privacy»[3]). Craig argues plausibly that the 'So' actually denotes a

---

[1] So-called for good reason. In the opening pages of the *Second Meditation*, Descartes makes it clear that he could not acquire the ideas of thought or sensation from contemplation of bodies (either his own or those of others). He wrote: « … according to my judgement, the power of self-movement, like the power of sensation or of thought, was quite foreign to the nature of a body; indeed, it was a source of wonder to me that certain bodies were found to contain faculties of this kind.» (R. Descartes *Meditations on First Philosophy* (ed.) J. Cottingham (Cambridge University Press, 1996) pp. 17 — 18) On such a view, introspection is presumably the only way in which one might give meaning to one's mental words.

[2] Wittgenstein concedes the possibility of contingently private languages: «[w]e could even imagine human beings who spoke only in monologue» (#243).

[3] In B. Hale & C. Wright *A Companion to the Philosophy of Language* (Oxford, Basil Blackwell, 1997)

fallacious inference (p. 128). From the fact that a language refers to «what can only be known to the person speaking» it does not follow that «another person cannot understand» it. This follows only on the assumption that A understands B only if A knows the objects B is referring to, and Craig rejects this assumption (pp. 130 — 1). Craig takes Wittgenstein to have unwittingly given two definitions of 'private language' (*viz.*, «what can only be known to the person speaking» and «what is intelligible only to the person speaking»), and regards the former definition as the more fruitful one. The interesting question then becomes whether a public language might also be private, in this sense of 'private'.

3.

Section #257 has been thought by some to harbour a self-contained argument against private language.[4] There Wittgenstein imagines a putative private linguist, who presumably has never acquired a public language, attempting to define his sensation words by inner ostension. He writes: «When one says 'He gave a name to his sensations' one forgets that a great deal of stage-setting is presupposed if the mere act of naming is to make sense.» However, three reasons tell against this 'stage-setting' argument.

First, the argument is at best suggestive. We need to be told *how* it is possible for us to have or acquire the necessary stage-setting and, correlatively, *why* it is impossible for a private linguist to have or acquire such stage-setting.

Second, it is generally agreed that a socially-isolated-from-birth Crusoe, alone on his desert island, can, if sufficiently ingenious, name and describe aspects of his physical environment. But, in that case, he must have in place whatever stage-setting is required for naming, and it is hard to see why such stage-setting cannot be available to the private sensation linguist. If the requisite stage-setting does not require the presence of others, as the Crusoe case demonstrates, why should the private sensation linguist be debarred from the practice of naming and describing his sensations?

Third, the stage-setting argument has nothing to say against someone already in possession of a public language choosing to ostensively define new sensation words, since such a person obviously has the required stage-setting.

4.

It is #258 which most commentators think of as containing 'the' Private Language Argument:

> Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign «S» and write this sign in a calendar for every day on which I have the sensation. — I will remark first of all that a definition of the sign cannot be formulated. — But still I can give myself a kind of ostensive definition. — How? Can I point to the sensation? Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation — and so, as it were, point to it inwardly. — But what is this ceremony for? For that is all it seems to be! A definition surely serves to establish the meaning of a sign. — Well, that is done precisely by the concentrating of my attention; for in this way I impress on myself the connexion between the sign and the sensation. — But «I impress it on myself» can only mean: this process brings it about that I remember the connexion right in the future. But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'.

A reconstruction of this argument against private language might run as follows:

---

[4] For example, H-J Glock *A Wittgenstein Dictionary* (Oxford, Blackwell, 1996), p. 312.

(1) In any (possible) language, there must be an is right/seems right distinction;

(2) In a private language, no such distinction can be drawn; so

(3) There cannot be a private language.

The idea behind (1) is presumably the unobjectionable one that meaning is normative. To say that a word has meaning is thereby to say that there are (or can be) uses of the word which are correct (right) and uses which are incorrect (wrong). What makes, e.g., 'table' a word (part of a language), rather than a meaningless squiggle, is that there are uses of it which are clearly correct, and uses of it which are clearly incorrect. It is just this distinction which Wittgenstein thinks cannot be made out by the private linguist: in trying to give meaning to a sign by 'inner' ostensive definition, «… I have no criterion of correctness. … whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'.» (#258) Hence the rationale for (2).

I think there are many problems with this argument, but here is one which I take to be decisive. As noted above, it is generally agreed that contingently private languages (such as a Crusoe language or a private code) are possible. Does the #258 argument not exclude such languages? That is, is premise (2) not indifferent to whether the private language is contingently private or logically private? After all, our socially isolated Crusoe, who invented his own language from scratch, can only rely on his memory and senses, just like the private sensation linguist. How then can Crusoe make out an is right/seems right distinction? It may be replied that Crusoe does indeed speak a language when he names the trees and mountains around him, and that his language admits of an is right/seems right distinction, in virtue of the fact that we could, in principle, learn his language and so make sense of what it would be for some of his uses to be mistaken.

However, a defender of the possibility of a logically private language can reasonably reply that this response begs the question: it simply assumes that an is right/seems right distinction is available *only* for languages which others can in principle learn. Why should this be conceded? Moreover, if Crusoe speaks a language, he surely does so in virtue of intrinsic facts about him, his world, and his intentions. He does not do so in virtue of the fact that *we might* understand and correct him. His speaking a language cannot rest solely, or even partly, on such modal considerations.

5.

According to the Cartesian Model, each of us defines the meaning of his sensation words in his own case. I define, e.g., 'headache' with reference to my headaches; you define your token of that word with reference to your headaches. Thus sensation words get their meaning in virtue of an act of 'inner' ostensive definition. The arguments of #257 and #258, since unsuccessful, have not undermined the Cartesian Model. Of course, this is not to say that there may not be good objections to the Cartesian Model, at least conceived as an historical account of how natural language sensation words get their meaning. Plainly, children do not typically learn words like 'headache' or 'tickle' in the manner suggested by the Cartesian Model. Learning to make third-person attributions is as much part of coming to learn the meaning of our sensation words as learning to make first-person attributions. Wittgenstein offers other, more sophisticated, objections: for example, how could someone defining 'pain' in terms of the Cartesian Model as much as form a conception of another's pain? «I have to imagine pain which I *do not feel* on the model of pain which I *do feel*» (#302) which is « … a none too

easy thing to do … .» (#302) The typical Cartesian reply — appeal to the Argument from Analogy — is given short shrift (#293).

Despite the Model's historical inaccuracy, answers to the following questions are still open. First, even if the Cartesian Model, as an historical account, is not the whole truth, might it be part of the truth? Second, might the Cartesian Model be (wholly) valid in some possible scenarios?

That there is a residual truth in the Cartesian Model is supported by the intuition that certain concepts (typically sensation and secondary quality concepts) seem to have an essentially experiential aspect to them. And this implies that, e.g., a colour-blind individual, however competent he may be in the use of the word 'red', could not be said to possess (or fully possess) our concept of red. Wittgenstein expresses the temptation to think that a word like 'red' has both a public meaning and a private one as follows: « … the word 'red' means something known to everyone; and in addition, for each person, it means something only known to him.» (#273)[5] Such intuitions are not universally shared amongst philosophers, but they are plausible. Given this, there may be some residual truth in the Cartesian Model.

What of the second, modal, question? Having criticised the #257 and #258 arguments, it's hard to see what objection there could be to a competent English speaker choosing to use the Cartesian Model in order to name and describe his sensations, perhaps for entry in a confidential diary. Indeed it's hard to see what objection there could be to a socially-isolated-from-birth Crusoe naming the mountains and trees around him, and then using the Cartesian Model to define his sensation words. If he can single-handedly generate norms in his thought and speech about his physical environment, and so have in place the necessary stage-setting, why can't he do the same with regard to his mental life?

6.

Whatever the verdict on the Cartesian Model, it is important to note a key assumption, made by Wittgenstein and most parties to the debate: that sensation words defined in terms of the Cartesian Model have logically private meanings. That is, it is assumed that the private sensation language of #258 is, or purports to be, a logically private language.

But is this assumption correct? Why think that the diarist of #258, if he speaks a language at all, speaks a logically private language? Two reasons are normally given. First, an epistemic reason: we can never *know* what meanings the diarist attaches to his words; so we cannot understand his words; so his words have logically private meanings. Second, a metaphysical reason: since sensations are private (i.e., unique and unshareable), and since our diarist defines his sensation words with reference to his own sensations, it follows that his sensation words have logically private meanings.

However, neither reason is compelling. As mentioned above, Craig makes plausible the idea that it is not a constraint on A's understanding B's words that A *knows* the objects B is referring to. Although Craig does not make this point in the context of criticising the #258 argument, it is relevant here. If the diarist names his own sensations, and I name mine too, there is no good reason to think that we cannot mean the same thing: «[t]here is no particular

[5] Of course, Wittgenstein is not urging us to yield to this temptation. See also # 315 where Wittgenstein asks «Could someone understand the word 'pain', who had *never* felt pain?»

reason to think that to understand correctly one must know that one understands correctly.»[6] And if it is possible for the diarist and me to mean the same thing, the meanings of his words cannot be logically private.

Second, John Cook goes to great lengths to argue against the idea of sensations as private objects.[7] The standard account of sensations assumes that two people cannot literally have the same sensation, however similar (in duration, intensity, etc.) those sensations might be.[8] According to Wittgenstein, this account misconstrues our language-game with sensations: « … it is … possible for us both to have the same pain.»(# 253). On the standard account, the sentence 'I have the same headache as my father' (where we both have a headache of exactly similar duration, intensity, etc), understood literally, expresses an impossibility. What we should say is: my father and I have similar, but distinct, headaches. It is this piece of reasoning which, in the name of Wittgenstein, Cook wants to ridicule. He suggests that this reasoning is as wrong-headed as the following: «the sentence 'I have the same build as my father' (where we have an exactly similar build), understood literally, expresses an impossibility. Builds are, literally, private. What we should say is: my father and I have similar, but distinct, builds.»[9]

According to Cook, the just mooted line of reasoning is a travesty, contrary to the 'grammar' of the word 'build'. In the case of builds, unlike that of tables and chairs, there is no space for a distinction between 'same F' and 'exactly similar but distinct Fs'. There is no (strict or literal) sense in which, though exactly alike in build, I fail to have my father's build. In the only sense there is, I have the same build as my father, which I can also express by saying that we have similar builds. It's not as if I have something (*my* build) which my father cannot have, however physically similar to me he may be. Builds are simply not private in that way.

Cook's suggestion is that the conception of sensations as logically unshareable, gained currency because the language-game with sensations was wrongly assimilated to that for ordinary physical objects (cf #253), with the latter's firm distinction between numerical and qualitative identity. A better model of the language-game for sensations is that for builds (or character, gaits, senses of humour, etc.). Though couched in the terminology of 'language-games', this is a conclusion about the nature of sensations, independent of the issue of private language. In which case, our diarist of #258, if he speaks a language at all, would not be speaking a logically private language, since its objects would no longer be thought of as logically unshareable. He would be speaking a contingently private language.

7.

In sum, neither the #257 argument nor the #258 argument undermines the coherence of the Cartesian Model. That model has thus not been refuted, and may indeed be a valid model

---

6 Craig, *op. cit*., p. 131.

7 «Wittgenstein on Privacy» in G. Pitcher (ed.) *Wittgenstein* (Macmillan, 1966)

8 Cook (*op. cit*., p. 300) cites Ayer as a typical defender of the standard view. Ayer writes: «it is logically impossible that one person should *literally* feel another's pain» (*The Problem of Knowledge* (Edinburgh, 1956), p. 202.)

9 Cook, *op. cit*., pp. 301 — 2.

in certain circumstances. In addition, the key assumption that the diarist's language of #258 is logically private was called into question.

Brian Garrett
Australian National University
<Brian.Garrett@anu.edu.au>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 63-69
A Trio on Truth
Copyright © by SORITES and Herbert Hrachovec

A TRIO ON TRUTH

Herbert Hrachovec

<Herbert.Hrachovec@univie.ac.at>

HENRY: Yesterday, looking at the log-files of an electronic discussion group I am running, I was in for a surprise. The president of the United States and his First Lady had subscribed. It was a very brief moment of elation, obviously. The notifications I had received could not have been true, even though they contained the correct adress of the White House and seemed perfectly ordinary messages. This hoax was just too obvious. But imagine Mr. Frank Costello subscribing as Mary Cooper. I would not have noticed at all. It is, in fact, well known that there is a lot of identity switching in Internet Relay Chats and MUDs. Chances are that Selma, your exiting contact in Cyberspace, is some male teenager making fun of you. Truth seems just off the point under those circumstances. Perhaps Baudrillard is right: it has been substituted by makebelieve and we should stop worrying and love the bomb.

THOMAS: Don't let yourself become confused. Baudrillard is but a hacker within theoretical discourse. You should trust him just as much as you rely upon those pranksters that faked President Clinton's subscription. Look at your instinctive reaction: You recognized more or less immediatly that someone was cheating here. This presupposes some reliance on the notions of correctness of procedures and well-foundedness of utterances on your part. One has to start from an understanding of truth to be able to complain about a lie. Simulation theory is wilfully exploiting our intuitions about truth in order to turn them into paradoxes. This move looks convincing under certain circumstances, but it is a cheap trick nevertheless. Deception derives its meaning from knowledge. Take away the possibility of knowledge and deception loses all its content. There is no way to avoid a theory of truth if you want to explore the philosophical implications of your experience as a listowner.

JUDY: Maybe Baudrillard went a little too far, but Thomas is certainly much too conservative. Let us agree on his claim that some account of truth is called for. This does not automatically imply the grasp of an elaborate set of foundational concepts. Thomas is talking about «correctness», but how can this be spelled out? I hope he doesn't believe that there are facts «out there» and truth consists in «getting them right». This kind of intuition is quite common, I agree. But it rests on extremely dubious theoretical foundations. How do you pick out states of affairs prior to describing them using true sentences? It seems that truth is just an honorary name we attach to utterances that serve a certain purpose, i.e. to guide us reliably through whatever environment we find ourselves in.

HENRY: Do you want to say that a sentence can be true on the Internet even though it contains an incorrect reference? Imagine I never find out about the spuriousness of Mary

Cooper. All linguistic actions directed towards this construct turn out to work in one way or the other. Would I be entitled to the claim that statements relating to Miss Cooper are true? This sounds extremely odd. I agree with Judy that there has been a lot of ideological distortion of the notion of truth and we have admittedly not yet agreed on any of its features in this discussion. But wouldn't it be fair to just abandon the use of this term rather than twisting it into a completely counter-intuitive construction? If I understand Judy, sentences can be more or less successful — and that is it. One might complain that this is not the whole story, but it is at least a coherent position. This does not seem to be the case with Baudrillard and his followers. They are invoking the whole set of classiscal intuitions only to subvert them. I do not deny that our language permits this strategy. Puzzles and paradoxes are sometimes instructive and amusing. But if there is any argumentative force in Baudrillard's utterances it has to come from their being true on some level. One should rather avoid the predicate if one is not prepared to own up to this.

THOMAS: Actually, big concepts are often disposed of in this somewhat disingenious way. After having discovered that there is no God, Freedom or Subject philosophers tend to propose substitute uses for those empty expressions. You can, for example, retain the wisdom of the Bible and yet refuse to follow its prescriptions. Turning «true» into a predicate on a par with «convincing» or «successfull in the long run» is an example of the same strategy. I regard this as conceptual cosmetics. There has, of course, to be a mechanism to challenge unquestioned metaphysical notions. But one should not disregard the basic intuitions at work here. Using the term «true» I want it to have a distinctive, non-derived meaning. Donald Davidson's scenario of «Radical Translation» is extremely helpful here. He envisages a person confronted with a group of strangers that speak an entirely foreign language. How would such a person react to their prompts? How, in fact, do we proceed with Maya languages or encrypted code? Davidson's answer: By tentatively reconstructing the chain of signifiers into an ensemble of sentences which are then checked against our own perception of reality and our interpretations of the foreign idiom. We employ our mastery of language, including an ability to proclaim a sentence as true, in order to understand foreigners or crack a code.

JUDY: Both of you have taken me to task for advocating a thoroughly pragmatic approach to truth. According to your opinion truth should not depend on the passing fortunes of particular linguistic environments. So it is quite a surprise to have Thomas praising Davidson's account, which puts truth explicitly into the context of particular persons, speaking a particular language. Davidson has recently rejected classical correspondence theories of truth. His reasons are precisely the reasons I have given before. A term's reference cannot be ascertained independently from the sentences in which it occurs. We have to handle terms according to our given understanding of the environment and there cannot be any guarantee that we are parsing the sentences — and consequently conceptualizing the world — in one uniquely valid way. Inscrutability of reference and indeterminacy of translation combine to make classical truth impossible. Davidson is actually quite close to hermeneutics, which is a notoriously relativistic enterprise. Trying to develop a maximally consistent picture of manifestations of unfamiliar agents, given a set of restricted and frequently puzzeling inputs, can be regarded as the job description of any cultural anthropologist.

HENRY: Your summary of Davidson's position is correct as far ays it goes, but you forgot to mention a crucial feature. Let me bring this out by returning to our previous example

regarding the internet as an alien life-form. Many confusing expressions are produced on terminals. There is no way to know, initially, whether this is the operating system's ordinary routine, an error report or an e-mail message. Starting from the presumption that at least some of those cyphers can be treated as names of real persons a person may begin to make sense of a number of strings. This, in turn, allows the agent to construct increasingly larger chunks of the output on the terminal as meaningful communication. This is standard Davidson so far. But now imagine an insidious hacker who has designed a program which constantly juggles around name-like expressions. The minute you begin to form a hypothesis concerning the existence of Frank Costello this name is changed to Mary Cooper in all the relevant contexts. And once you conjecture that there has been a substitution another switch occurs. You would feel frustrated, that's for sure. But there is a deeper lesson to be learned here.

TOMAS: Your point seems to be a methodological one. Just as truth without extralinguistic warrants is but a shadow of its former self, naming would be severely impaired under those circumstances. One has, in fact, to put it stronger. It would be entirely misleading to call such exertions «naming». There is a semblance of meaning here, but it derives from external contexts that allow for a language-game like «naming» to be established. In other words: Something has to work before we can even begin to think about discussing possible effects of those disruptive activities. That's why I am so weary about considerations exploiting the ordinary meaning of «truth» to turn it against itself. Davidson's way of putting this is to say that most of the utterances directed towards our environment must be true. I read this not as an empirical estimate but rather as some kind of transcendental claim. «Truth» is built on a propositional activity that cannot effectivly be relativized within the respective language. Removing this feature does not change or extend the concept; it just destroys it.

JUDY: And what would be so terrible about getting rid of truth-talk? I notice that none of you has come up with an account that supports «truth» in the traditional metaphysical or epistemological fashion. Thomas seems to be insisting on some foundational surplus of the truth predicate, but he has failed to show us how it differs from my pragmatic understanding of our communicative activities. We presume some sentences to be true, i.e. to provide reliable guidance for our dealing with the world. So what? Those sentences do their work for us and we might feel that we would not like to miss them. But this is a contingent decision nevertheless. Another person, or I myself, at a different point in time, may arrive at dramatically different conclusions. You cannot have it both ways, namely in effect abandon the classical theory by taking singular language games as your point of departure — and retain the reassuring objectivity provided by a «God's eye» position.

HENRY: There is indeed some disagreement on how to classify Davidson's position. Richard Rorty makes him into a wholesale pragmatist whereas Davidson himself insists on being a realist at various places. I am not interested in exegetical matters here, so I shall propose my own reading of the situation. To put it in a nutshell: I think that one has to have it both ways, even if this sounds like a paradox. The concept of truth is such that it calls for an attempt to put together two lines of thought that seem to exclude one another. Davidson is right in pointing out the unconditional character of truth-talk and we have to square this with Judy's insight, namely that we are perfectly able to relativize it in every single instance. Something is already settled before we even begin to make assertions about Mr. Costello. And yet, none of the sentences referring to this person is

cast in stone. There is no ultimate guarantee that he cannot turn out to be Mary Cooper after all. The challenge is to make those ends meet.

THOMAS: Here is an idea on how to approach this problem in the Tarski-Davidson tradition. Tarski taught us one extremely important thing about the conceptual location of truth: It has to be situated at the crossroads of object-language and meta-language. Forget, for the moment, all the elaborate details of Convention T and recursive definitions of the T-predicate. Everything hinges on his decision to operate within a two-languages framework. Truth is a predicate that cannot be treated adequately by restricting oneself to one of those levels. To treat it adequately you must not get stuck within the one language you happen to speak. If you do, the need to talk about truth will never arise. There will be no occasion to rise above your statements and proclaim them to have a certain relationship towards the world. But you cannot completely immerse yourself in a meta-language either, since this language is defined by reference to an outside idiom. You may attempt to incorporate the object language into the meta-language. There has, however, still to be a mechanism for distinguishing between the two languages. In the absence of such a distinction the metalanguage collapses into any language this side of the problem of truth. «To make ends meet», as Henry has called it, one would have to exploit the changeover made possible by Tarski's setup. Combining unconditionality and relativity begins to make sense if one reflects on the fact that switching languages has been built into the problem from the very start.

JUDY: This business about the unconditional still escapes me, but I find some of your remarks quite interesting. You are apparently moving in my direction when you describe the impact of truth by means of linguistic strategies. Proceeding from Tarski's assumptions your observations strenghten the case of playing down all the venerable claims usually raised in connection with Truth. One might deconstruct the classical intuition in the following way: unconditional truth is a certain kind of phantasma arising from the careless transition from one given language to a meta-language. Imagine someone who has taken one step up from a given discourse. We can see her holding on to an impossible postulate, expressing an untroubled confidence into some of her utterances, knowing quite well that this urge is triggered by going beyond the realm of unconditional confidence in the first place. It is like reflecting on mutual trust among partners. Once you make an issue out of it, there have been reasons for mistrust. So I agree that your two-language framework throws some light on how confusing truth-talk can turn out to be. But I reject the metaphysical approach to trust as well as truth. We are beyond any simple origins once we talk about those subjects.

HENRY: No doubt about it. I accept your last sentence as it stands. But certain essential parts of the story have again been skipped. You insist on our being at a distance from unconditional uses of language. And there is in fact a disturbing tendency for people to become dogmatic when their initial assumptions are put into question. They often lack the resources to adapt themselves to a more flexible view of the issues considered. Arguing against rightwing politicians I use precisely your kind of argument. But there is more to this, and it is important to bring it out for methodological as well as for political reasons. It's quite simple, really and we touched on it before. No amount of second-level considerations will spare you the burden of establishing first-level confidence. You can question each of your trusted assumptions at one time or the other. You can also question some of them all the time. But questioning all of them all of the time is not an option. It is insanity.

THOMAS: We have been here before. So let me offer some clarification before we go into an argumentative loop. Up to now we have compared intuitions about the unconditional and about relativity somewhat superficially. In arguing for the former I have to be more precise. What exactly is being claimed to be unconditional here? Reflecting on our conversation I find that I have not been clear enough about this. There are two quite different contexts to be considered. The first one is a transcendental argument (sort of). Truth demands unquestioned success of some communicative behavior, any other way lies madness. This seems irrefutable to me. There is, however, a second line of thought and things are more complicated there. Analyzing transcendental qualities of truth-talk should not make us forget that it is sentences that are pronounced to be true. And there is serious doubt whether there can be a transcendental argument for any particular sentence to be true. So we are facing a gap here. A rather speculative consideration on the one hand and a matter of actual linguistic practice on the other. There is no easy attribution of properties across the divide. It is a long way from spoken sentences to reconstructing the conceptual obligations incurred by truth-talk. To put it succinctly: How does one derive metaphysical content from formal considerations? I can, admittedly, not name an unconditionally true sentence, even though I have argued in favor of unconditionality.

JUDY: So what's the use of your excursions into transcendental philosophy? The divide you are talking about seems to separate old-fashioned foundational reflections from the investigation of how we actually use our language. And since I know of no proposition that remains true under any circumstances whatsoever I just can't see the point of your insistence on something exempt from the flux of time and place. You have actually admitted that it could not be a sentence. What sort of entity is it, then? There is small comfort in an in-principle-stability that does not render any tangible results. Our discourse is, of course, guided by many regularities and norms. But they are themselves subject to change. The best way to approach them is via pragmatics.

HENRY: O.k., I'll give you a more or less pragmatistic account of how to arrive at ultimatly true sentences. I agree on the following observation: There are no privileged sentences as long as our empirical involvement with the environment is concerned. So I concede, for example, that every single name on the Internet could be missdirected. And you are right to imply that no amount of external metaphysical or transcendental pressure will fix this problem. But we have by now established a crucial distinction in our handling of various discursive disturbances. Errors, disappointments and improvements are *always* preceded by language use within a status quo. You simply cannot call a sentence into question without having placed it within a given framework of understanding and practice. You cannot avoid holding that there is a Mr. Costello before discovering that you may be wrong. Now this initial conviction is obviously not ultimatly true. But note the following: the way this term's use might be challenged is very different from how a new name is introduced. In the latter case I can just add an expression to my lexicon, whereas in the former case my empirical findings might have repercussions on the language employed. I may have to attend to some malfunctions. This activity is quite different from straight empirical usage. «Mr. Costello has written» might be an identifiable string of letters across various contexts. But it triggers a very different response depending on its placement within the variety of discursive practices. Its employment as an actual proposition has to be distinguished from its examination within a corrective loop.

THOMAS: If I understand you correctly you want to mirror my transcendental detatchment into pragmatical language use. You seem to say that recognizing errors and improving on given knowledge demands at least a threefold view on sentences, depending on whether they are used in ordinary circumstances, or challenged, or scrutinized. Given these distinctions it does indeed make sense to insist on a more subtle treatment of truth-talk. In claiming truth for a sentence we would have to allow for various alternative settings. Is a given sentence part of our taken-for-granted linguistic attitudes — or is it singled out for special scrutiny? It is indeed a matter of discussion to what extent this «sentence» remains the same across contexts. Now, this enables me to rephrase my Davidsonian account. No transcendental moves (in the strict sense) have to be invoked to arrive at the claim that the majority of our sentences must be true. It suffices to say the following: The language-game of error-correction can only be played within the more encompassing game of asserting how things are. And this provides us with ultimate truth in the following sense: For every case of doubt an overwhelming number of undoubted propositions have to be in place.

JUDY: Your use of «ultimate truth» is very idiosyncratic here. It just seems to amount to saying that a lot of sentences go unchallenged for lenghty periods of time. This does not give them special status nor does it delineate an area of language-use that is somehow exempt from change. Some things don't change an awful lot, this is all the comfort I get from your remarks. By the way: Do you realize that your argument has a reactionary touch? It can easily be directed against critical inquiry. You are doubtlessly familiar with a certain type of attitude suppressing innovation. Some people feel that things are the way they are just because they have to be that way. On innumerous occasions attempts to change the status quo have been rejected by appeals to tradition masquerading as self-evidence. I am concerned that your emphasis on the more persistent features of our use of language work against curiosity and experimentalism. So, while I am prepared to accept the point that there are several distinctivly different strata within language, I'd still resist any attempts to strenghten the immobility of the most unsophisticated ones.

HENRY: I've seen this coming, but rest assured, your criticism is unwarranted. It is, in fact, one-dimensional pragmatism that is in danger of supporting any given set of power-relations. Thomas' considerations seem much more challenging to me. Coming up with something new and unexpected is quite cheap nowadays. Its the stuff advertisments are made of. Novelty is part of the most entrenched socio-economic order. The line of thought I have been advocating sets different priorities. It explores sentences which carry the weight of some discursive formation. Assertions that cannot be moved around at one's fancy. Think of the following remark: «If 'Costello' is no name I do not know what names are.» Or else: «If this is not a letter by Mr. Costello I don't know what you are talking about.» On such occasions a name or a sentence are doing double duty. They serve as tokens of ordinary expressions and they indicate a point of no return. It's been discussed by Wittgenstein: A certain stick in Paris is a measuring-standard and its length is exactly one meter. Throughout our idioms there are expressions functioning in this peculiar way. I consider it of the utmost importance to find out about this double deployment, testing its power and recognizing how it works in any given case. Substituting a different city: to what extent do the Maastricht criteria define the framework of a common economic approach of the EU — and to what extent are they themselves subject to changing economic strategies within the community?

THOMAS: Truth occurs at the interface of two languages or two usages of a language, this seems to be agreed upon. Now, Judy takes this as an unspectacular event, like switching from face-to-face conversation to talking to a person on the phone. I think there is much more to it. Something goes largely unnoticed in the ascent to the meta-level. That simple talk is often insufficient is in itself a remarkable fact. It is amazing how we manage to develop complicated linuistic strategies. There are doubts, demands for reassurance, guarantees are offered. All of this is worked out between level one and two. And one remarkable question pervades those transactions: How can speakers communicate their straight intentions within a setup that overrules this straightness from the start. This is what the problem of truth boils down to. It has its origins in ordinary language. But communication sometimes breaks down, sending shockwaves through our entire cognitive arrangements. Sentences are stripped of context and seem to be in need of external validation. Quite often it simply works and they are re-integrated in an encompassing linguistic practice. These are the paradigmatic cases of disquotational theories. The explanation of truth for a sentence consists in giving an account of how it is successfully incorporated into another operative language. I'm in agreement as far as this story goes. But I want to add one more prospect. Duplicating languages misses one point: Why are there multiple languages in the first place? Holding on to a distinguished notion of truth is just a way to resist papering over this rupture. Mr. Costello wrote a letter. Let us pursue such sentences in their variety of uses. One of them amounts to saying: If you do not take this for granted we have nothing in common any more.

JUDY: Have you been around before the Tower of Babel broke down? I just want to cope with some of the confusion that its collapse left behind.

HENRY: For me the shock of Babel resonates through language.

THOMAS: I am told that the towers of medieval cathedrals are embellished with elaborate sculptures very high up. No one can see them from the ground. Those artists haven been risking truth.

Herbert Hrachovec
University of Vienna
Department of Philosophy
<Herbert.Hrachovec@univie.ac.at>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 70-84
Futility and the Meaning of Life Debate
Copyright © by SORITES and Brooke Alan Trisel

FUTILITY AND THE MEANING OF LIFE DEBATE

Brooke Alan Trisel

<triselba@cs.com>

The concept of futility has figured prominently in the debate about whether there is a «meaning of life.» Futility is exemplified by the ancient myth of Sisyphus. Sisyphus, it will be recalled, was condemned by the gods to push a large stone up to the top of a mountain, whereupon the stone would immediately roll down the mountain, at which time Sisyphus would again push the stone back up the mountain, only to have it roll down again. This struggle by Sisyphus to get the stone to the peak of the mountain goes on forever. In recounting the myth, Albert Camus writes: «His scorn of the gods, his hatred of death, and his passion for life won him that unspeakable penalty in which the whole being is exerted toward accomplishing nothing.»[1]

Some have argued that the condition we find ourselves in is no different from the one faced by Sisyphus and that this becomes evident when we set aside our life affirming internal perspective and view life from a detached, external perspective. For example, Richard Taylor asserts: «The two pictures — of Sisyphus and of our own lives, if we look at them from a distance — are in outline the same and convey to the mind the same image.»[2] However, Taylor concludes his analysis by arguing that it is best to view life from an internal perspective and that life's meaning is not bestowed from «without,» but comes from «within.»

The notion of futility is also reflected in the provocative writing Ecclesiastes from the Old Testament, as evidenced by the following example: «So I turned to all my works that my hands had done, and my toil that I toiled to accomplish, and, lo, all is vanity and a pursuit of wind, and there is no advantage under the sun» (Ecclesiastes 2.11).[3] The noted pessimist Schopenhauer expresses a similar belief when he writes:

> That the most perfect manifestation of the will to live represented by the human organism, with its incomparably ingenious and complicated machinery, must crumble to dust and its whole essence and all

---

[1] Albert Camus, *The Myth of Sisyphus & Other Essays*, trans. Justin O'Brien (New York: Vintage Books, 1955), p. 89.

[2] Richard Taylor, «The Meaning of Life,» *Good and Evil*, rev. ed. (Amherst, N.Y.: Prometheus Books, 2000), p. 329.

[3] «Ecclesiastes,» *The Anchor Bible*, trans. Choon-Leong Seow (New York: Doubleday, 1997), p. 118. The author of Ecclesiastes remains unknown. Ecclesiastes has traditionally been attributed to Solomon. However, many bible scholars including Seow believe that this is implausible.

> its striving be palpably given over at last to annihilation — this is nature's unambiguous declaration that all the striving of this will is essentially vain.[4]

One who believes that human endeavor is futile will hereinafter be referred to as a «futilitarian,» following the standard usage of this word.

The following two questions will be explored in this article. First, what does it mean when someone claims that «life is futile» and is this claim valid? Second, if human striving is futile, as futilitarians contend, can life still be worth living? Although «meaninglessness» has been explored in great detail,[5] the concept of futility, as expressed in the context of the meaning of life debate, has received very little scrutiny. However, an extensive literature has developed in the field of bioethics regarding «medical futility.» It will be useful to take note of these discussions before addressing the question of whether life is futile.

I. Medical Futility

Physicians have sought an «objective» way of determining when treatment should be withheld or withdrawn from a patient. The concept of medical futility has been proposed as one such method. The physician Lawrence Schneiderman, a proponent for using evaluations of medical futility to guide medical decision making, asserts that «a futile action is one that cannot achieve the goals of the action, no matter how often repeated.»[6] In giving an example of a futile treatment, Schneiderman argues that it is futile to provide nutritional support to a patient who is in a persistent vegetative state and that this is futile «for the simple reason that the ultimate goal of any treatment should be improvement of the patient's prognosis, comfort, well-being, or general state of health.»[7]

Schneiderman indicates:

> The futility of a particular treatment may be evident in either quantitative or qualitative terms. That is, futility may refer to an improbability or unlikelihood of an event happening, an expression that is quasi-numeric, or to the quality of the event that treatment would produce.[8]

Schneiderman proposes two standards for determining when treatment is futile. Under the quantitative standard, which is expressed in probabilistic terms, a treatment is futile «when physicians conclude … that in the last 100 cases, a medical treatment has been useless, they should regard that treatment as futile.»[9] Under the qualitative standard, «any treatment that merely preserves permanent unconsciousness or that fails to end total dependence on intensive

---

4 Arthur Schopenhauer, «On the Vanity of Existence,» *Essays and Aphorisms*, trans. R.J. Hollingdale (London: Penguin Books, 1970), p. 54.

5 See, e.g., Richard Taylor, «The Meaning of Life,» op. cit.

6 Lawrence J. Schneiderman, Nancy S. Jecker, and Albert R. Jonsen, «Medical Futility: Its Meaning and Ethical Implications,» *Annals of Internal Medicine*, 112 (1990), p. 950.

7 Ibid., p. 950.

8 Ibid., p. 951.

9 Ibid., p. 951.

medical care should be regarded as nonbeneficial and, therefore, futile.»[10] He argues that quantitative and qualitative determinations of futility relate to the same underlying notion that the result is not commensurate to the effort. In his words, «The effort is, on the part of the agent, a repeated expenditure of energy that is consistently nonproductive or, if productive, its outcome is far inferior to that *intended* [emphasis added].»[11]

Schneiderman's descriptions of futility, and his proposed standards for judging medical futility, clearly demonstrate that the concept of futility is a combination of and reflects three other more basic concepts: (1) ordinary causation; (2) failure; and (3) the temporal related notion of repetition. Futility is conceived of as a *repeated failure* of the means to «bring about» or «produce» an «intended» effect; the means are persistently insufficient at producing the intended effect.

Many have argued that the notion of medical futility is value-laden and problematic[12] and, therefore, that it should not be utilized as the basis for physicians to make decisions, especially unilateral decisions, about the provision of treatment. Robert Truog writes: «The fact that this concept has appeared in law and policy may seem to indicate that it is clearly understood and widely accepted. In reality, however, the notion of futility hides many deep and serious ambiguities that threaten its legitimacy as a rationale for limiting treatment.»[13]

Since futility is thought of as a failure to bring about the «goal» of treatment, whether or not treatment will be judged futile depends, in large part, on how one conceives of and defines the goal. For example, suppose that a patient has been in a vegetative state for nine months as a result of a serious head injury. The patient's physician has concluded that treatment is futile, and therefore that life-support should be withdrawn, since it is unlikely that the patient will ever regain consciousness, recover a significant amount of physical or intellectual functioning, and be free of intensive medical support. After sharing this conclusion with the family of the patient, the family inquires whether there is *any* possibility that the patient will ever regain consciousness. The physician indicates that this is a remote possibility, but that the patient would likely continue to be extremely debilitated and dependent on intensive medical care. The family concludes that treatment is not futile since there is a possibility, however small, that the patient will regain consciousness if treatment is continued.

Standards, including those used to evaluate futility, are derived from our goals. The «goal» of the family, in the above example, is for the patient to regain consciousness, whereas the goal of the physician is for the patient to regain consciousness and some physical functioning, and to be independent, at some point, of intensive medical care. Since the goal of the physician is loftier (appropriately so, the physician would argue) than the goal accepted by the family, the physician adopts a higher standard than the family for judging whether treatment is effective and should be continued. Consequently, the physician and the family

---

[10] Ibid., p. 952.

[11] Ibid., p. 951.

[12] See, e.g., Baruch A. Brody and Amir Halevy, «Is Futility a Futile Concept,» *The Journal of Medicine and Philosophy*, 20 (1995), pp. 123-144.

[13] Robert D. Truog, Allan S. Brett and Joel Frader, «The Problem With Futility,» *The New England Journal of Medicine*, 326 (1992), pp. 1560-1564.

reach different conclusions about whether treatment is futile. What can be learned about futility, in general, from the deliberations regarding medical futility? I believe there are three things of importance. First, futility is a combination of the concepts of ordinary causation, failure, and repetition, as argued above. Second, the concept of futility is value-laden, as argued by critics of medical futility. Why is this so? It is value-laden because it reflects the notion of failure, which is value-laden. Furthermore, futility is conceived of not just as repeated, failed causation, but as the failure to cause a desired or «intended» effect or to realize a goal. If an effect is intended or desired, or a goal is established and pursued, this suggests that it is *valued*. Evaluative standards are derived from goals which, in turn, are based on values. Through its association with goals, which can vary among individuals in terms of how a goal is defined and understood, futility ends up being a value-laden concept.

The third, and perhaps most important, thing that can be learned about futility from the discussions about medical futility is that judgements about whether an activity is or is not futile will vary according to the level of a person's expectations; the higher one's expectations are, the greater the likelihood that an action or activity will be considered futile.

II. Is Life Futile?

According to William Lane Craig: «If God does not exist, then life is futile.»[14] How do futilitarians reach the bleak conclusion that life is futile? There are some distinctive characteristics of the perspective adopted by futilitarians that predispose them to conclude that «life is futile.» They view life from an external perspective, as if they were looking at it from a distant vantage point.[15] Our internal perspective is naturally limited since we are earthbound and unable to personally experience what happened before we were born or what will happen after we die. The external perspective is not subject to these limits since it involves imaginative thought. As a result, this perspective tends to be very broad in scope (temporally and spatially), extending beyond our lifetimes and this planet to encompass remote times and places. For example, Craig writes: «Compared to the infinite stretch of time, the span of man's life is but an infinitesimal moment; and yet this is all the life he will ever know.»[16]

From this distant viewpoint, a futilitarian sees not just himself or herself, but all of humanity. Consequently, they ask questions about life in general, as opposed to questions about individuals. For example, they ask «Is there a meaning of *life*?» and «Is *life* futile?» instead of personal questions such as «Is *my* life meaningful?» and «Is *my* striving futile?» Because the questions that arise from this external perspective are about life in general, the answers that are given to these questions tend to be generalized to all of life. Thus, a futilitarian concludes that life is futile, not just for a few individuals, but for everyone.

Many of our goals can be achieved as, for instance, when treatment does eradicate a disease, thereby extending a person's life. However, by temporally expanding the scope of our perspective, it raises the threat of futility. For example, if we broaden our perspective far enough to encompass times before and after the existence of human life, we see, as did the

---

[14] William Lane Craig, «The Absurdity of Life Without God,» *The Meaning of Life*, ed. E.D. Klemke (Oxford: Oxford University Press, 2000), pp. 53-54.

[15] For an in-depth analysis of the external perspective, see Thomas Nagel, *The View From Nowhere* (Oxford: Oxford University Press, 1986), esp. pp. 208-231.

[16] Craig, op. cit., p. 40.

author of Ecclesiastes, that «All came from dust and all will return to dust» (Ecclesiastes 3.20).

By reflecting on the thought that we will eventually all return to dust, a picture upheld by modern science, it prompts us to wonder about the following question: Will all of our toil, effort, and striving lead to anything of «significance» and, if not, then why do we continue to expend such effort? Continuing with the types of questions raised by viewing life from a broad, external perspective, Craig asks: «If death stands with open arms at the end of life's trail, then what is the goal of life? To what end has life been lived? Is it all for nothing? Is there no reason for life? And what of the universe? Is it utterly pointless?»[17]

Before the question of whether life is futile can be properly addressed, it is necessary to have a clear understanding of the concept of futility, which can be obtained by reflecting on the following example. Suppose that there is a race, that a runner crosses the finish line, and that the other runners who are still making their way around the track witness this. Is continuing the race «futile» for these runners? *This depends on their expectations* in running the race. If a runner's goal was to finish first, and if this runner would not be satisfied with anything less than a first place finish, then continuing the race is futile for this runner. It is futile because there is nothing that this runner could do to achieve the desired result. Even if this runner sprinted the rest of the way around the track, the goal of winning the race is unachievable and therefore continuing the race is a futile activity for this runner. However, if a runner's goal in entering the race was to improve upon a personal record, to finish in the top three, or simply to enjoy the experience of participating in the race, then completing the race is not futile for this individual. Although someone else has won the race, it remains possible for this runner to realize his or her goal.

«Medical futility» has been criticized because of its ambiguity yet, when compared with the notion of existential futility, it no longer seems that unclear. In the field of medicine, there may be disagreement regarding the likelihood with which a particular treatment will bring about the intended result. However, the treatment is generally well defined and most people understand that the treatment is the means for achieving the «end.» Regarding the goal of treatment for a patient who has a life-threatening condition, at times there may be conflict between physicians and the families of patients about the exact nature of the goal. However, it is understood by the various stakeholders that the goal has something to do with extending and improving the quality of human life.

Thus, in evaluating medical futility, there is at least *some* clarity about what the means and the ends are. However, this is not true regarding claims about the futility of life. Is life a means to an end, as religious oriented futilitarians assume, or is living one's life the end?

Effectiveness is the opposite of futility. If, for example, a treatment is effective at curing cancer, then the treatment is not futile. The question «Is life futile?» is extremely obscure. This becomes evident by reversing the question and asking «Is life effective?» The first response will likely be «effective *at what*?» A treatment is considered effective if it brings about the «intended» result (e.g., alleviates suffering). The question «Is life effective?» and the reverse question «Is life futile?» both presuppose, and misleadingly suggest, that «life» is a means to an end, in the same way that a medical treatment is, and that there is an intended result or purpose of life.

---

[17] Craig, op. cit., p. 44.

Even the vague question «Is philosophy futile?» is clearer than the question «Is life futile?» Philosophy is often thought of as a means for obtaining «ultimate» knowledge. If it is understood as such, then the question becomes whether philosophy can in fact lead to this intended result. However, before this question can be addressed, the goal must be clarified and defined. If one presumes that the aim of philosophy is to attain a *complete* understanding of the relations between subjective reality and the rest of the world or to solve all of the problems in the world, then this person will surely end up believing that philosophy is futile. However, if an individual has more realistic expectations about what philosophy can achieve, then it will likely be concluded that philosophy is not futile.

Schneiderman suggests various reasons why he believes that people sometimes insist on receiving futile medical treatment. One explanation for this, he argues, is that «means are confused with ends,»[18] meaning that people conceive of receiving treatment from sophisticated medical technology as the goal, instead of recognizing that technology is merely the means for achieving the true goals of medicine. If it is believed that the «goal» is to receive treatment, then this can lead one, inappropriately, to pursue this goal with great fervor even though such treatment may not improve the patient's condition and therefore may be considered futile by the physician. In reflecting on life, including the question of whether human striving is futile, an opposite type of misconception often occurs: ends are confused with means. Living one's life is thought of, not as an end, but as the means for achieving some «higher» or «larger» cosmic or divine purpose.

If a person works in an organization, is unclear about the purpose or mission of that organization, and does not see the relation between work activities and the organization's purpose, perhaps because the organization is large or undergoing rapid change, then this individual may become frustrated and feel that his or her efforts are futile. Similarly, if it is assumed that life was created for a reason and is a means to an end, but that it is unclear exactly what that goal is, then this may lead one to conclude that striving is futile. Alternatively, if living one's life to the fullest that one can is thought of as the «end,» then, because this goal is intelligible and achievable, one is unlikely to conclude that «life is futile.» Thus, one's underlying belief about whether living is a means to an end, or an end that is worthy itself, will affect one's conclusion about whether striving is or is not futile.

The words «meaningless,» «purposeless,» and «futile» recur throughout the writings of pessimists. «Meaning» and «purpose» are often used interchangeability as, for example, when someone asks: «Is there a meaning or purpose of life?» Although this question can seem bewildering, using «meaning» and «purpose» interchangeably in this question does not generally result in confusion since these words, in this context, have a similar meaning.

The words «futile» and «meaningless» are also frequently used interchangeably. For example, when Craig uses «futile» in the statement: «If God does not exist, then life is futile. If the God of the Bible does exist, then life is meaningful,»[19] he seems to mean that life would be «meaningless» without God. Regarding the writings of pessimists, Paul Edwards indicates:

---

[18] Lawrence J. Schneiderman, Nancy S. Jecker, and Albert R. Jonsen, «Medical Futility: Response to Critiques,» *Annals of Internal Medicine*, 125 (1996), p. 671.

[19] Craig, op. cit., pp. 53-54.

«They usually speak of the 'futility' or the 'vanity' of life, and presumably they mean by this both that life is not worth living and that it has no meaning.»[20]

Although «meaningless» and «purposeless» have a similar meaning in the context of discussions about whether there is a meaning of life, there is a significant difference in the meaning of «futile» and «purposeless.» Futile means that there is a repeated failure to realize a purpose *that one envisions* whereas purposeless signifies the *absence* of a purpose.

A failure to recognize the difference and to distinguish between futile and purposeless can lead one to conclude, quite mistakenly, that life is futile if it has no purpose. If there is no «objective» purpose of life and life was not created for a reason by a superior being or an «intelligent» universe, as I believe, then there is no way that we could be failing to achieve such a purpose. Therefore, life is not futile in any objective sense independent of human subjective evaluation.

Life could be objectively futile only if there was an objective purpose of life and we were unsuccessful, for whatever reason, at realizing this purpose. For example, if human life was created by a god for the singular purpose of rescuing a dying alien civilization and, despite repeated attempts to do so, we were ineffective at accomplishing this goal, then this god might conclude that human life is futile. But since we were not created for such a purpose, it does not make sense to claim that «life is futile» any more than it would make sense to claim that «life is effective.»

«Life» is not futile, but it is «purposeless» in the sense that it was not created as the means for achieving a cosmic or divine purpose. This does not mean, however, that each of our lives is devoid of meaning or purpose since, as Kurt Baier[21] and others have convincingly argued, one can pursue worthwhile purposes and create meaning *in* life, even if there is no «purpose of life.»

Although «life» is not futile, striving to achieve a particular goal can be futile for an individual. Because the question «Is life futile?» misleadingly suggests that life is a means to an end, it would be better to rephrase the question to ask: «Is striving futile?» It then becomes clear that this depends on *what* one is striving to achieve.

III. Futilitarian Expectations

Since there is no preordained purpose of life, or at least there is no evidence of such a purpose, the only purposes that we could be failing to achieve are ones that *we have conceived*. What are the purposes that futilitarians believe that we are failing to achieve? To evaluate whether there is any merit to the subjective determinations of futilitarians that human striving is futile, we must first have a clear understanding of what their expectations are regarding life. For example, until it was clarified what the runners' expectations were regarding the race in the earlier discussed example, it was not possible to answer the ambiguous question: «Is continuing the race futile?» Analogously, when futilitarians claim that human striving is futile, there is no way of knowing what they really mean by this statement, or of evaluating the validity of this claim, unless they also reveal what their expectations are.

---

[20] Paul Edwards, «Meaning and Value of Life,» *The Encyclopedia of Philosophy*, ed. Paul Edwards (New York: Macmillan, 1967), vol. 4, p. 467.

[21] Kurt Baier, «The Meaning of Life,» *The Meaning of Life*, ed. E.D. Klemke (Oxford: Oxford University Press, 2000), pp. 101-132.

Without this essential information, the claim that «life is futile» is an empty statement. It conveys no information other than suggesting that the futilitarian is not meeting his or her expectations, whatever they are.

Unfortunately, those who claim that life is futile do not state their goals clearly, but only indirectly through the standards they adopt to judge life. Rather than explicitly stating that ϕ and ψ are their desires or goals, they claim that life is doomed or meaningless and/or futile without ϕ and ψ. For example, Craig writes: «If God does not exist, then both man and the universe are inevitably doomed to death. Man, like all biological organisms, must die. With no hope of immortality, man's life leads only to the grave.»[22]

Death is thought of as preventing us from realizing a goal, but what goal? Craig suggests that achieving personal immortality is the goal. Thus, death is preventing us from realizing never-ending life. Since we are «biological organisms,» this goal of living forever is clearly unrealistic. Furthermore, even if personal immortality could be achieved, this would not automatically make a person's life meaningful. Indeed, part of what makes the life of Sisyphus seem meaningless to an outside observer is the fact that it persists forever. Craig recognizes this point, as indicated by the following remarks:

> But it is important to see that it is not just immortality that man needs if life is to be meaningful. Mere duration of existence does not make that existence meaningful. If man and the universe could exist forever, but if there were no God, their existence would still have no ultimate significance.[23]

Taylor indicates: «Activity, and even long, drawn out and repetitive activity, has a meaning if it has some significant culmination, some more or less lasting end that can be considered to have been the direction and purpose of the activity.»[24] In addition to personal immortality, this «significant culmination» is what futilitarians seem to be seeking. They want to be able to look back at life, from the distant vantage point they are viewing life from, and to say that it amounted to something. Life did not simply perpetuate itself for a while and then crumble to dust leaving no traces that it had ever existed.

Even if there was a significant culmination, then what? Would life suddenly be rendered purposeless since the «purpose of life» had been achieved? Thus, from the external perspective adopted by futilitarians, even a significant accomplishment might be unable to eliminate doubts about the significance of life. Furthermore, it is unclear what, if anything, would qualify as a «significant» achievement.

We could construct a space station extending from one end of the galaxy to the other and yet, from an external perspective, this would not be considered «significant» since the space station, like life itself, will eventually be annihilated. Even if we had the grandiose powers sometimes given to us by science fiction writers, and were able to alter the course of the universe — perhaps extending its longevity — it is still doubtful that this would be considered a significant culmination for the following reason. If all of life will end at some point, as it surely will, it is inconsequential how long a cold, dark, and impersonal universe will persist after this point.

---

22 Craig, op. cit., p. 40.

23 Ibid., p. 42.

24 Taylor, op. cit., p. 325.

Indeed, as long as the universe is indifferent to our values and accomplishments, it is hard to imagine any achievement that would be considered «significant» from this external perspective. Consequently, one might be tempted to conclude that doubts about the significance of life would be erased, and that life could be considered significant, if there was a personal god who created life for a reason and who considers life important. Of course, this longing to be recognized by a superior being is reflected in Christian theology where it is written: «So it was; and God saw all that he had made, and it was very good» (Genesis 1.30-31).

To the best that can be determined, the principal goal of futilitarians is to achieve personal immortality. In addition, some religious minded futilitarians seek reassurance from an independent source that life matters and is significant. If these are their expectations, then it is not surprising that they would find human striving to be futile, in the absence of a personal god who has the power to resurrect us from death, since there is no way we can overcome death on our own to realize the goal of living forever. Furthermore, since nature is impersonal, it is unable to declare or affirm that life is of any significance.

What if we received some reassurance from a superior being that life is significant? Would futilitarians then conclude that life is not futile after all? On the contrary, due to lingering doubts about the significance of life, and because of doubts that will surely surface regarding the superior being, such as whether this entity is in fact self-sufficient and unlimited and therefore deserving of respect, I suspect that they would then further *increase* their already unrealistic expectations. Thus, it would then be claimed that human striving is futile unless there is personal immortality and we receive some additional evidence — perhaps a second opinion — that life is significant. If one's expectations are a moving target, as described, and set such that they are always out of one's reach, then all striving will seem futile.

Camus, in a well-known passage, indicates:

> There is but one truly serious philosophical problem, and that is suicide. Judging whether life is or is not worth living amounts to answering the fundamental question of philosophy. All the rest — whether or not the world has three dimensions, whether the mind has nine or twelve categories — comes afterwards.[25]

Suicide is not the only serious philosophical problem, but it certainly ranks among the highest in terms of questions that need to be confronted. One possible reaction to futility is to *give up*. For example, a runner when faced with futility — perhaps after falling during a race — may decide to quit the race since there is no longer any possibility of achieving the desired result of winning the race. With medical futility, «giving up» on the goal of improving the condition of the patient involves withholding treatment or withdrawing treatment if treatment had already been initiated.

C. Stephen Evans, a philosopher and proponent of Christianity, characterizes, inaccurately I believe, existentialism as «the philosophy of despair.» In describing existentialism, he writes:

> The paradigm of despair is the moment when death is certain and final and there is no recourse. Despair as a way of life is living death. It is the knowledge that every action or decision I take is utterly futile. To live in despair is to be alive, but aware that my life is of no more consequence than a corpse. No experience, decision, or accomplishment is of any significance; whether a deed is done or undone is

---

[25] Camus, op. cit., p. 3.

> infinitely trivial. Consistent despair leads to suicide, which is a serious problem for Camus and Sartre, one to which neither gives any ultimately satisfactory answer.[26]

Evans goes on to suggest that the only way to overcome the futility and despair written about by some existentialists is to have faith that God exists. Disputing the reasoning of futilitarians that *all* efforts are futile because death prevents us from realizing the goal of eternal life, Baier argues:

> Surely, there are many things we do that are not futile, that are effective and successful. All is futile only if all our actions aim at eternally perpetuating our lives. But surely this is not so and would be wholly irrational if it were so, given that we know very well that we cannot attain this goal.[27]

If it is believed that all striving is futile because life will end some day, one might consider giving up on life. However, suicide would be an irrational act for a person with such a belief. The fact that futilitarians want life to go on forever — at least those who are religious oriented — demonstrates that they value life and find it worthwhile. Baier makes a similar point when he argues: «The sadness that lies in the thought that life ends in death shows, not that nothing in life is worth having but, on the contrary, that death may bring to an end something that would be well worth continuing; that it must end does not show that it is not worth having while it lasts.»[28]

If life were truly miserable and not worthwhile, then futilitarians would not bemoan death as they do, and certainly would not yearn for eternal life. Since life is valued, taking one's life would be to destroy that which is valued simply because the meaning and benefits (e.g., happiness) that are derived from living will not continue forever. It would be as if one were to assert: «If I cannot have the benefits associated with living go on forever, then I no longer want any benefits.»

Instead of contemplating suicide based on the belief that all striving is futile since life will end, a more rational course of action, and one that should be pursued before judging life, would be to examine one's expectations since, in large part, they determine whether striving will be considered futile. In this self-examination, it will be useful for each of us to ask ourselves the following questions: Have I chosen realistic goals, taking into consideration that we are finite «biological organisms» in a natural world? Why do I value the goals I have chosen? Are these goals really as important as they seem? Have I unwittingly turned a personal desire into a standard for judging all of life? Do I have control over the means necessary for achieving this standard? We have no control over whether it is physically possible to live forever, nor do we have any control over whether there is a god who believes that life is significant. Therefore, it would be imprudent to adopt these two conditions as a standard for judging whether human striving is futile.

IV. Can Pursuits Be Futile, Yet Worthwhile?

Nihilists and futilitarians seem to assume that life cannot be worthwhile if there is no «purpose of life» or «life is futile.» The first premise has been subjected to penetrating criticism, as noted earlier. Having been created for a purpose is unnecessary for one's life to

---

[26] C. Stephen Evans, *The Philosophy of Despair: Existentialism and the Quest for Hope* (London: Probe Books, 1984), p. 65.

[27] Kurt Baier, «Threats of Futility: Is Life Worth Living,» *Free Inquiry*, 8 (Summer 1988), p. 51.

[28] Ibid., p. 52.

be worthwhile. What about the second premise? Is it a valid assumption or can living be worthwhile even if one's goals cannot be achieved?

Questions that have been discussed as part of the meaning of life debate include: «Does life have a meaning or purpose?,» «Is life futile?,» and «Is life worth living or worthwhile?» The first two questions are narrowly focused, in comparison to the third question, since they are concerned solely with purposes — whether there is a worthy purpose of life and whether one's envisioned purposes are achievable. The question about the worthwhileness of life is not limited to goal related experiences.

Determining whether an activity is «worthwhile» involves weighing the tangible and intangible benefits associated with performing the activity against the costs. If the benefits outweigh the costs, then it is generally concluded that the activity is worthwhile. For example, if it is believed that the benefits derived from attending a concert will exceed the costs, then this pursuit is considered worthwhile. This type of evaluation is often crude, implicit, and guided by emotions as well as rational thought. The same is true about an evaluation of whether life is or is not worthwhile. Thus, this form of evaluation differs from a formal, economic cost-benefit analysis, where the costs and benefits associated with an activity are made explicit and expressed in monetary amounts.

An evaluation of worthwhileness is a *much broader* type of evaluation than is an evaluation of futility. To see the difference between these types of evaluations, let us return to the race example. One runner is observed crossing the finish line as the slower runners make their way around the track. Since a runner has won the race, completing the race is futile for other runners whose goal was to finish first. Nevertheless, the runners typically complete the race. Why do they continue? If we asked them, I think they would give a variety of reasons. Some would continue for pride, because determination — even when defeat is a certainty — is a trait valued by many people, or simply because of the pleasure and exhilaration derived from running in the race. Others may continue to avoid certain costs such as the feeling of shamefulness that is often associated with quitting.

In effect, the runners have made two different types of evaluations, one regarding futility and another about worthwhileness. Runners' generally conclude that completing the race is worthwhile even though it may be considered futile. Thus, a runner can arrive at inconsistent judgements about whether to continue the race. This inconsistency is a reflection of the fact that there is a significant difference in the scope of the two types of evaluations.

Realizing the goal of winning the race is only one of the potential benefits of participating in the race. Even if this goal cannot be achieved, there may be other benefits associated with completing the race. These other benefits are taken into account in an evaluation of worthwhileness, but not in an evaluation of futility. A futility evaluation focuses primarily on the question of *effectiveness*: whether the means will be effective at producing the intended result. To a much lesser extent (it may be ignored altogether in some analyses), the costs (e.g., «toil») involved in achieving the intended result may also be considered in the evaluation. The broader evaluation of worthwhileness considers, not only the costs involved in pursuing the activity and whether the means will or will not be «effective» at bringing about the desired result, but other costs and benefits (e.g., joy of running in the race) of the activity that are unrelated to effectiveness. Because an evaluation of worthwhileness takes into account benefits not considered in an evaluation of futility, continuing the activity may be considered futile (i.e., ineffective), yet worthwhile.

To return to the question of suicide, how should one go about deciding whether to continue living? It is of utmost importance to choose the best approach in deciding this question since a particular method can unduly influence one's decision and because suicide is irreversible. Should one decide whether to go on living based on the results of a futility evaluation, a worthwhileness evaluation, or perhaps with another method? A futility evaluation involves answering the question «Is striving to achieve the envisioned goal(s) futile?,» whereas a worthwhileness evaluation addresses the question of «Do the benefits associated with living exceed the costs?» Which question should guide the decision?

Before addressing the above question, it will be useful to consider why medical futility has been proposed as a way to assist medical decision making and the advantages and disadvantages of this method of evaluation. A perceived advantage of a futility evaluation is that it avoids some of the problems associated with an economic cost-benefit analysis, including controversial ethical issues associated with placing a dollar value on the length or quality of human life and the difficulty of identifying and quantifying costs and benefits.[29] Since a futility evaluation is concerned solely with the narrow issue of whether or not a particular treatment will be effective for a given patient, the preceding problems are circumvented, to some extent, and this type of evaluation may be considered more «objective» than other types of evaluations.

A futility evaluation may seem more objective at first glance. However, it can in fact be as subject to personal bias as an evaluation of worthwhileness since it involves an assessment of whether the envisioned «goal» (which individuals may conceive of differently) will be achieved. Furthermore, the limited scope of a futility evaluation is a disadvantage as well as an advantage. Narrowing the scope of the evaluation may reduce intersubjective variability among those involved in the evaluation, but as a result it may exclude important values that would be taken into consideration in a broader evaluation and thus there is a tradeoff involved.

When a decision will be made on behalf of *other people* (e.g., provision of medical care to a patient), to help assure that the decision will be as fair and unbiased as possible, limiting the scope of an evaluation can be preferable in certain situations. However, the question about whether one should continue living is a different kind of question, one that we, as individuals, decide for ourselves. This personal question should be addressed using the broadest form of evaluation possible, which is an evaluation of worthwhileness.

A person can be «effective» and «successful,» by conventional standards, and yet miserable and find living not worthwhile. Conversely, one can be happy and find living worthwhile despite being ineffective or only modestly effective at achieving goals. Since an evaluation of worthwhileness takes into account experiences ignored by a futility evaluation, inconsistent judgements regarding whether or not one should continue living that are obtained from these two types of evaluation are likely to be, and should be, resolved based on the results of the worthwhileness evaluation. Furthermore, because an evaluation of the worthwhileness of life is extremely broad, it subsumes an evaluation of futility. Any costs and

---

[29] For a detailed discussion of the various methods used to assist in health care decision making, including cost-benefit analysis, cost-effectiveness analysis, and cost-utility analysis, see Henry M. Levin and Patrick J. McEwan, *Cost-Effectiveness Analysis: Methods and Applications* (London: Sage Publications, 2001) and *Cost-Effectiveness in Health and Medicine*, ed. Marthe R. Gold et al. (Oxford: Oxford University Press, 1996). These sources do not discuss medical futility. However, there are some similarities between a futility evaluation and a cost-effectiveness evaluation. One similarity is that both methods employ the concept of «effectiveness» instead of the broader concept of «benefits» used in a cost-benefit analysis.

benefits associated with achieving or failing to achieve one's goals are already factored into an evaluation of worthwhileness, which makes a futility evaluation superfluous.

Having and achieving purposes is undeniably an important part of life and is something that can enrich and give direction to our lives. However, there are other praiseworthy aspects of life which have nothing or very little to do with goal-directed activity. In deciding whether one should continue living, these experiences should not be overlooked or discounted.

V. Preventing Futility

To sum up, «life» is not futile in any sense apart from human subjective evaluation. Furthermore, just because the goal of living forever is unachievable does not mean that *every* action we take is futile or, in other words, that every goal is unachievable. Striving can be futile from a personal perspective, but whether or not it is futile can vary among individuals, over time, and from one goal to another, and largely depends on how high an individual has set his or her expectations and, to a lesser extent, on the capabilities of the individual.

Since no purpose has been imposed on humankind from without, we are free, to a large extent, to choose our own purposes in life, as many existentialists have emphasized. Not only are we free to choose *which* purposes we will and will not pursue, but we are *free to choose how high* to set our expectations.

Futility is not an inescapable part of the human condition, as futilitarians suggest, nor is a deity necessary for one to overcome a life of futility. The likelihood of experiencing futility is correlated with how high we have set our expectations. Therefore, we have a significant amount of control over whether or not our efforts will be futile. A person who lived life without having *any* goals would never experience frustration or failure. However, since this person would also never experience a sense of accomplishment, choosing to live an aimless existence is a high price to pay to avoid futility. A better way to avoid futility is to set realistic goals.

There is a delicate balance involved in setting expectations. Goals should be challenging, but achievable. A goal that is easily met provides little, if any, sense of accomplishment since it does not test one's capabilities. An unachievable goal also does not provide a sense of accomplishment, but for the opposite reason: because it exceeds human capabilities. The goal cannot in fact be achieved no matter how much effort is put forth.

By setting realistic goals and assuring that we have control over the means for achieving these goals, we can help to *prevent* our pursuits from becoming futile. However, if a person's overriding goal in life is to achieve personal immortality, and this goal is turned into a standard for judging futility, then all of one's efforts will surely seem futile. As a result, much energy will be spent *reacting to* and trying to cope with futility. In choosing unrealistic goals, one also indirectly chooses a life of futility, which can lead to frustration and despair. When futilitarians wonder why striving is futile, they need not look much further than to their own towering expectations.

Even by embracing realistic goals, achieving these goals might not be possible due to circumstances beyond a person's control. One way of reacting to futility, as discussed, is to give up. This could mean giving up on the few goals that one was unable to achieve or giving up on life altogether.

Instead of choosing the latter option just because one's expectations cannot be achieved, it would make more sense to consider pursuing other goals. Not only may these goals turn out

to be achievable, but, more importantly, pursuing these goals can foster the creation of meaning in life.

It is also important to explore and understand how the decision to «give up» was reached in the first place. The questions «Is there a purpose of life?» and «Is striving futile?» are too narrowly focused to provide a solid foundation for making a decision about whether one should continue living. This decision should be made taking into consideration *everything* associated with life, not just goal related experiences, and should be guided by the question «Is living worthwhile?»

If futilitarians would pause to consider this broader question, they will likely conclude, as I believe the author of Ecclesiastes ultimately did, that living can be worthwhile even if striving is futile (i.e., one's goals cannot be achieved). The author indicates: «So I have commended joy because there is nothing better for people under the sun, but to eat, drink, and enjoy» (Ecclesiastes 8.15).[30]

An evaluation of futility leaves out those experiences that many of us cherish the most: being with family and friends, appreciating the beauty and grandeur of the rising and the setting of the sun, and marveling at the fact that life exists, just to name a few. There is more to life than just having and achieving purposes.

Brooke Alan Trisel
4443 Prairie Pine Court
Hilliard, Ohio, U.S.A. 43026
Phone Number: (614) 777-9452
<triselba@cs.com>

[30] For other statements supporting that the author concluded that living is worthwhile, see Ecclesiastes 3.12-13 and 11.7.

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 85-109
On The Fourfold Root Of Philosophical Skepticism
Copyright © by SORITES and Mark Walker

ON THE FOURFOLD ROOT OF PHILOSOPHICAL SKEPTICISM[1]

Mark Walker

<mark@markalanwalker.com>

1.0 A Puzzle?

Philosophical skepticism is a view, or a family of views, about human knowledge in general. Michael Williams, for example, writes that «in broadest terms, the philosophical skeptic holds, or at least finds irrefutable, the view that knowledge is impossible.»[2] A more modest characterization casts philosophical skepticism as a challenge to justify our claims to knowledge.[3] One of the central features of philosophical skepticism is the generality of its scope: in the limit it is a thesis about knowledge in general, although often it is directed at a more specific target such as knowledge of the external world. In any event, it is clear that the philosophical skeptic cannot proceed by challenging epistemic claims individually, if for no other reason than this project would be too time consuming, indeed, potentially interminable. Of course, historically a much more direct approach has been attempted: discussions of skepticism are often directed towards the justification condition for knowledge.[4] The skeptic's strategy then can be understood as attempting to show that we may not know what we think we know because one of the necessary conditions for knowledge is never (or almost never) satisfied. As Grayling observes:

> The general problem with justification is that the procedures we adopt, across all walks of epistemic life, appear highly permeable to difficulties posed by scepticism. The problem of justification is therefore in large part the problem of scepticism; which is precisely why discussion of scepticism is central to epistemology.[5]

---

[1] I would like to express my appreciation to Professors Nick Griffin and David Hitchcock for helpful comments on an earlier version of this paper.

[2] «Skepticism» in *The Blackwell Guide to Epistemology*, edited by J. Greco and E. Sosa, Massachusetts: Blackwell, (1999), p. 35.

[3] A. C. Grayling, «Epistemology», in *The Blackwell Companion to Philosophy*, edited by Nicholas Bunnin and E. P. Tsui-James, Cambridge, Massachusetts: Blackwell, (1996), p. 51.

[4] This seems true at least of modern forms of skepticism, which is my main concern here. For discussion of the relation of modern skepticism to the Pyrrhonian tradition see, Myles Burnyeat, «Idealism and Greek Philosophy: What Descartes Saw and Berkeley Missed,» *Philosophical Review*, (1982), and Michael Williams, «Scepticism Without Theory», *Review of Metaphysics* (March 1988): 547-588.

[5] *Op. cit.*, p. 46.

But at this level of abstraction, it might seem perplexing why the justification condition is singled out; for the traditional definition of knowledge is that it is equivalent to justified, true belief.[6] Given that the skeptic's goal is to undermine confidence in our epistemic claims, and that there are additional necessary conditions for knowledge, why is so much emphasis placed on the justification condition? After all, undermining any of the other conditions ought to work with equal effect. That is, unquestionably there is nothing about the *logical* relations among the various components of knowledge that may be used to differentiate them as intrinsically interesting for skeptics. If, for example, justification were the sole necessary condition for knowledge, and the other conditions figured in analyzes of sufficient conditions, then, the reason for attention to the justification condition might be apparent. However, given our present understanding of knowledge, then, necessarily, if any one of these necessary conditions is not satisfied then knowledge is impossible. Of course there may be reasons other than its logical relation to knowledge for focusing on justification — perhaps there is some reason to suppose that justification is the weak link in the structure of knowledge. Whether this is so remains to be seen.

I mention this «puzzle» — if it is indeed that — only to illuminate the general strategy I hope to explore here. My argument will be that the same sorts of considerations that motivate skeptical questioning of the justification component of knowledge can be extended to the other definiens of knowledge. Our first task then is to examine the structure and definition of knowledge in order to shed light on other potential candidates for skeptical questioning. In section 3 I will attempt to articulate four types of skepticism based on this definition. Section 4 argues that these four types of skepticism can be supported by invoking counter-possibilities or «defeaters». Section 5 concludes with a brief look at some applications and consequences of this taxonomy of skepticism.

I should perhaps issue a caveat at this stage. If it is not already apparent this is a «big-picture» paper in the sense that it covers an enormous amount of area. Of necessity then exploration of many details of particular arguments will have to be sacrificed. I believe that, if my argument is persuasive at all, it is in the manner in which it unites what are perhaps the two largest epistemic enterprises of the late twentieth century-defining 'knowledge' and responding to skepticism-into a single synoptic vision.

2.0 Defining 'Knowledge'

The proposal then is to analyze various forms of philosophical skepticism in terms of the necessary conditions of knowledge. However, an obvious obstacle stands in the way: Gettier's criticism of the traditional definition of knowledge seems to leave the concept of knowledge partially undefined. I will argue, however, that there is a definition of knowledge of sufficient clarity available — at least for our purposes.

Let us, then, first look at the Gettier problem. The traditional definition of knowledge is often explicated along these lines:

S knows that p = Definition:

---

[6] Edmund Gettier's criticism of this definition indicates that there is an additional condition for knowledge, which brings the total to four. «Is Justified True Belief Knowledge?», *Analysis*, vol. 23, (1963), pp. 121-3. I discuss the Gettier problem below. It is perhaps worth noting that I shall assume here that there are only four necessary conditions for knowledge, although I suppose we cannot rule out the possibility of counterexamples leading to a search for a fifth condition for knowledge.

i) S believes that P

ii) It is true that P

iii) S is justified in believing that P.

Gettier offered a couple of counterexamples by way of refutation of the claim that these conditions are sufficient for knowledge. However, we will use an example adapted from Russell to illustrate this point.[7] Let us suppose that Fred has worked in an office for forty years and has observed a clock working reliably for the same period. One morning Fred glances at the clock and as a result believes that it is nine o'clock. What he does not know is that the clock stopped working exactly twelve hours ago. Thus, it is true that it is nine o'clock, Fred believes that it is nine o'clock, and Fred is justified in his belief that it is nine o'clock. However, the intuition of most is that Fred does not know that it is nine o'clock. Something seems to be missing.

Over the years we have seen a number of ingenious attempts to specify what this something missing is, the search for the elusive «fourth condition».[8] This is not our quarry. Rather, for our purposes, a much more modest proposal is sufficient. The leading idea is that a natural understanding of cases like that of the clock is that the agent has a justified, true belief, but the belief is not X, where 'X' names precisely that awareness articulated by the Gettier type examples. That is, the concept X articulates the common understanding that is shared in response to these types of examples. Following in the great taxonomical tradition in biology we may name this concept X, 'gettier'.

Assuming (as we shall) that the traditional analysis of knowledge is correct,[9] the following four conditions, then, are necessary and sufficient for knowledge:

S knows that p = DF

i) S believes that p

ii) It is true that p

iii) S is justified in believing that p.

iv) S's belief that p is gettier.

An obvious response to this definition is that it does not solve the Gettier problem, it (at best) names the problem. This criticism is correct if we are speaking about constructing a *theory* of knowledge, although not if we are attempting merely to make available a *definition* of knowledge. To see our way to this point notice what is implicit in what was said above: the traditional definition of knowledge is defined in terms of primitive concepts, not in terms

---

7 *Human Knowledge: Its Scope and Limits*, New York: Simon & Schuster, (1948), p. 154.

8 *Cf.* L. T. Zagzebski, «The Inescapability of Gettier Problems», *Philosophical Quarterly* 44, no. 174 (1994), pp. 65-73.

9 Clearly a complete defense of this argument would require an examination of the traditional (but not uncontroversial) view that justified, true belief is equivalent to knowledge. D. Lewis, for example, questions the necessity of the justification condition in «Elusive Knowledge,» *Australasian Journal of Philosophy*, 74 (1996), pp. 549-67, and H. A. Pritchard argues that knowledge and belief are mutually exclusive in *Knowledge and Perception*, Oxford: Claredon Press (1950).

of theories about these more primitive concepts. For example, it is not typically taken as a point against the traditional definition of knowledge that the condition «it is true that p» is neutral (or less charitably, ambiguous) among the competing theories of truth. Perhaps a full-blown theory of knowledge would require a full theoretical articulation of the constituent terms, in which case the debate over the nature of truth, e.g., correspondence versus coherence etc., would be to the point. But this is not what the traditional account sought. An exact parallel between the fourth condition and the other three would have us cite some «ready-made» concept. If the parallel holds we should be able to cite some primitive concept analogous to 'truth', 'justification' and 'belief' to fill in the fourth spot. The actual literature on the fourth condition, however, looks more like the sorts of debates that take place when we are trying to articulate a full-blown theory, rather than merely citing some «ready-made» concept.

Naturally one might think that if there were such a ready-made concept, then it would have been a relatively simple matter to cite the concept, and the «Gettier problem» would have been solved long ago. The fact that rather complex theories for the fourth condition have been proposed, it might be argued, ought to be taken as some evidence that we lack such a concept. But do we actually lack such a ready-made concept? I think not. I propose that we see Gettier's accomplishment as pointing out that there is a concept that we have been able to understand and use, but not mention. It is this concept that allows us to identify Gettier-type cases and discuss various theoretical solutions to the Gettier problem, just as our concept of truth allows us to discuss various theories of truth. So naming this concept means we have a concept that we can now both use and mention, and it allows us to complete the task of defining knowledge. The project of providing a full-blown theory of the fourth condition remains as incomplete as before. Much of the literature on the Gettier problem, then, can be seen as a contribution to the task of articulating a theory. What is important to realize, however, is that there is no «Gettier problem» with respect to the *definition* of knowledge anymore than there is a problem with its sibling concepts, i.e., 'truth', 'justification', and 'belief'. Whether this definition is acceptable, ultimately, will depend in large measure whether it proves useful. I hope to show that it can be put to good use below.[10]

An analogy from the history of philosophy might help secure this point. Aristotle in the *Nicomachean Ethics* asks what the highest good attainable by actions is and asserts that most would say that it is 'happiness'. As Aristotle points out, although this may answer one question it leaves unresolved the explication of happiness: some think that it is the possession of health, others say wealth or honor etc.[11] We can imagine that the concept of gettier plays a similar role: we may agree that knowledge has as one of its definiens the concept of gettier, but disagree when it comes to explicating what gettier is.

---

[10] I don't expect many will be persuaded by this suggestion for how to use 'gettier'. My case should, I hope, improve as we examine more global types of skepticism. However, nothing in the argument that distinguishes various types of skepticism turns on accepting this understanding of 'gettier'. Where I use the term 'gettier', for instance, one can simply read this as a place-holder for 'the philosophical theory which supplements the traditional analysis of 'knowledge' in a way which takes account of the Gettier type counter-examples'.

[11] Aristotle, *Nicomachean Ethics,* translated by M. Ostwald, Indianapolis: Liberal Arts Press, (1962), 1095a.

We ought to pause here to consider a serious objection to this definition, namely: that the concept of gettier is completely parasitic on that of knowledge. It is serious, since if this objection could be maintained, it would seem we have good reason for rejecting the proposed definition of knowledge. After all, we have an understanding of the other concepts used to define knowledge, independent of our understanding of knowledge, certainly it seems that we should demand this much of the gettier definien. This objection, however, is unsuccessful as the concept of gettier too has a life independent of knowledge; for we may ask whether a statement rather than a purported item of knowledge is gettier. Adjusting our previous example, let us suppose that Fred by some mental quirk (brought on by too many philosophy courses according to his friends) distrusts clocks. Fred thus does not know that it is nine o'clock, according to the traditional analysis, because he does not believe that it is nine o'clock. When Sally, Fred's boss, asks the time he replies that it is nine o'clock — Fred realizes that his skepticism about clocks is not shared or appreciated by his boss. Fred's statement is justified — as we have said the clock has worked flawlessly for forty years — and the statement is true. Given that the clock stopped exactly twelve hours previous, it seems that Fred's statement is not gettier. The fact that we understand whether a statement is gettier or not independent of whether the statement is known shows that the concept of gettier is not completely parasitic on that of knowledge. The appropriate parallel here is how we can understand a statement being either true or justified independent of whether the statement is known. At least in this respect, the concept of 'gettier' does not differ from its more illustrious sibling concepts.

3.0 Types of Philosophical Skepticism

In this section I will attempt to delineate four types of philosophical skepticism based on the four components of knowledge: justificatory skepticism, gettier skepticism, noetic skepticism, and aletheia skepticism. The conditions that justificatory and gettier skepticism are associated with are self-explanatory. Noetic skepticism concentrates on the belief condition of knowledge and aletheia skepticism questions the truth component.

3.1 Justificatory Skepticism

As its name implies, justificatory skepticism focuses on the justification condition for knowledge. As will soon become apparent, justificatory skepticism is more or less coextensive with what is customarily considered the entirety of philosophical skepticism. Historically, there are two main ways in which this idea is commonly expounded: one is what has been referred to as the 'regress problem' but, as Williams argues, is more aptly termed 'Agrippa's trilemma'.[12] The other is what might be thought of as Cartesian skepticism. Let us take these in turn.

3.11 Agrippa's Trilemma

Agrippa's trilemma traces its historical roots to the Ancient Greeks who first noticed that skepticism seems a natural consequence of reflection on the process of justifying our epistemic claims. When confronted with a persistent skeptic it seems we are faced with a very unappealing trilemma. Suppose a justificatory skeptic asks us to defend some epistemic claim P. We might choose to defend it on the basis of some further claim Q, but the skeptic will then ask about the justification for Q itself. If we invoke R in support of Q then, short of some alternate strategy, we are faced with an infinite regress. The difficulty of course is that

---

[12] «Skepticism», *op. cit.*

only two alternatives present themselves: either we invoke P again in our defense of R, but this invites the criticism of begging the question. The only alternative is to insist that some claims or beliefs are self-justifying, e.g., let us suppose that R justifies Q and Q justifies P, but R itself does not require any further justification, R is self-justifying. R, in other words, is an axiom. This suggestion at least has the correct *form* for silencing the skeptic. If we can get the skeptic to accept our chain of reasoning and the axiom then we will have met the challenge. The trouble of course is that this foundational response to skepticism has generally been considered a failure.[13] Roughly, the difficulty seems to be in discovering claims that are both specific enough to provide justification for statements like «I am sitting in front of a fire» and beyond skeptical doubt. Certain logical truths like «not every statement is true and false»[14] are perhaps beyond skeptical doubt, (or at least the best candidates for this role), but it is difficult to see how such abstract claims on their own could be used to support specific claims. On the other hand, if we turn to propositions that look like they might be more useful in supporting the fireside claim, e.g., a benevolent deity would not allow me to be radically deceived about such matters — these look like they are much more vulnerable to skeptical doubt. In any event, this is a thumbnail version of one understanding of the failure of the foundationalist answer to the skeptic.

3.12 Cartesian Skepticism

The other form of justificatory skepticism is most famous and familiar from Descartes' evil demon thought experiment.[15] Descartes' point, to put it into a contemporary idiom, is that the evidence we have for our epistemic claims is often radically underdetermined, i.e., the evidence is compatible with (and perhaps equally supportive of) any number of hypotheses. As Descartes suggests in the *Meditations*, he has very good sensory evidence to suppose that he is sitting in front of a fire, but this evidence is also compatible with the hypothesis that a powerful evil demon is attempting to deceive him that he is sitting in front of a fire. A modern variant on this thought experiment — made popular in such movies as The Matrix — speculates that what appears real to us is merely a virtual world created by some supercomputer. The most gruesome versions of this thought experiment ask us to image that our brains have been surgically removed and placed in a vat of nurturing nutrients. The computer interfaces via implants designed to mimic the sorts of information that our sense organs would send along the nervous system. Thus, the evidence that I have for the belief that I am presently sitting at a desk writing this paper is also logically compatible with the suggestion that my brain is floating in a vat full of nutrients and a computer is generating for me the illusion that I am sitting at a desk. If we take such scenarios seriously then it seems that much of what we believe and claim to know might in fact be false.

---

[13] *Cf.* Rorty's *Philosophy and the Mirror of Nature,* Princeton: Princeton University Press, (1979). But see: T. Triplett «Recent Work on Foundationalism,» *American Philosophical Quarterly* 27, (1990), pp. 93-116.

[14] This is what Putnam refers to as the minimal principle of noncontradiction, «There is at least one a priori truth,» reprinted in Putnam's *Realism and Reason*, Philosophical Papers vol. 3 Cambridge Univ. Press, (1983).

[15] Myles Burnyeat in «Idealism and Greek Philosophy: What Descartes Saw and Berkeley Missed,» *op. cit.*, argues that this form of skepticism makes its first appearance in Descartes' work.

It should be emphasized that this skeptical suggestion, that it is possible that all our beliefs (about the external world) might be false, is a point about justification. The argument from the underdetermination of evidence demonstrates that (standard) attempts to justify our epistemic claims typically do not imply the truth of the statements they are intended to justify. Take for example what we might think of as the 'standard view of the world', which, for at least a certain segment of the world's population, consists in the conjunction of what Sellars termed the 'manifest' and 'scientific images', which correspond roughly to the beliefs and theories of common sense and contemporary science.[16] Cartesian skepticism says that our justification for believing the external world is more or less like we think it is in fact implies a disjunct along the lines of:

D: Either the world is more or less as standard view describes, or the brain-in-the-vat scenario is true, or the evil demon hypothesis is true, etc.

For as we have said, the evidence that we take to support our knowledge claims about the external world is logically compatible with the claim that we are deceived. Cartesian skepticism, then, can be seen as challenging us to eliminate the skeptical disjuncts of D. Of course this is precisely what the Cartesian skeptic suspects cannot be done, i.e., we have no way to justify our common sense and scientific view such that they are uniquely implied by our justificatory practices. That is, the Cartesian argues that there does not appear to by any a priori reason why we should favor the standard view, nor does the empirical evidence provided by our senses allow us to favor the standard view. Whether we ought to accept the Cartesians' assessment of the epistemic terrain remains controversial. (I discuss (but do not defend) the Cartesian view in more detail in section 4.1).

The conclusion of both types of justificatory skepticism is that the ability to justify putative items of knowledge is beyond our abilities, i.e., the justification condition cannot be satisfied, and hence, knowledge is impossible.[17] I hope that this brief exposition is sufficient to indicate that these two types or families of skepticism are more or less coextensive with what is normally considered philosophical skepticism.[18] I say 'families' because it is clear that other famous skeptical hypotheses, e.g., Russell's suggestion that it is difficult to prove that the world wasn't created only five minutes ago, skepticism about induction, or skepticism about other minds, have the same general form as Cartesian skepticism.

3.2 Gettier Skepticism

---

[16] Wilfrid Sellars, «Philosophy and the Scientific Image of Man,» in *Science, Perception, and Reality,* London: Routledge and Kegan Paul, Ltd., (1963).

[17] I shall often discuss the four types of philosophical skepticism in terms of the strongest formulation of philosophical skepticism discussed, namely, that 'knowledge is impossible'. Naturally, these skepticisms might be formulated in a weaker manner as suggested by Grayling's formulation noted above.

[18] Ultimately, it may be wondered to what extent these two types of justificatory skepticism are in fact independent. After all, if we could address the foundational problem of Agrippa's trilemma then we might have an answer to the Cartesian skeptic. Indeed, this is the precisely the strategy Descartes opted for. Conversely, it seems that often the most effective means for undermining responses to the Agrippa skeptic is to point out how evidence for putative axioms are underdetermined by the evidence. However, for our purposes, nothing is lost in distinguishing between these two types of justificatory skepticism.

The first nonstandard philosophical skepticism that we should examine is gettier skepticism.[19] Gettier-type examples typically involve very local epistemic claims, as we saw in the clock example above. The gettier skeptic asserts that it is quite possible to imagine much more radical thought experiments, as the 'Ed the lucky brain-in-a-vat' example demonstrates. Let us suppose that Ed is the hapless victim of a «brain kidnapping». The proverbial mad scientist has surgically removed his brain and placed it in a vat. The mad scientist has set-up the program such that in the morning when Ed awakes Ed will believe that he is reading his copy of the «Sydney Morning Herald» when in fact he will be reading the scientist's vat-image imitation. The scientist plans to plant all sorts of implausible stories to test Ed's gullibility, e.g., that President Bush is a (closet) card-carrying Trotskyite, that scientists have discovered that the moon is actually made of cheese (other than a thin surface crust of rock). However, that night there is a mammoth electrical storm and a bolt of lightning hits the computer controlling Ed's «vat world». By massive coincidence, the computer software is scrambled such that the program will in fact provide Ed with a lot more truths than the scientist had intended. It turns out that the vat-image imitation of the morning paper is word for word identical to the real «Sydney Morning Herald». Ed might read for instance that the stock market is down and thus believe the truth that the stock market is down, when in fact the stock market is down. The computer's software might engender the belief in Ed that it is a cloudy day, and it is in fact a cloudy day, and so on. We might suppose that Ed has as many true beliefs about the world (which do not involve him) as he would have if he had not been kidnapped. Ed's beliefs about the current state of the world — the stock market, the weather etc. — even though they are true and justified, fail to qualify as knowledge. His beliefs are not gettier. They depend on the most unlikely of circumstances, e.g., the lightning hitting the computer in just a certain way. Thus, a skeptic may grant that a brain-in-a-vat might have true, justified beliefs, but deny that such a brain would in fact know. For the brain's beliefs might not be gettier. Gettier skepticism is a form of philosophical skepticism then because it advises that knowledge is impossible because we have no reason to suppose that the gettier condition of knowledge is satisfied.

3.3 Noetic Skepticism

Noetic skepticism differs from the previous skepticisms in that it questions whether we are capable of formulating the appropriate types of belief in the first place — never mind the subsidiary question of whether these beliefs are true or gettier. It claims that the hypothesis that correctly describes the truth might be beyond the «reach of our minds» — to use Nagel's

---

[19] I thus disagree with Michael Williams (*Groundless Belief*, Connecticut: Yale University Press, (1977) p. 6):

> How does Gettier's problem-the problem of formulating necessary and sufficient conditions for knowledge-relate to the dominant concerns of traditional epistemology? On the face of things, only distantly.

On the other hand, readers of an earlier version of this paper pointed out the relevance of an interesting article by Hetherington (1996). Unbeknownst to me, it seems that Hetherington had already noted the skeptical potential of Gettier cases. There are, however, substantive differences in our exposition: First, Hetherington seems to be skeptical about the possibility of reductive definitions of knowledge like the one offered above (see his discussion on pp. 96-7, footnote 12). Second, I believe there is a greater parallelism between Gettier skepticism and Cartesian skepticism than he allows (see especially p.88). Finally, Hetherington exploits a parallel between Cartesian dreaming arguments while the arguments we will look at focus on brain-in-the-vat and evil demon skepticism.

formulation.[20] The idea that we are conceptually limited is one familiar from the history of philosophy and can be traced back at least as far as Heraclitus:

> (Fragment 83) The wisest man will appear an ape in relation to God, both in wisdom and beauty and everything else.
>
> (Fragment 79) Man is called childish compared with divinity, just as a boy compared with man.[21]

This line of thought has recently been recast within a naturalistic framework. Jerry Fodor writes:

> …so long as the class of accessible concepts is endogenously constrained, there will be thoughts that we are unequipped to think. And, so far, nobody has been able to devise an account of the ontogeny of concepts which does not imply such endogenous constraints. This conclusion may seem less unbearably depressing if one considers that it is one which we unhesitatingly accept for every *other* species. One would presumably not be impressed by a priori arguments intended to prove (e.g.) that the true science *must* be accessible to spiders.[22]

Heraclitus' analogies and Fodor's reflections on the epistemological consequences of evolutionary theory are suggestive of a quite radical form of skepticism: namely, that we are congenitally incapable of knowing very much of anything.[23] There are two ways of articulating noetic skepticism: One is what I shall term the 'proper subset argument', the other is the 'Kantian argument'.

3.31 Proper Subset Noetic Skepticism

This formulation of the argument takes its cue from Heraclitus' ontogenetic analogy: human understanding is said to stand to an omniscient understanding as a child's understanding stands to ours. Obviously the younger we imagine the child in the comparison the less flattering it looks to our level of intelligence and knowledge. Plausibly, the average

---

[20] A somewhat analogous distinction is made by Nagel in his *The View From Nowhere*, Oxford: Oxford University Press, (1986), p. 90:

> In the last chapter we discussed skepticism with regard to knowledge. Here I want to introduce another form of skepticism — not about what we know but about how far our thoughts can reach. I shall defend a form of realism according to which our grasp on the world is limited not only in respect of what we can know but also in respect of what we can conceive. In a very strong sense, the world extends beyond the reach of our minds.

His distinction is not exactly the same as the one discussed here. One difference lies in the fact that Nagel seems to suggest at certain points that the world does in fact transcend our ability to conceptualize it whereas the skepticism here asserts merely that we leave open the possibility of such a transcendence. More importantly, Nagel does not consider the possibility the more radical versions of the idea that our thoughts cannot reach reality in itself. In other words, he may have something similar to the proper-subset version of noetic skepticism in mind here. Further evidence for this is the fact that the chapter from which this quote is taken concentrates on an analogy almost identical to that of Heraclitus' ontogenetic analogy (see below).

[21] *The Worlds of the Early Greek Philosophers*, edited by J. B. Wilbur and H. J. Allen, (Buffalo: Prometheus Books 1979), p. 72.

[22] *The Modularity of Mind*, (Cambridge: The MIT Press, 1983), pp. 125-6. I would be impressed if the spiders themselves made the arguments — although I am not sure I would believe the arguments.

[23] Noetic skepticism might be articulated as a biological thesis: there are congenital limitations on the sorts of thoughts that we might think; or as a socio-historical thesis: there are limits to what we can think at this particular point in our cultural development. In the text I concentrate for the most part on the stronger biological thesis.

child of 15 is capable of understanding most of what an average adult knows; so let us suppose, for illustrative purposes only, that a 15 year old can understand and know 80% of an average adult's knowledge. This comparison might arguably be said to generate merely a local skepticism when applied to our own case. For if we stand to an omniscient being as the 15 year old stands to us, then we have the conceptual capacity to understand 80 % of an omniscient being's knowledge. However, if the comparison is between an adult and a child of 2, and the average child of 2 knows (let us suppose) 20 % of what the average adult knows, then we have the basis for generating a global skepticism. For now we must imagine that it is possible for us to know only 20 % of what the omniscient being knows, that is, 80 % of the omniscient being's knowledge is unknowable by humans. The discrepancy seems at least as large if we switch to Heraclitus' phylogenetic analogy, for it seems quite likely that apes know much less than 20% of what a normal human adult knows. How far we can push this analogy is an interesting question. Stephan Körner suggests that our level of intelligence compared to a superintelligence might be like that of a worm's to a human.[24] Worms are almost entirely ignorant about the nature of the universe so what values should we set here, or does the analogy become unintelligible at this point? It will take us too far a field to investigate these questions. In any event, to the extent that we accept the proper subset argument, it seems in harmony with the philosophical skeptic's recommendation for epistemic modesty.

3.32 Kantian Noetic Skepticism

A second line of argument derives from Kant who developed a version of noetic skepticism about the external world considered as a thing in itself. This is not the place to explore the subtleties of Kant's position; we will have to be content with a thumbnail version of his noetic skepticism: in this connection the upshot of Kant's argument of the first *Critique* is that some of the most basic features of our conceptual scheme may serve to distort the nature of things in themselves.[25] For example, one of Kant's suggestions is that our space/time intuitions may misrepresent the true nature of things in themselves, for we cannot rule out the possibility that God might intuit the world in a different manner, i.e., we have to contend with the possibility that He might not possess our sensuous form of intuition, but instead, work with an intellectual intuition. Kant also seems to think that God knows the world directly with his intellectual intuition, He does not have a discursive understanding like us since: «…all his

---

[24] Stephan Körner, *Fundamental Questions in Philosophy*, (Harmondsworth: Penguin, 1969).

[25] Clearly, this «ontological» reading of the notion of the things in themselves offered is only one interpretation. The notion of the thing in itself, for example, has also been interpreted along epistemological lines by N. Rescher in «Noumenal Causality», in L. W. Beck edits, *The Proceedings of the Third International Kant Congress*, Dordrecht: D. Reidel Publishing, 1972. R. Butts provides an interesting methodological account in *Kant and the Double Government Methodology*, Boston: Reidel Publishing, 1984. P. F. Strawson suggests that there is no coherent account of the notion of things in themselves to be found in Kant, v. *The Bounds of Sense*, London: Methuen, 1966. Those that hold to what I am terming the 'ontological reading' are numerous (perhaps they even constitute the majority). To cite but a few: R. Solomon, *In the Spirit of Hegel*, Oxford: Oxford University Press. (1983), pp. 296-7; S. Körner, *Kant*, New York: Penguin Books, (1955), p. 96; R.C.S. Walker, *Kant*, London: Routledge & Kegan Paul, Ltd., (1978), pp. 89-90; A. C. Ewing, *A Short Commentary on Kant's Critique of Pure Reason*, London: Methuen, (1938), pp. 188-9, D. P. Dryer, *Kant's Solution for Verification in Metaphysics*, Toronto: University of Toronto Press, (1966), p. 519; J. N. Findlay, *Kant and the Transcendental Object*, Oxford: Claredon Press, (1981), p. 350; and R. Pippin, *Kant's Theory of Form*, New Haven: Yale University Press, (1982), p. 203; M. Westphal's «In Defense of the Thing in Itself,» *Kant-Studien*, (1968), Heft 1, 118-41.

[God's] knowledge must be intuition, and not *thought*, which always involves limitation.»[26] Kant argues that we cannot say much about how this form of intuition might work; it seems to be beyond our conceptual grasp, indeed, he says that we can say nothing positive about this alternate means of intuiting the world.[27] Can we do any better? Not much, perhaps, for as with the proper subset argument, it seems that we must resort to analogies to clarify this skepticism, for obviously we cannot describe in detail that which lies beyond the limits of our thought. An analogy that might be of some use is that of a map, a two dimensional representation of our planet, versus a three dimensional representation like a globe. It is well-known that many standard maps tend to distort somewhat the exact spatial relations among the parts of our planet, e.g., Greenland often tends to get represented as much larger than it really is. Not only are maps usually inherently less than perfectly accurate they are also incomplete, e.g., the shortest distance between any two points on the planet will almost always yield the wrong answer because maps ignore the three dimensional topography of our earth, e.g., the shortest distance to China from Canada, as many children digging in their backyard know, is through the earth. To apply this analogy to our cognitive situation we might image that our understanding of the external world (in itself) is like that of a map, whereas an omniscient being's is more like that of a globe. If this is the case then we can see how our understanding is both conceptually capped — we fail to understand certain concepts just as a map fails to represent the three dimensional aspects of the world — and it may be distorting, just as maps regularly represent parts of the planet as either too large or too small. Notice too that our analogy helps us understand how this form of noetic skepticism can be explained in a manner consistent with naturalism. We may suppose that the additional information of a «globe» type representation of reality — assuming that it is even available in the gene pool — may have conferred little or no additional survival value to our ancestors while taking up cognitive resources. That is, maps may be less than perfectly accurate and conceptually incomplete, but they can be extremely useful. If this analogy holds, then, we can make sense of the idea that the true representation of the external world is beyond our cognitive capacity, and furthermore, that our form of representing the external world may be false, but nevertheless, our form of representing the external world may be extremely useful.

3.4 Aletheia Skepticism

Aletheia skepticism focuses on the truth condition for knowledge: it says that truth is (for the most part) beyond our ability to grasp.[28] F. H. Bradley is perhaps the most illustrious inspiration for aletheia skepticism.[29] Bradley argued that «There is no possible relational scheme which in my view in the end will be truth…I had long ago made it clear (so I

---

[26] *Critique of Pure Reason*, translated by Norman Kemp, London: Macmillan, (1929), B 71.

[27] *Ibid.* B 307-312.

[28] Like noetic skepticism, aletheia skepticism might be articulated as a biological or socio-historical constraint.

[29] Whether Bradley himself qualifies as an aletheia skeptic is a difficult exegetical question. He does say, for example, that the 'the chief need of English philosophy is, I think, a skeptical study of first principles…' (*Appearance and Reality.* 2nd edition, Oxford: Oxford University Press (1969), p. xii). But he seems to allow that feeling might be somewhat revelatory of the ultimate truth of reality (*Essays on Truth and Reality, op. cit.*, p. 159) which seems incompatible with aletheia skepticism.

thought) that for me no truth was quite true…»[30] In other words, all human attempts to ascertain the truth fail. If we accept this thesis, then from our definition of knowledge, philosophical skepticism follows as a consequence. If one agrees with Bradley then skepticism follows, for one of the necessary conditions for knowledge is unattainable by humans. Of course this begs the self-referential question: is a Bradley inspired aletheia skepticism itself true? Presumably Bradley would be forced to say that it is not quite true. It would probably be premature for the anti-skeptic to celebrate, since it does not follow that anti-skepticism has ascertained the truth on this point either, for Bradley claims that no truth is quite true. Bradley had a doctrine of degrees of truth that might help here. Specifically, while Bradley claims that no human attempts to ascertain the truth can ever be quite true, nevertheless, he argues that some claims are closer to the truth than others. Thus, it may be that aletheia skepticism is closer to the truth than anti-skepticism.

Bradley's argument for the claim that «no truth is quite true», in outline, is that human cognition is shot-through with contradictions, but reality in itself is not contradictory; rather reality is a self-consistent whole:

> Truth is an ideal expression of the Universe, at once coherent and comprehensive. It must not conflict with itself, and there must be no suggestion which fails to fall inside it. Perfect truth in short must realize the idea of a systematic whole.[31]

Hence, it follows, according to Bradley, that humans are incapable of grasping the truth, which, as we have noted, implies philosophical skepticism.

Chomsky suggests a more naturalistic route to aletheia skepticism, he writes,

> What is the relation between the class of humanly accessible theories and the class of true theories? It is possible that the intersection of these classes is quite small, that few true theories are accessible. There is no evolutionary argument to the contrary. Nor is there any reason to accept the traditional doctrine, as expressed by Descartes, that human reason is a «universal instrument which can serve for all contingencies.» Rather, it is a specific biological system, with its potentialities and associated limitations. It may turn out to have been a lucky accident that the intersection is not null. There is no particular reason to suppose that the science-forming capacities of humans or their mathematical abilities permit them to conceive of theories approximating the truth in every (or any) domain, or to gain insight into the laws of nature.[32]

The general strategy of the aletheia philosophical skeptic is the same as the other types of philosophical skepticism: the argument is that knowledge is impossible because one of the necessary conditions for knowledge, viz., truth, is beyond the reach of humans.

Before moving on we should dwell for a moment on the question of the relation between noetic and justificatory skepticism. It might be objected that these two types of skepticism are not as distinct as advertised: both ultimately, it seems, want to conclude that formulating true beliefs about reality is impossible for humans. Certainly there is something to this objection, given that it seems that we might rightly hope that a full philosophical account of 'belief' and 'truth' will speak to how the concepts intersect. But even if one believes with the view that

---

[30] *Essays on Truth and Reality*, London, 1914, p.239.

[31] *Ibid.,* p. 223.

[32] Noam Chomsky, *Rules and Representations*, New York: Cambridge University Press, (1980), pp. 251-2.

there are strong conceptual linkages between belief and truth,[33] still there may be some point to distinguishing these two skepticisms; if for no other reason than to mark the strategy they used to reach the conclusion that true beliefs about reality are impossible for humans. As we saw in the discussion of noetic skepticism, Fodor approaches the conclusion that true belief is impossible on the basis of the idea that there are concepts and thus beliefs that we cannot formulate. Chomsky, on the other hand, in the quote above, emphasizes that it is a *true* theory that is beyond our grasp. We can emphasize the divergence here by noting that for the noetic skeptic it does not follow that truth is beyond our grasp: it is logically consistent with noetic skepticism that we have some non-belief based access to the truth, e.g., imagine that someone believed that art provides a «feeling» or «intuition» about the true nature of reality, but that humans are not capable of formulating this access to truth in terms of belief.[34] On the other hand, this non-belief access to truth is incompatible with aletheia skepticism for it is the doctrine that the truth about reality is beyond our grasp; not simply that true belief is beyond our grasp. Conversely, noetic skepticism is the thesis that reality transcends our ability to believe it, so it follows that true belief is beyond our grasp. But so too is false belief. Our situation, if the noetic skeptic is correct, is like that of a monkey's vis-à-vis physics: it is not that we think that monkeys have a false view about physics — that they are Newtonians when they should be Einsteinians — rather they do not have physics-type beliefs at all. A noetic skeptic, in contrast, can allow that we have beliefs in the relevant domain, it is simply that these beliefs are false. All this demonstrates is that noetic skepticism and aletheia skepticism are logically distinct. As intimated, in practice we might suppose that the sorts of considerations that motivate noetic and aletheia skeptics will be very similar, e.g., the sorts of considerations that Fodor and Chomsky raise are not entirely dissimilar.

4.0 The Realm of the Defeaters

We have had occasion to note that justificatory skepticism is typically seen as co-extensive with philosophical skepticism itself. And although I have delineated three other types of skepticism, naturally this in itself does not show that these other skepticisms are viable or indeed philosophically interesting in their own right. In other words, nothing said thus far suggests that skeptics are not right to focus on the justification condition, for it may be the weak link. However, in this section I will argue that the same considerations used to motivate Cartesian justificatory skepticism can be employed with equal effect by the other three types of skepticism. Specifically, Cartesian skeptics suggest that we can conceive of alternatives to our epistemic claims that act as «defeaters» to the knowledge claims. Our first task then will be to consider the general form of the defeater stratagem and then apply it to the other three skepticisms.

4.1 Defeaters and Cartesian Skepticism.

The first step of the argument is to extend the analysis of Cartesian skepticism developed in section 3.12. A. C. Grayling suggests this formulation of the formal aspect of Cartesian skeptic's underdetermination argument:

---

[33] As Davidson, for example, has argued in a series of influential papers collected in *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, (1984).

[34] This sort of position has historical antecedents with F. Schelling's early work.

> Stated succinctly and formally, scepticism is the observation that there is nothing contradictory in the conjunction of statements *s* embodying our best grounds for a given belief *p* and the falsity of *p*.[35]

As we shall see, Graylings formulation (at best) characterizes justificatory skepticism. Confining our attention for the moment to justification we may nonetheless wonder why this observation should present a problem for our ordinary justificatory practices, unless one assumed some sort of (deductive) foundational model of justification. Such a foundational model would insist that the only adequate form of justification is one where *P* can be deduced from some first principles. Such a form of justification would rule out the sort of skepticism Grayling is speaking of, but of course such a view is considered by most to be far too demanding. On the other hand, if we accept this model then it seems we must adopt justificatory skepticism; given, as we noted above, that very few statements (if any) meet this stringent requirement. However, as Stroud and others have argued, this view may actually mistake cause and effect, i.e., foundationalism is an attempt to answer skepticism rather than a cause of skepticism itself.[36]

In fact, Stroud's understanding of Cartesian skepticism suggests that it seems to follow as a natural extension of our everyday justificatory practices. Take a simple example. Suppose Anna asks, «Does anyone know where Dawn is?» and Barry answers, «Yes, she is at the bar». Our normal justificatory practices suggest that if Barry does in fact know that Dawn is at the bar then he has ruled out potential defeaters to his knowledge claim (or is in a position to rule out such defeaters). For example, if Anna says, «Are you sure? I know Dawn is dead-set against drinking, but her (identical) twin sister Kimberly is a barfly if there ever was one.» Let us suppose that Barry is not in a position to eliminate this alternative, (he did not know that Dawn had a twin sister); it seems quite natural to say that Barry does not know that Dawn is at the bar. So in this case we seem to have a very straightforward version of an extremely local version of skepticism. (Such skepticism is of course not a type of philosophical skepticism since its scope is too narrow).[37] More formally, the skeptical result seems to be generated by the following line of thought:

C: If Barry knows that Dawn is at the bar then Barry knows that it was not Kimberly (Dawn's twin) that he saw at the bar.

Since Barry is not in a position to rule out the relevant defeater — that it was Kimberly that he saw — Barry does not know that Dawn is at the bar. We can see then that this sort of local skepticism fits the model suggested by Grayling: Barry's evidence for the claim that Dawn is at the bar is consistent with it being false that Dawn is at the bar. If Barry is to legitimately claim to know that Dawn is at the bar then it seems he must be in a position to reject the defeater, that is, it seems that it is a condition of Barry's knowing that Dawn is at the bar is that he knows that the person he saw there was not Kimberly.

---

[35] A. C. Grayling, «Epistemology», op. cit., p. 53.

[36] Williams says («Epistemological Realism and the Basis of Scepticism» *Mind,* xcvii, no. 387, July 1988, note 10) that the question of whether foundationalism is best seen as a presupposition or a by-product is due to Barry Stroud.

[37] On the other hand, this use of 'skepticism' is well within «everyday» usage of the term. It would be appropriate to say, for example, «Barry claims to know that Dawn is at the bar, although I am skeptical.»

More controversial of course is whether a global skepticism can be generated in the same manner. Formally at least, the Cartesian philosophical skeptic's view is identical to the local skepticism sketched above. Indeed, there is nothing that would prohibit using Cartesian-type thought experiments to underwrite a local skepticism:

C': If Barry knows that Dawn is at the bar then Barry knows that he is not a brain-in-a-vat.

For if Barry does not know that he is not a brain-in-a-vat then he cannot rule out the defeater that what he thinks he saw was in fact merely a computer generated image of a person, i.e., his evidence has nothing to do with whether there was a real person named Dawn at a real bar. Of course the Cartesian skeptic will argue that Barry does not know that he is not a brain-in-a-vat. In any event, if we accept C' as a natural extension of our everyday justificatory practices, and we despair of showing that Barry knows that he is not a brain-in-a-vat, then it follows that Barry is in no position to claim that he knows that Dawn is at the bar.[38] Obviously, whether Cartesian-type defeaters demonstrate that what we do not know is a controversial matter — by 'Cartesian-type defeaters I mean thought experiments like the brain-in-a-vat or the evil demon. What is clear is that *if* it is permissible to invoke such defeaters then they can easily be used to generate a global form of skepticism:

C1: If we know that the standard view is true in the main then we must know that Cartesian-type defeaters are false.

If one of these Cartesian-type defeaters is true, if, for example, we are brains-in-a-vat, then our claims to know the manifest and scientific image would be undermined: our knowledge claims would at best be about a computer generated world rather than the world outside of the macabre experiment that we are unwilling participants in. Cartesian-type defeaters are those skeptical cases where the truth, belief, and gettier components of knowledge are not the subject of skeptical challenge but simply the justification component.

Clearly there are two broad strategies that one might use against the skeptic: one can accept the conditional and show that the skeptic's rejection of the consequent is mistaken, or one can deny the conditional. With respect to the former, one possibility would be to construct a transcendental argument along the lines suggested by Putnam to argue that we know we are not brains-in-a-vat.[39] One way to deny the conditional is to follow contextualists who argue (roughly) that to insist that we eliminate Cartesian-type defeaters is to raise the standards beyond what is accepted by our everyday justificatory practices.[40] Whether this defeater stratagem is viable remains controversial within philosophy; and it is certainly not my intention to enter into this debate. Rather, let me emphasize that the rapid sketch I have provided here is simply meant to *describe* the defeater stratagem, not to *defend* it. What I will

---

[38] See Stroud *The Significance of Philosophical Skepticism, op. cit.* for details on why the skeptic believes that we have no evidence that we are not brains-in-a-vat.

[39] *Reason, Truth and History*, Cambridge: Cambridge University Press, (1981), chapter 1. Putnam himself does not target skepticism with this argument but realism. For a critical examination of various attempts along these lines to make in-roads against skepticism see A. Brueckner's «Content Externalism» in *Transcendental Arguments,* edited by R. Stern, Oxford: Claredon Press, 1999, pp. 229-250.

[40] An overview and review of some of this literature can be found in K. DeRose's «Contextualism: An Explanation and Defence» in *The Blackwell Guide to Epistemology*, edited by J. Greco and E. Sosa, Massachusetts: Blackwell (1999), pp. 187-205.

defend is the claim that, *if* we accept the defeater stratagem, it can be employed by the other three types of skepticism.

4.2 Defeaters and Gettier Skepticism

Let us consider then how the defeater stratagem can be applied to gettier skepticism. To emphasize how this differs from Cartesian skepticism recall Grayling's formulation of Cartesian skepticism: there is nothing contradictory in the conjunction of statements: s embodies our best grounds for a given belief p, and the falsity of p. There are two distinct reasons to reject Grayling's formulation; assuming that it is correct to interpret it as a necessary condition for skepticism. First, as we saw above in the 'Ed the lucky brain-in-a-vat' example, it is quite possible to formulate a global form of skepticism that does not invoke the possibility that our epistemic claims are false. After all, in the case of the kidnapping of Ed we supposed that his beliefs were true even after he had been envatted. Thus, this case does not rely on the observation that our grounds for P are consistent with the falsity of P. Notice too that this example is a type of defeater: if we know that the standard view is true in the main then we must be in a position to eliminate the possibility that we are in exactly the sort of situation described in Ed's case.

A second means to demonstrate that gettier skepticism is a viable independent form of philosophical skepticism is to contemplate the example of the 'Brothers Evil'. Descartes imagined a single evil demon; let us suppose that there are two demons, brothers as it turns out. They bickered over how to deceive people until the Ancient Greek civilization gave them the idea to divide their territories up into Thought and Being. One brother is in charge of manipulating the thoughts and beliefs of people, the other manipulates the world. Twice the error — so they thought. Now imagine that the hapless humans of this example reach a point of ideal justification of the manifest and scientific image, i.e., they have reached what is sometimes referred to as the Piercean end of inquiry. The day of the party celebrating this epistemic victory, brother Thought surgically removes everyone's brain and places them in a vat and hooks up a program intended to radically deceive. One of the beliefs he engenders is that all the bodies of the solar system are made out of different cheeses, the Moon is feta, Mars cheddar and so on. Brother Being working independently had his own plans. (The two brothers had not talked since the publication of *The Critique of Pure Reason* when brother Thought taunted Brother Being by calling him «mere noumena»). Quite independently brother Being had decided to radically rearrange the universe in order to mock the end of enquiry celebrations. As fate in philosophy examples would have it, one of the first things he did was to turn all the bodies of the solar system into cheese, the moon was turned into feta, Mars cheddar, and so on. Of course it turns out that all the bizarre beliefs about the world that brother Thought engendered were exactly matched by bizarre changes in the world orchestrated by brother Being.

To ensure that considerations of truth and justification have been removed let us suppose that we endorse a version of anti-realism that says, necessarily, if our theory is ideally justified then it is true. This form of anti-realism would then maintain that there is a contradiction in the conjunction of statements s embodying our best grounds for a given belief p and the falsity of p. This in itself would demonstrate that Grayling's understanding of skepticism has been refuted in the brothers Evil example. For having reached the end of inquiry there is no room, according to this version of anti-realism, to suppose that our ideally justified theory is not true. If one accepts Grayling's formulation as a necessary condition for philosophical skepticism then this form of anti-realism has defeated philosophical skepticism. Yet if we accept there are alternate formulations of skepticism then this is not the end of the matter. For the

philosophical skeptic still has the «fall back» position of saying that, although our ideally justified theory is true and believed by us, nevertheless, we may not know this theory. The reason of course is that the theory may not be gettier, as the above example demonstrates. The skeptic in effect says: «granted that given anti-realism, our ideal justification leads to truth no matter what the outcome of events, nevertheless, there are better and worse ways this could happen. It would be better if our justification connected in a way that was gettier rather than non-gettier.» It is perhaps worth noting here that the gettier skeptic is able to drive a wedge between justified true belief, on the one hand, and knowledge on the other, because what is meant by 'ideally justified' here is the sort of justification humans might reasonably be expected to amass. To insist that by 'ideally justified' we mean that all the logical possibilities have been eliminated is to insist on a standard of justification that looks unattainable by humans, for it would seem to presuppose that we have a «God's eye view» of the universe. Anti-realists have generally wanted to provide epistemology with a «human face».[41] It is precisely because of this relaxed standard — i.e., not insisting that an ideal theory must eliminate all logical possibilities — that gettier skepticism is able to get a toehold. From the anti-realist's perspective, without this «humanizing» move — i.e., constructing epistemological theories that acknowledge our finitude — we will forever be mired in an unproductive skepticism. What the brothers Evil example shows is if anti-realists hope to eliminate skepticism then they will need to do more than assert that our ideal theory must be true, they must assert that it is also gettier. Perhaps the easiest way for a Piercean anti-realist to achieve this is to stipulate that at the end enquiry our beliefs are not only true and justified they are also gettier. This would guarantee that we have knowledge and not just justified true belief. Without this stipulation there is still a potential gap between justified true belief and knowledge that the skeptic might easily exploit.

Gettier skepticism, then, can be motivated by a natural extension of the Cartesian's strategy to explore defeaters as a means to undermine knowledge: we have seen how it is possible to construct an alternative or defeater to one anti-skeptical strategy that does not require us to question whether the knowledge claim is true or justified. The gettier skeptic, then, maintains that

C2: If we know that the manifest and scientific images are true in the main then we must know that gettier-type defeaters are false.

Gettier-type defeaters are those skeptical cases where the truth, belief, and justification for a putative knowledge claim are not in question; rather, the question is one of whether the putative knowledge claim is gettier. The conclusion to draw is that gettier skepticism is independent of Cartesian skepticism and that it can be supported using the same defeater stratagem.

4.3 Defeaters and Noetic Skepticism

The defeaters stratagem can also be used in support of noetic skepticism, indeed, I think it can be shown that noetic skepticism is epistemologically prior to Cartesian skepticism in the same way that Cartesian skepticism is epistemologically prior to the epistemological stance of common sense and science. I will argue that it may be a very real possibility that we cannot eliminate defeaters that we cannot conceive.

---

41 H. Putnam, *Realism with a Human Face,* Harvard University Press, (1990).

The first step in this argument is to consider how Cartesian skepticism understands the realm of alternative hypotheses or defeaters. A natural interpretation of Cartesian skepticism is that the class of defeaters is extremely broad; any alternative that is conceivable seems to be up for consideration. Clearly the realm of conceivable alternatives is larger and more encompassing than that which is associated with the epistemological point of view of everyday life. Indeed, as our discussion above indicates, in effect the Cartesian skeptic's point is that the realm of defeaters countenanced by our everyday justificatory practices are not sufficiently expansive in their scope.

But is the realm of conceivability the appropriate one, that is, is it expansive enough to include all potential defeaters? There are, I believe, good reasons to think that this realm is also too restrictive. Consider the following sort of example: suppose that there are just two contested hypotheses about the nature of the universe. The standard model, or big-bang theory, H1, describes the universe as evolving from some primordial singularity. The steady state hypothesis, H2, proposes that the universe has always been more or less as it is now. Of course the steady state hypothesis has been rejected by contemporary science; but we can imagine a Cartesian skeptic arguing that we do not know that H1 is true because there is at least a conceptual possibility that we are in error. While our best available empirical evidence supports H1, it is logically possible that H2 is true. We might be asked by a skeptic to imagine an evil demon has arranged all sorts of false clues; e.g., the «alleged» background radiation of the universe leftover from the big bang was simply planted there by the epistemic fiend in an attempt to mislead us. The Cartesian skeptic does not suggest that we are ignorant of the conceptual alternatives, for they allow that we might entertain the possibility that H2 is true. A noetic skeptic could agree that H1 and H2 describe the only two hypotheses about the universe that are worthy of human scientific scrutiny but this does not rule out the possibility that the «complexity» hypothesis H3 is true. It suggests that the theory that best describes the universe must posit a billion billion billion billion billion billion initial conditions, and each of these initial conditions requires at least the same number of bits of information to describe it. Such a hypothesis, let us suppose, is far too complex for any human to conceive, yet it is logically incompatible with the truth of H1 or H2. The noetic skeptic then argues that the possibility of H3 demonstrates that we might forever be ignorant about the truth of the universe. So just as the Cartesian skeptic says that everyday epistemic practices are not sufficiently critical about the scope of relevant alternatives, so the noetic skeptic says that the Cartesian skeptics are not sufficiently critical in their evaluation of the relevant alternatives or defeaters. We would need some guarantee that we are in a position to conceive of the relevant alternatives. Perhaps the Cartesian would have been better off to tackle the problem head-on by appealing to the realm of what is logically possible. However, this option suffers from the mirror imagine of the previous. For if the Cartesian says that the alternatives that are to be rejected are all those that are logically compatible with the evidence then we need some assurance that we are in fact capable of surveying these potential defeaters. The sort of possibility raised by H3 suggests that we may lack the cognitive wherewithal to appreciate in sufficient detail every theory in this set.

There is an obvious and potentially decisive objection here: there is no straightforward way of describing the sorts of defeaters that might lie beyond human abilities to cogitate. After all, if they could be described then a fortiori they would not be beyond human ability to cogitate. In comparison with Cartesian skepticism, noetic skepticism suffers a serious disadvantage in that it is not able to describe in any detail the potential defeaters. Clearly part of the power of Cartesian skepticism is precisely the fact that the sorts of thought experiments Descartes engages in are easy to conceive. Yet to insist on a detailed understanding is to

misunderstand noetic skepticisms, for it is precisely the possibility that we cannot grasp all the relevant possibilities that underlies noetic skepticism. While the noetic skeptic cannot describe in detail the sorts of possibilities that are conjectured to lie beyond our ability to conceptualize, noetic skeptics are not without resources. It is possible for us to see how others failed to conceive of relevant defeaters. One famous example is that historically it seemed reasonable to claim to know that Euclidean geometry best describes the world. A skeptic in the eighteenth century might have pointed out against this position that we cannot rule out the possibility that there are alternate geometrical configurations of the universe that are beyond our ability to conceptualize in any detail.[42] Of course this skeptic would have been at a disadvantage in that she could not describe in any detail the content of these alternate geometries, although, as we now know, subsequent developments in geometry (and physics) proved the skeptic right. In other words, from our epistemically advantaged point we can see that the possibility of such defeaters was indeed correct, given that we now accept a non-Euclidean geometry as the geometry of the universe. It seems we are in no position to eliminate the possibility that some advance civilization stands to us in a similarly epistemically advantaged position. Furthermore, it seems that we are in no position to suppose that there are not congenital limitations to the sorts of possibilities that we can conceive in any detail, just as the realm of possibilities that children and animals are limited from our perspective. A monkey, it seems, cannot even conceive of the thought experiment that it might be a brain-in-a-vat nor perhaps can the average (human) three year old. Given our finitude, it seems that we have every reason to suppose that there may be any number of defeaters beyond our ability to cognize just as there are any number beyond the comprehension of animals and children.

This problem is only exacerbated when we think in terms of how the evil demon or the mad scientist from the brain-in-a-vat scenario might be used in support of noetic skepticism. The mad scientist, for example, is often thought to interface with the kidnapped brain at the point where a brain would normally receive information input from the nervous system. The nerve endings that would normally extend to our sense organs and other parts of our body are said to receive artificial input directed by a supercomputer. But suppose the evil demon or mad scientist took the deception one step further and actually «reprogrammed» our reason such that we have a «blind-spot» in our reason for precisely those concepts and beliefs that describe our actual world. If we allow the possibility that we have purposely been redesigned by some malicious agent so as not to be able to conceive the world then we have a defeater that seems impossible (for humans) to defeat.[43]

The Cartesian skeptic of course says that the justificatory practices of everyday life are not sufficiently critical, that we can conceive of all sorts of defeaters that lie beyond the realm considered relevant to everyday justificatory practices. Likewise, noetic skeptics believe that the realm considered by the Cartesian skeptic is not sufficiently critical, that there may be a number of defeaters that lie beyond our ability to conceive.

---

[42] This skepticism might be a form of socio-historic noetic skepticism: cultural developments have led to the ability to conceive of alternatives to Euclidean geometry.

[43] For simplicity sake I take the set of defeaters of our everyday justificatory practices to be static and univocal. Of course it may well be that the realm of the defeaters is larger or smaller depending on the context of our everyday practices, e.g., the realm of defeaters may be larger where a lot depends on getting the knowledge claim right, as in a murder trial.

4.4 Defeaters and Aletheia Skepticism

The discussion of aletheia skepticism here can be brief because the argumentative structure exactly parallels that of the noetic skeptic: the aletheia skeptic suggests that there are truths beyond our grasp and we cannot rule out the possibility that a defeater lies within this set. Again, the Cartesian skeptic's position does not look sufficiently critical given that we want to pursue the defeater stratagem vigorously. To see this recall the conditional used to support Cartesian skepticism:

C1: If we know that the standard view is true in the main then we must know that Cartesian-type defeaters are false.

The aletheia skepticism suggested by Bradley's claim that «no truth was quite true», points toward the following conditional

C4: If we know that the standard view is true in the main then we know that aletheia-type defeaters are false.

By 'aletheia-type defeaters' I mean of course the sorts of possibilities that Bradley and Chomsky raise: that we may have no or little access to the truth. It seems appropriate then to draw the following conditional conclusion: If the defeater stratagem is viable, and if Bradley's aletheia skepticism is correct (has a higher degree of truth than its negation) then knowledge of the standard view is impossible.

5.0 Prospects for the Fourfold Root of Skepticism

I have attempted to make two points in this paper: First, that a full taxonomy of philosophical skepticism shows that there are alternatives to Cartesian skepticism. Second, if viable, the defeater stratagem can be used to support these alternative skepticisms. Both of these points (obviously) could bear further scrutiny than is possible here. With respect to the former it might prove profitable to examine the historical antecedents of these skepticisms. We noted briefly a possible connection between Kant and Bradley with noetic and aletheia skepticism respectively. Descartes is of course connected with 'Cartesian skepticism', but even Descartes at times seems to have noetic skepticism in mind, although this is not often noted.[44]

Another avenue to explore is how contemporary answers to skepticism figure into this classification. I will mention three, in very broad and rapid brush strokes, to provide some idea of the fecundity of this taxonomy. One example is how causal theories of content or belief are thought to answer skepticism about whether we know that we are not brains-in-a-

---

[44] In the *Meditations* (*The Philosophical Works of Descartes,* volume I, translated by E. S. Haldane, and G. R. T. Ross, Cambridge: Cambridge University Press, (1931), p. 158) Descartes writes:

> But when I took anything very simple and easy in the sphere of arithmetic or geometry into consideration, e.g., that two and three together make five, and other things of the sort, were not these present to my mind so clearly as to enable me to affirm that they were true? Certainly if I judged that since such matters could be doubted, this would not have been for any other reason that it came into my mind that perhaps a God might have endowed me with such a nature that I may have been deceived even concerning things which seem to me most manifest. But every time that this preconceived opinion of the sovereign power of a God presents itself to my thought, I am constrained to confess that it is easy to Him if He wishes it, to cause me to err, even in matters in which I believe to have the best evidence.

In this and other passages Descartes comes close to the noetic (or Kantian) idea that it is not that we may simply be mislead about the nature of sensory evidence, but the nature of our minds might prohibit us from grasping the truth. Unlike Kant, Descartes does not seem to grasp that there are two quite separate problems here.

vat.[45] One anti-skeptical view put forward is, very roughly, that there is a contradiction in saying that 'I am a brain in a vat'. For if I am a brain in a vat then 'vat' means in my language something like a type of computer generated image; but I am not a brain in a computer generated image. Many, I think, have found puzzling the idea that the argument that says that, if I am a brain-in-a-vat then I cannot think that I am a brain-in-a-vat, is an anti-skeptical argument.[46] Certainly this puzzle remains even if one accepts that my utterance of 'I am a brain-in-a-vat' is false. One possibility to explore is that, even if this response is successful against the Cartesian skeptic, one may wonder how robust it is against the noetic skeptic. On this understanding the Cartesian skeptic might be seen as saying that one can conceive of the defeater that 'I am a brain-in-a-vat'. The causal theory of meaning or belief refutes this suggestion by saying that this statement must be interpreted as false. On the other hand, this response to the Cartesian skeptic seems to bring to the fore noetic skepticism: if my beliefs or meanings are circumscribed by the realm of what I causally interact with then it seems there are any number of thoughts and beliefs that I am unequipped to think — including a detailed picture about how I am being deceived. If one accepts this taxonomy of skepticism it suggests a line of defense for the philosophical skeptic against such arguments.

Another example is John McDowell's discussion of skepticism in *Mind and World*. McDowell starts out looking like he is going to discuss noetic skepticism, but as the discussion proceeds it is clear that he has justificatory skepticism in mind.[47] This is perhaps most evident in part of his response to skepticism. One of the problems he struggles with is the idea that we share with animals sense organs that are (in some sense) equivalent in their information gathering abilities, for example, other creatures besides humans have stereoscopic color vision. A crucial difference is that we use the information supplied by our sense organs to justify knowledge claims whereas animals do not. One question McDowell investigates is how this is possible. His answer to this is quite complex, but for our purposes we can concentrate on his proposal (inspired by Aristotle and Kant) that we have a «second» nature that is fused with our animal nature. With this taxonomy of skepticism we can see that, even if McDowell has formulated a reply to the Cartesian skeptic, he has not answered the noetic or aletheia skeptic. For they will wonder about the abilities of this «second» nature vis-à-vis other possible knowers. How do we know, for example, that in addition to an animal and second nature the Alpha Centaurians do not have a «third» nature and the Beta Centaurians a «fourth» nature and so on that eclipses our nature in the same manner that our «second» nature eclipses that of animals? These sorts of questions are not addressed by McDowell. Our taxonomy of philosophical skepticism suggests that they ought to be if the answer to the skeptic is to be complete.

---

[45] I am thinking in particular of the anti-skeptical arguments inspired by Putnam's seminal discussion, *op. cit.*

[46] *Cf.,* John McDowell's reply to Davidson' version of a content externalism reply to skepticism. He says that this «…response to the brain-in-a-vat worry works the wrong way round. The response does not calm the fear that our picture leaves our thinking possibly out of touch with the world outside us. It gives us a dizzying sense that our grip on what it is that we believe is not as firm as we thought.» (*Mind and World*, second edition, Cambridge, Massachusetts: Harvard University Press, (1996), p. 117.

[47] *Cf.* «Introduction» to *Mind and World*, *ibid.*

A more complete answer to skepticism can be found in the work of Donald Davidson.[48] Davidson discusses what we are calling 'justificatory skepticism' most notably in «The Method of Truth in Metaphysics» and «A Coherence Theory of Truth and Knowledge».[49] Although he does not discuss it under the rubric of skepticism, Davidson also has a reply to the noetic and aletheia skeptic in his argument from «On the Very Idea of a Conceptual Scheme».[50] The upshot of this well-known argument is that we cannot make sense of the idea of languages that are true and untranslatable into our own. If this argument is successful then the idea of other knowers like the Alpha Centurians with a transcendent conceptual scheme is incoherent.[51]

The other line of inquiry is the role of the defeater stratagem. As noted, the argument for the viability of these other skepticisms is conditional on the acceptability of the defeater stratagem. This of course raises the question of whether the defeater stratagem is itself viable. Indeed, it seems quite possible that the line of argument canvassed here might be understood by some to constitute a reductio ad absurdum of the defeater stratagem. The idea would be to agree that the defeater stratagem forces us to move beyond justificatory skepticism to noetic and aletheia skepticism, but that these skepticisms are absurd (or perhaps simply pointless). But even if the defeater stratagem is rejected, this in itself does not show that philosophical skepticism in all its varieties is defeated. For it still must be asked whether there are other means to motivate these skepticisms that do not rely on the defeater stratagem. I have argued else, for instance, that naturalism itself (independently of the defeater stratagem) invites noetic skepticism.

References

Bradley, F. H. 1914: *Essays on Truth and Reality*, London.

---------- 1969. *Appearance and Reality.* 2nd edition, Oxford: Oxford University Press.

Brueckner, A. 1999. «Content Externalism,» in *Transcendental Arguments,* edited by R. Stern, Oxford: Claredon Press, pp. 229-250.

Burnyeat, M. 1982: «Idealism and Greek Philosophy: What Descartes Saw and Berkeley Missed,» *Philosophical Review*.

Butts, R. 1984. *Kant and the Double Government Methodology*, Boston: Reidel Publishing.

Chomsky, N. 1980. *Rules and Representations*, New York: Cambridge University Press.

---

[48] In saying that Davidson's answer is more complete I mean that his response covers more of the types of philosophical skepticism that we have discussed. It should not be taken to mean that Davidson is more successful at refuting the philosophical skeptic, this is another matter entirely.

[49] «The Method of Truth in Metaphysics», reprinted in *Inquiries into Truth and Interpretation, op. cit.* and «A Coherence Theory of Truth and Knowledge», in *Truth and Interpretation,* edited by E. Lepore, Oxford: Blackwell, (1986), pp. 307-19.

[50] Reprinted in *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, (1984), pp. 183-198.

[51] Both Rorty («The World Well Lost» in *The Journal of Philosophy*, LXIX (1972), pp. 649-665) and Nagel, *op. cit.* see this implication of Davidson's argument.

Davidson, D. 1984. *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press.

---------- 1986. «A Coherence Theory of Truth and Knowledge,» in *Truth and Interpretation,* edited by E. Lepore, Oxford: Blackwell, pp. 307-19.

DeRose, K. 1999. «Contextualism: An Explanation and Defence,» in *The Blackwell Guide to Epistemology*, edited by J. Greco and E. Sosa, Massachusetts: Blackwell, pp. 187-205.

Descartes, R. 1931. *The Philosophical Works of Descartes,* volume I, translated by E. S. Haldane, and G. R. T. Ross, Cambridge: Cambridge University Press.

Dryer, D. P. 1966. *Kant's Solution for Verification in Metaphysics*, Toronto: University of Toronto Press.

Ewing, A. C. *A Short Commentary on Kant's Critique of Pure Reason*, London: Methuen, 1938

Findlay, J. N. 1981. *Kant and the Transcendental Object*, Oxford: Claredon Press.

Fodor, J. 1983. *The Modularity of Mind*, Cambridge: The MIT Press.

Goldman, A. 1967: «A Causal Theory of Knowing,» *Journal of Philosophy*, 64, pp. 355-72.

Gettier, E. 1963: «Is Justified True Belief Knowledge?», *Analysis*, vol. 23, pp. 121-3.

Grayling, A. C. 1996: «Epistemology», *The Blackwell Companion to Philosophy*, edited by Nicholas Bunnin and E. P. Tsui-James, Cambridge: Massachusetts: Blackwell.

Körner, S. 1955. *Kant*, New York: Penguin Books.

---------- 1969. *Fundamental Questions in Philosophy*, Harmondsworth: Penguin).

Lehrer, K. 1965: «Knowledge, Truth and Evidence», *Analysis*, 25, 168-75.

Lewis, D. 1996. «Elusive Knowledge», *Australian Journal of Philosophy*, 74, pp. 549-67.

McDowell, J. 1996. *Mind and World*, second edition, Cambridge, Massachusetts: Harvard University Press.

Nagel, T. 1986. *The View From Nowhere*, Oxford: Oxford University Press.

Nozick, R. 1981: *Philosophical Explanations*, Cambridge Mass: Belknap Press.

Pippin, R. 1982. *Kant's Theory of Form*, New Haven: Yale University Press.

Putnam, H. 1981. *Reason, Truth and History*, Cambridge: Cambridge University Press.

---------- 1983. There is at least one a priori truth, reprinted in *Realism and Reason*, Philosophical Papers vol. 3 Cambridge Univ. Press.

---------- 1990. H. Putnam, *Realism with a Human Face,* edited by J. Conant, Harvard University Press.

Rescher, N. 1972. Noumenal Causality, in L. W. Beck edits, *The Proceedings of the Third International Kant Congress*, Dordrecht: D. Reidel Publishing.

Pritchard, H. A. 1950. *Knowledge and Perception*, Oxford: Claredon Press.

Rorty, R. 1979. *Philosophy and the Mirror of Nature,* Princeton: Princeton University Press.

Russell, B. 1948. Huma*n Knowledge: Its Scope and Limits*, New York: Humanities Press.

Sellars, W. 1963. Philosophy and the Scientific Image of Man, in *Science, Perception, and Reality* London: Routledge and Kegan Paul, Ltd.

Solomon, R. 1983. *In the Spirit of Hegel,* Oxford: Oxford University Press.

Strawson, P. F. 1966. *The Bounds of Sense*, London: Methuen.

Stroud, B. 1984. *The Significance of Philosophical Scepticism,* Oxford: Claredon Press.

Triplett, T. 1990. Recent Work on Foundationalism, *American Philosophical Quarterly* 27, pp. 93-116.

Walker, R. C. 1978. *Kant*, London: Routledge & Kegan Paul, Ltd.

Westphal, M. 1968. In Defense of the Thing in Itself, *Kant-Studien*, Heft 1, 118-41.

Williams, M. 1988a. Scepticism Without Theory, *Review of Metaphysics* (March): 547-588.

---------- 1988b. Epistemological Realism and the Basis of Scepticism, *Mind,* xcvii, no. 387.

---------- 1999. Skepticism. In *The Blackwell Guide to Epistemology*, edited by J. Greco and E. Sosa, Massachusetts: Blackwell.

Wilbur, J. B., and Allen, H. J. (eds.) 1979. *The Worlds of the Early Greek Philosophers*, Buffalo: Prometheus Books.

Zagzebski, L. T. 1994. «The Inescapability of Gettier Problems», *Philosophical Quarterly* 44, no. 174, pp. 65-73.

Mark Walker
Trinity College, University of Toronto
Ontario, Canada
<mark@markalanwalker.com>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 110-116
Fodor's Epistemic Intuitions of Analyticity
Copyright © by SORITES and Wayne Wright

FODOR'S EPISTEMIC INTUITIONS OF ANALYTICITY

Wayne Wright

<wrightwt@comcast.net>

I. Holism

Semantic holism has it that the semantic properties of an individual expression are determined by that expression's relations to *all* of the other elements in a system of expressions to which it belongs. A standard argument for holism about the semantic properties of concepts goes as follows:[1]

(1) The meaning of a concept C is determined by its intrinsic/constitutive connections to other concepts within the cognitive economy of a thinker T.

(2) There is no principled distinction between the concepts to which C is intrinsically/constitutively connected and those to which it is not so connected.

Therefore,

(3) The meaning of a concept C is determined by T's *entire* stock of concepts.

(1) can be derived from either a conceptual role theory of meaning or verificationism, and (2) is supposed to follow from the failure of the analytic/synthetic distinction.

Holism is often thought to have several highly counterintuitive consequences, particularly regarding the incommensurability both of different thinkers' concepts and of the concepts of an individual thinker at different times.[2] Granting that there are no formal problems with the holism argument, the anti-holist's options seem limited to denying either (1) or (2).[3] Jerry Fodor (1987, 1990, 1998) has responded to the holism argument by denying (1), which leads to his *atomistic* theory of concepts. According to Fodor, the meaning of a concept C is independent of the meaning of any other concept. It follows from this that concepts, contrary to much of philosophical tradition, are not complexes consisting of meaning-constitutive connections with other concepts. This claim about the semantics of concepts has an important consequence for the requirements on concept possession: according to atomism, possessing an individual concept C does not require that one possess any other concept.

---

[1] See Fodor and Lepore (1992), pp.23-26.

[2] For a detailed discussion of the difficulties confronting holism, see Fodor and Lepore (1992).

[3] One might argue that the consequences of holism form a sufficient reductio of the position, and thus one need not concern herself with mounting an attack against the holism argument's premises. This strategy will be set to the side in what follows.

In this paper, I will not directly address Fodor's atomism. Instead, I want to discuss what Fodor (1998) sees as the principal inducement for responding to the holism argument by denying (1) and thereby opting for atomism, the repeated failure of efforts to refute (2) by attempting to make the analytic/synthetic distinction principled. If (1) is maintained and (2) is denied, *molecularism* would result; according to the molecularist, the semantic properties of a concept C are dependent on the semantic properties of some (but not all) of a thinker's other concepts. If an analytic/synthetic distinction could be maintained, the molecularist could distinguish conceptual connections which are meaning-constitutive from those which are not by saying that a concept C's meaning depends on those concepts to which it is analytically related, and any concept to which C is not analytically related does not figure in determining C's meaning. Since many still cling to strong intuitions of analyticity, there is at least some reason for supposing that there are meaning-constitutive interconceptual connections, despite the implausibility of holism.

Fodor sympathizes with such intuitions of analyticity to a certain extent, but denies that they have anything to do with a semantic property of concepts. Instead, he claims that such intuitions are really informing us as to an *epistemic* property of concepts, which is not sufficient for doing the work required to reject (2) of the holism argument. Since holism is unacceptable, the only remaining option is to deny (1) and embrace atomism. In what follows, I will argue that Fodor fails to make a convincing case against analyticity, and thus he cannot rule out molecularism, which is the primary anti-holistic competitor to his own theory.

II. Fodor and Quine

When inquiring into whether there are any analyticities, one is often struck by the following sort of consideration: BACHELOR and UNMARRIED MALE are connected in a way which BACHELOR and LIVES ALONE are not.[4] Put another way, *bachelors are unmarried males* looks to be a different kind of truth than *bachelors often live alone*. Such cases often lead philosophers to harbor intuitions on behalf of analyticity, despite the strong influence Quine's attack on analyticity continues to exert. The pro-analyticity sentiment is bolstered by the suspicion, often felt even in quarters sympathetic to Quine, that it is unclear both what Quine's arguments really were and how successful his arguments should be taken to be. Pairing strong intuitions on behalf of analyticity with the apparent lack of a decisive argument against analyticity might lead some to suppose that we are justified in appealing to analyticity to explain the difference between 'bachelors are unmarried males' and 'bachelors often live alone'; the former, and not the latter, is an analytic truth.

Fodor agrees that the intuitions on behalf of analyticity are strong. However, he claims that the alleged confusion over Quine's arguments is mistaken.

> I am, myself, inclined to think it's pretty clear after all how Quine's main argument against analyticity is supposed to run: namely, that nobody has been able to draw a serious and unquestion begging distinction between conceptual connections that are reliable because they are intrinsic/constitutive and conceptual connections that are reliable although they aren't; and that it would explain the collapse of this project if there were, in fact, no such distinction. (Fodor (1998), p.71)[5]

---

[4] I adopt the practice of capitalizing words which are used to denote concepts.

[5] Hereafter, references to Fodor (1998) will appear in the text, and will be given by page number alone.

It should be noted that it is unclear to what extent Fodor actually endorses the argument he attributes to Quine. Fodor says that his informational semantics will show *why* Quine was right, but he does not clearly indicate that he finds Quine's argument at all convincing. There are places, however, where Fodor seems to not only agree with Quine's negative conclusion regarding analyticity, but also to think that Quine's argument is successful. In Section IV, I will argue that even if it turned out that Fodor wanted to endorse Quine's argument, he cannot do so and remain consistent with what he says about a prioricity and necessity.

A deeper concern about this passage is that the argument attributed to Quine does not constitute a genuine refutation of the analytic/synthetic distinction. Simply put, the fact that we cannot (or, at least, have not yet) draw(n) a principled analytic/synthetic distinction should not by itself be taken to entail that there is, in fact, no analytic/synthetic distinction. That we are unable to draw a principled analytic/synthetic distinction may very well be a necessary condition of there being no real analytic/synthetic distinction, but it is far from obvious that it is a sufficient condition. Further, why suppose that distinguishing between conceptual connections which are meaning-constitutive and those which are not requires that we have on hand a principled analytic/synthetic distinction? Considering the strength of the intuitions of analyticity, there seems to be no *prima facie* reason why we could not take paradigm cases of intrinsic conceptual connection, such as seems to hold between BACHELOR and UNMARRIED MAN, as meaning-constitutive and paradigm cases of conceptual independence, such as seems to hold between BACHELOR and LIVES ALONE, as not meaning-constitutive. *At most*, Fodor's argument manages to shift the burden of proof back on to the supporters of analyticity. However, granting for the moment that the burden-shifting is successful, Fodor still needs some further argument if he is to have a knockdown refutation of the analytic/synthetic distinction, which is what he seems to think he has and what his endorsement of informational semantics requires.

III. What our intuitions of analyticity are supposed to be about

Fodor claims that his atomistic informational semantics is a fallback position from the perceived failure of inferential role semantics (p.146). The repeated failure of attempts to make the analytic/synthetic distinction principled is supposed to lead us to despair over there ever being an acceptable molecularist version of inferential role semantics; according to Fodor, any inferential role semantics will lead directly to holism.[6] As already noted, however, Fodor's problem is that, despite the problems surrounding the analytic/synthetic distinction, there still remain strong intuitions of analyticity. If informational semantics truly is just a fallback position, it has to be shown that there is something which needs falling back from; mere hand-waving about there being difficulties for one's competition will not suffice. Thus Fodor needs to account for there being intuitions of analyticity in a way which shows them to offer no evidence for supposing that there really are meaning-constitutive interconceptual connections.

The standard, Quinean way of deflating analyticity intuitions is to say that cases which appear to exhibit analyticity are instead exhibiting «theoretical centrality.» This might work for 'F = MA' and the like, which could appear to be analytic simply because a great deal of the rest of physical theory hangs on it. However, as Fodor points out, 'bachelors are unmarried males' and 'Tuesdays come before Wednesdays' do not exhibit any such theoretical centrality. As a matter of fact, such truths seem to rest at the outermost edges of the web; they are so highly conventionalized that nothing of theoretical import rides on them at all. Moreover, that

---

[6] See Fodor and Lepore (1992), especially p.31.

they are so isolated from nearly everything else seems to make 'bachelors are unmarried males' and 'Tuesdays come before Wednesdays' *better* candidates for analyticity than 'F = MA'. They look to be *mere* conceptual truths. So, theoretical centrality, by itself at least, cannot be all that there is to a deflationary account of analyticity.

This sort of consideration forces Fodor to supplement the appeal to theoretical centrality, and he chooses to do so with a version of Putnam's notion of a «one-criterion» concept. A one-criterion concept is one for which «there is just one way to tell that [it applies]» (p.80). For Putnam, one-criterion concepts, and not those which exhibit theoretical centrality, are the sorts of concepts which figure in analytic inferences. Putnam's reason for cutting things up this way looks to be grounded in the already noted difference between truths such as 'bachelors are unmarried males' and 'F = MA'. Take BACHELOR as an example of a one-criterion concept: there's just one way to tell whether Smith is a bachelor, which is by checking whether he is an unmarried male, so 'bachelors are unmarried males' is analytic. Thus on Putnam's story, the good news is that it turns out that there really are analyticities, but the bad news is that analyticity is confined to a group of concepts which are of little, if any, philosophical interest. It will turn out that there are conceptual truths involving BACHELOR and TUESDAY, but not involving JUSTICE, GOD, or TRIANGLE.

Unfortunately for Putnam, put to use as an account of analyticity itself, one-criterionhood fails. Fodor points out that the 'just one way to tell' test for analyticity works only if you already have in place some means of individuating criteria.[7] Whether BACHELOR has only one criterion (viz., 'unmarried man') «depends, *inter alia*, on whether 'unmarried man' and 'not married man' are synonyms» (p.82). But, following Quine, 'synonymy' and 'analyticity' are both members of the same troublesome circle of interdefinable terms, so any appeal to synonymy in the explication of analyticity is bound to be circular. Thus Putnam's one-criterionhood does nothing to remove the standard Quinean worries about analyticity.

Since Fodor's project is not to give an account of the metaphysics of analyticity, but instead is to explain our intuitions of analyticity consistent with there being no real analyticities, he feels that he can live with the sort of circularity involved in the notion of a one-criterion concept which dooms Putnam's use of it (p.82). For Fodor, the reason that 'bachelors are unmarried males' and 'Tuesdays come before Wednesdays' seem analytic is that concepts such as BACHELOR and TUESDAY are so highly conventionalized that there are few ways to introduce them to someone who does not already have them, and we come to think of those few ways as being particularly intrinsically connected to that concept. Thus analyticity is not a semantic feature of concepts, as the molecularist insists it is, but is instead epistemic. Since the only apparent way of determining whether someone is a bachelor is to check whether he is an unmarried male, 'bachelors are unmarried males' seems to be analytic, whereas since there are other ways of finding out whether something is a dog besides checking whether it is an animal, 'dogs are animals' seems less, or perhaps not at all, analytic.

One concern which Fodor's Auntie has about this strategy is that if there's one way to tell that something is an X, then there are bound to be *lots* of ways to tell that something is an X. Thus it turns out that there are no one-criterion concepts after all, and Fodor has done nothing to account for intuitions of analyticity. If I can tell that Smith is a bachelor by checking to see if he is an unmarried male, I can tell that Smith is a bachelor by having Jones check whether he is an unmarried male. And, if I can tell that Smith is a bachelor by having

---

[7] Actually, Fodor points out that Jerrold Katz pointed this out to him.

Jones check whether he is an unmarried male, I can tell that Smith is a bachelor by having Jones ask Williams to check whether he is an unmarried male. This can go on indefinitely. Fodor responds to this worry by saying that «some ways of telling pretty clearly depend on others. It's …the pretty clearly independent ones …that you are supposed to count when you decide whether something's a one-criterion concept» (pp.82-3). According to Fodor, whatever way I devise of determining whether Smith is a bachelor will depend (in some way) on determining whether he is an unmarried man, and whatever way I devise of determining whether today is Tuesday will depend on determining whether today is the second day of the week.

Although I agree to some extent with Fodor regarding the particular example Auntie brings up, his response to the general point raised by Auntie's objection fails (Auntie never was one for details, I suppose). Contrary to what Fodor says, I can think of ways in which I might determine whether today is a Tuesday which do not depend on my performing inferences involving SECOND DAY OF THE WEEK. If I know that today is 9 December, and I know that 2 December was a Tuesday, and I know that a week contains seven days, then I can determine that today is Tuesday without ever knowing what place Tuesday (or any other day, for that matter) has in the order of the days of the week; SECOND DAY OF THE WEEK does not at all figure in my inference-process. If I'm working as a telemarketer, and I'm given a list of people to call, and I'm told all of the people I am to call are bachelors, and I find Smith's name on that list, then I can determine that Smith is a bachelor without ever deploying UNMARRIED MAN in my inference-process. Just as there is a «plethora» of reliable ways of determining that something is water besides checking whether it's H_2O, the only apparent limit on the number of reliable ways one might determine whether today is a Tuesday or Smith is a bachelor is one's imagination.

Knowing that there are lots of ways of determining that something is a bachelor or a Tuesday, however, doesn't sway one's intuitions of the analyticity of 'bachelors are unmarried males' and 'Tuesdays comes before Wednesdays' (or, at least, it doesn't sway my intuitions). For that matter, that there are plenty of ways to determine whether something is a dog which do not involve inferences in which ANIMAL is deployed doesn't make 'dogs are animals' seems any less analytic, despite Fodor's claims otherwise. Fodor's appeal to one-criterionhood to account for intuitions of analyticity seems quite unsuccessful. And, as already noted, if Fodor cannot come up with a way to explain away intuitions of analyticity by showing them to be of something other than analyticity, then there is simply no reason to follow him in rejecting analyticity and thereby opting for his atomism.

IV. Fodor and Quine, again

Getting back to a point raised in Section II about what to make of Fodor's discussion of Quine's argument against analyticity, it seems as though any appeal to Quine by Fodor would be quite strange, because it would undercut other important parts of his own theory. Fodor, like Quine, is clearly opposed to analyticity, but, unlike Quine, Fodor is congenial towards a prioricity and necessity. The oddness of Fodor's *even mentioning* Quine in his own discussion of analyticity is brought out in the following passage:

I think it's reasonable, on the philosophical evidence, to suppose that such conceptual connections *don't* exist. Quine was likely right about conceptual connections, even though he was wrong about necessity and a prioricity, both of which are, so I suppose, very important and perfectly real. (p.87)

The problem with this sort of thinking is that 'necessary' and 'a priori' put you directly in the troublesome circle of interdefinable terms, which also includes 'analyticity', that Quine criticized in «Two Dogmas».[8] In fact, Quine showed how, if you had a prioricity and necessity, you could get analyticity. Thus if necessity and a prioricity are real, and Fodor's argument is Quine's argument, it is fair to ask: What's so unreal about analyticity? So, it is clear that Fodor is not relying on Quine to make his case against analyticity. One has to wonder what Fodor *could* be relying on.

I certainly do not mean to say that the general position Fodor takes is an impossible one to maintain. One could follow Kripke (1980), who goes to some length to show that necessity, analyticity, and a prioricity are *very* different notions and that we would be wise to keep them separate. So, there's no *prima facie* reason to doubt that necessity and a prioricity are okay and analyticity is not. But, to hold that, one of the things you absolutely need to have is a convincing argument against analyticity. Well, if Fodor's anti-analyticity argument isn't Quine's anti-analyticity argument, then what is his anti-analyticity argument?

Fodor rightly points out that if his informational semantics turns out to be true, then analyticity is *ipso facto* ruled out. But, since atomism is only a fallback position from the perceived failure of the analytic/synthetic distinction, a satisfactory deflationary account of our strong intuitions of analyticity is needed before Fodor's position should be accepted. Otherwise, Fodor begs the question; because the viability of informational semantics rides on the success of the anti-analyticity argument, it seems to me clearly illegitimate to use informational semantics as a premise in an argument against analyticity. As was noted at the end of Section III, the appeal to one-criterionhood will not provide the needed deflationary account of analyticity. Thus all Fodor really seems to have established is that informational semantics and analyticity are incompatible. However, that does not in any way constitute a substantial conclusion; we knew that from the start. Thus unless Fodor has tucked up his sleeve an anti-analyticity argument of his own (one which does not have informational semantics as a premise), he has done nothing to show that molecularism, the anti-holistic position competing with his atomism, is false. Given Fodor's failure to decisively rule out the primary competitor to his theory and the fact that his informational semantics is burdened with problems of its own, it appears very much as though molecularism is still an open option and that there is no reason at all to buy into Fodor's atomism.

VI. References

Block, N. 1986. Advertisement for a Semantics for Psychology. *Midwest Studies in Philosophy* 10: 615-78.

Block, N. 1993. Holism, Hyper-analyticity and Hyper-compositionality. *Mind and Language* 8: 1-27.

Fodor, J. 1987. *Psychosemantics*. Cambridge: MIT Press.

Fodor, J. 1990. *A Theory of Content and Other Essays*. Cambridge: MIT Press.

Fodor, J. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Fodor, J. and Lepore, E. 1992. *Holism: A Shopper's Guide*. Cambridge: Blackwells.

[8] And to which 'synonymy' belongs, which makes it unclear how Fodor could swallow Katz' objection to Putnam's notion of a one-criterion concept.

Kripke, S. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Quine, W.V. 1953. *From a Logical Point of View*. Cambridge: Harvard University Press.

Wayne Wright
Philadelphia, PA. USA
<wrightwt@comcast.net>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #14 — October 2002. Pp. 117-121
Wittgenstein: Transcendental Idealist?
Copyright © by SORITES and John M. Weyls

WITTGENSTEIN: TRANSCENDENTAL IDEALIST?

John M. Weyls

<Weyls@juno.com>

In Jonathan Lear's and Barry Stroud's essay «The Disappearing We,»[1] Lear presents Wittgenstein as transcendental idealist and parallels him with Kant. Stroud, while willing to grant some degree of Kantianness to Wittgenstein, is unwilling to press the parallel as far as Lear does. I will argue that both Lear's account of Wittgenstein as Kantian, and Stroud's objections as to the extent to which the parallel can be taken, are fraught with difficulties. I will attempt to show that the difficulties center on what I take to be Wittgenstein's paradoxical relationship with synthetic a priori judgments. If, like Kant, Wittgenstein holds them to undergird the sciences, then, contrary to what he maintains, he is not entitled to hold that concepts different from the ones we are used to are intelligible. On the other hand, if Wittgenstein rejects them and, consequently, their foundational status, he is committed to either one of two views, both of which he seems to reject — that mathematical statements are revisable in light of empirical facts, or that they are mere tautologies.

I will try to summarize Kant and Wittgenstein briefly. Kant was concerned to establish the objective validity of certain concepts. He found that while philosophers such as Hume were correct to emphasize an epistemology tied to empirical representations, the problem of not having an empirical representation for a causal concept unavoidably led to skepticism. If Hume was right, associating causes with effects was merely a psychological process, one from which conclusions drawn about causal efficacy were inductively obtained.

However, Kant rejected a metaphysics that lacked a touchstone in experience. He says in the introduction to the first *Critique*,

> once we are outside the circle of experience, we can be sure of not being *contradicted* by experience. The charm of extending our knowledge is so great that nothing short of encountering a direct contradiction can suffice to arrest us in our course; and this can be avoided, if we are careful in our fabrications — which none the less will still remain fabrications (A4/B8).

Kant's project, in part, consists in an attempt to bridge the gap between metaphysical opposites, that is, between subjective conditions of thought and what we would like to think is objectively the case. He thinks that by showing subjective conditions to be the very conditions necessary for experiencing objects, their objective validity is secured. Synthetic a priori judgments articulate these conditions; synthetic because they enlarge our knowledge about experience, a priori because they are the necessary conditions under which experience itself is possible. Accordingly, the concepts conveyed by these judgments are constructed a priori, as they reflect these conditions; they express, with the strictest universality, how

---

[1] Jonathan Lear and Barry Stroud, «The Disappearing We,» *Proceedings of Aristotelian Society*, supplemental vol. 58 (1984): 219-42.

experience is, must be, and will be. By negotiating between, on one side, what we take to be objective and, on the other, subjectivity, Kant thinks he stakes out a more tenable middle position, one that reconciles sense and mind.

Wittgenstein, too, negotiated between two camps. Steve Gerrard argues that:[2]

> The history of philosophy can partially be characterized by what Hilary Putnam has called the *recoil* phenomenon: an oscillation between two extreme positions, with each camp reacting to the untenable part of the other, resulting, finally, in *two* untenable positions … on one side there are those who deny objectivity in all fields in all ways; there are only incommensurable narratives. On the other side are those who attempt to secure objective validity, but do so at the cost of clothing it in metaphysical mystery … Wittgenstein argued against both sides. His ultimate achievement in the philosophy of mathematics was to stake out a defensible intermediate position between two untenable warring factions (CCW 171).

It's not hard to see that both Kant and Wittgenstein are essentially reacting against the same things. That is, psychologism on one hand, and, if not objectivity itself, the *metaphysics* of objectivity on the other. Like Kant's, Wittgenstein's is a critical philosophy. Newton Garver goes as far as to say that «One cannot read any of his works without a powerful sense of familiar ideas being subjected to devastating scrutiny» (CCW 162). But if I am right in what follows, Wittgenstein's arrows of devastating scrutiny in fact turn out to be boomerangs.

First, consider one particular conclusion Lear draws from analogizing Wittgenstein with Kant. Lear stresses a strong tie between Kant's synthetic unity of apperception and his own claim that for Wittgenstein, the inclination to partake in a community of language users or rule followers arises from being 'so minded' to share the same cognitive dispositions.[3] That is, they agree in a «form of life.»[4] For Kant, the synthetic unity of apperception results from a spontaneous act that is capable of generating the representation 'I think' concomitantly with every other representation.[5] More simply, representations are accompanied by a consciousness, e.g. mine, conceived as that required to carry out the syntheses from which the representations are the result. If we construe Wittgenstein's investigations as broadly transcendental, Lear argues:

> we find ourselves as speakers and understanders of language that is used both as a means of thought and of communication, and ask 'what must be the case for *this* to be possible? … One answer is that the representations to which I am able to append an 'I understand' must also be capable of accepting a 'we are so minded' (230-31).

and further,

> we come to see that there is no concept of being 'other minded.' The concept of being minded in any way at all is that of being minded as we are. To put it in Kantian terms: language is that in the concept of which the manifold of our representations is united (233-34).

---

[2] Steve Gerrard, «A Philosophy of Mathematics Between Two Camps,» in *The Cambridge Companion to Wittgenstein*, ed. Hans Sluga and David Stern (New York: Cambridge UP, 1996), 173.

[3] Lear and Stroud, 229-30.

[4] Ludwig Wittgenstein, *Philosophical Investigations*, 3rd ed., trans. G. E. M. Anscombe (New Jersey: Prentice Hall, 1958) paragraph 241.

[5] Immanuel Kant, *Critique of Pure Reason*, trans. Norman Kemp Smith (New York: Random House, 1958) B 132.

For Lear, othermindedness is rejected by Kant and Wittgenstein equally; it amounts to not synthesizing the sensible manifold according to conditions Kant thinks necessary for such synthesis, or, in Wittgenstein's case, not sharing a «form of life.» Kant's experience is Wittgenstein's language. Othermindedness means having neither.

Lear suggests that on Wittgenstein's account, something like the following is the case: '7 + 5 = 12' — Kant's exemplar of a synthetic a priori judgment — is the kind of proposition we are disposed to assert. When we study the *Investigations*, we come to assert that 'we are so minded' to assert '7 + 5 = 12.' After realizing that we could not be 'other minded,' we also realize that there is no alternative possibility to '7 + 5 = 12.'[6]

I want to press the idea that Wittgenstein accepts the synthetic a priori in much the same way Kant does — as foundational to both mathematics and natural science — i.e., serving as objectively valid principles upon which their epistemic claims rest, only to uncover a problem that holding such a position involves. Stroud quotes Wittgenstein to refute Lear's claim that Wittgenstein's philosophy provides no room for alternative ways of thinking, and I point to the same passage to make a different argument. Wittgenstein says that if the formation of concepts can be explained by facts of nature, we can, for purposes of illustration, «invent fictitious natural history» and that

> I am not saying: if such-and-such facts of nature were different people would have different concepts (in the sense of a hypothesis). But: if anyone believes that certain concepts are absolutely the correct ones, and that having different ones would mean not realizing something that we realize — then let him imagine certain very general facts of nature to be different from what we are used to, and the formation of concepts different from the usual ones will become intelligible to him (PI p230).

Now if Wittgenstein's investigations are transcendental, then it's reasonable to think that he accepts some form or other of a priori concepts. But if he accepts them in the way Kant does, i.e., as being comprised of both analytic and synthetic propositions, then he is not entitled to imagine general facts of natural history, except in a trivial sense, to be different from the ones we are used to. All Wittgenstein is entitled to imagine are *a posteriori* differences — different particular facts, such as, for example, that Vesuvius did not erupt in 79AD or that he is not the author of the *Investigations*, and the different 'very general facts' that we may imagine inductively follow from the different empirical content of particular representations. One can imagine, say, different quantitative magnitudes of empirical constants employed in mathematical hypotheses, e.g. a gravitational constant of 12.8m/s squared instead of 9.8, or a virtually infinite number of qualities predicated to any object one can think of that may differ in every respect from those encountered in experience. But these differences are trivial in the sense that what is objective in the concepts remains unchanged: Whatever their empirical content, the concepts conform to principles conceived within the limits of what is possible under subjective conditions of thought. If we press the Kant analogy, such concepts might conform to quantity, quality, relation, and modality. Finally, Wittgenstein's remarks concerning the non-revisabilty of mathematical statements suggests support for his acceptance of a priori principles: «Mathematics as such is measure, not thing measured» he says (RFM III-75). In line with the above comment about mathematical hypotheses, Wittgenstein

---

6 Lear and Stroud, 238.

conceives mathematics as changing only in light of mathematical considerations, not empirical ones.[7]

While Stroud's objection to Lear's account succeeds by rightly pointing out that Wittgenstein *says* 'alternative' concepts are possible, his position fails. If Wittgenstein rejects the objective validity of synthetic a priori judgments, then he is entitled to imagine 'othermindedness.' But if this is so, then he must accept that either the judgments giving rise to the fundamental concepts employed in natural science and the mathematical hypotheses they entail are potentially revisable in light of empirical considerations, or that these judgments are not synthetic, and the 'very general facts' of nature are reducible to tautologies — a position that Wittgenstein came to reject before writing the *Investigations*.[8] If such judgments are not a priori, we have the former; if they are not synthetic, we have the latter. Keeping in mind what he says about the nonrevisability of mathematical statements, along with his rejection of a kind of early Russellian Platonism,[9] neither view is one that I think Wittgenstein would embrace.

Finally, there are, of course, similarities that emerge when comparing Wittgenstein with Kant. Lear is right to point to an analogy between Kant's synthetic unity of apperception and the conditions under which he thinks, in Wittgenstein's sense, make rule-following and language-use possible. Gerrard's discussion brings to light Wittgenstein's attempt to negotiate the interminable debate among ideologues of competing epistemologies; and the «devastating scrutiny» to which Wittgenstein subjected subjectivist and objectivist ideas alike smacks of Kant. But any substantive comparison must take into account whether, and the sense in which, Wittgenstein holds the synthetic a priori. Neither Lear nor Stroud address this issue squarely. Unmistakably, Wittgenstein is Kantian. But the question as to whether his philosophy is remains.

References

Garver, Newton. «Philosophy as Grammar.» In The Cambridge Companion to *Wittgenstein*, eds. Hans Sluga and David Stern, 139-70. New York: Cambridge UP, 1996.

Gerrard, Steve. «A Philosophy Between Two Camps.» In *The Cambridge Companion to Wittgenstein*, eds. Hans Sluga and David Stern, 171-97. New York: Cambridge UP, 1996.

Lear, Jonathan, and Barry Stroud. «The Disappearing We.» *Proceedings of the Aristotelian Society*. supplemental vol. 58 (1984) 219-42.

---

[7] Wittgenstein, *Remarks on the Foundations of mathematics*, eds. G. E. M. Anscombe, G. H. von Wright, and R. Rhees, trans. G. E. M. Anscombe (Oxford: Blackwell, 1956) III-75.

[8] A. N. Prior, «Logic, History of: Modern Logic: Frege to the Present,» in *The Encyclopedia of Philosophy*, vol. 4: 559-60.

[9] The comparison of Russell to Plato is drawn with respect to the *logicism* project, i.e. Russell's attempt, along with Whitehead and Frege, to reduce the whole of mathematics to logic. Since the propositions of logic are tautologous, their argument entails that the same is true of mathematical statements. As in Plato's epistemology, the claim is that mathematical statements say nothing at all about the sensible world. Though Wittgenstein's philosophy of mathematics changed over time, he always repudiated *logicism*. For more on Wittgenstein's philosophy of mathematics, see Steve Gerrard's «A Philosophy of Mathematics Between Two Camps,» in *The Cambridge Companion to Wittgenstein*, eds. Hans Sluga and David Stern (Cambridge UP, 1996) 171-97.

Kant, Immanuel. *Critique of Pure Reason*. Translated by Norman Kemp Smith. New York: Random House, 1958.

Wittgenstein, Ludwig. *Philosophical Investigations*. 3rd ed. Translated by G. E. M. Anscombe. New Jersey: Prentice Hall, 1958.

Wittgenstein, Ludwig. *Remarks on the Foundations of Mathematics*. Edited by G. E. M. Anscombe, G. H. Wright, and R. Rhees. Translated by G. E. M. Anscombe. Oxford: Blackwell, 1956.

John M. Weyls
Philosophy, SIUC
Carbondale, IL.
<Weyls@juno.com>

SORITES

An Electronic Quarterly of Analytical Philosophy

ISSN 1135-1349

COPYRIGHT NOTICE AND LEGAL DISCLAIMER

© 1996 The SORITES Team

Please, read!

(1) ***SORITES*** is not in the public domain. In accordance with international Law (especially the Berne Convention for the Protection of Literary and Artistic Works and the Universal Copyright Convention), this issue of ***SORITES*** is Copyright-protected throughout the Planet.

(2) The Copyright of this issue of ***SORITES*** taken as a whole is held by the electronic publisher (the SORITES team).

(3) The Copyright of the papers published in ***SORITES*** is retained by the individual authors, except that: (i) no part of any such paper may be printed or displayed elsewhere or incorporated into a book, an anthology or any other publication of any sort until **SORITES** has accorded the author(s) permission to that effect [which will be done routinely and quickly, provided ***SORITES*** is therein clearly and explicitly mentioned as the primary source]; and (ii) the authors agree to abide by the other terms and conditions contained in this Copyright Notice.

(4) The authors of the included papers and the electronic publisher, the **SORITES** team, — whether jointly or separately, as the case may be — hereby reserve all rights not expressly granted to other parts in this Copyright Notice.

(5) In compliance with Spanish Law, this issue of ***SORITES*** has been legally registered, three diskette-copies being deposited with the competent authorities, namely the «Deposito Legal» office of the Autonomous Community of Madrid, c/ Azcona 42. (Legal Deposit Registration: M 14867-1995.)

(5) A licence is hereby granted without fee for anybody to freely make as many unmodified copies as they wish of this issue of ***SORITES*** **IN ITS INTEGRITY**, give such copies to anyone, and distribute this issue of ***SORITES*** via electronic means or as printed copies, **PROVIDED** no part thereof is altered or omitted, and especially **NEITHER THIS COPYRIGHT NOTICE NOR THE COPYRIGHT BOXES ON TOP OF THE DIFFERENT PAPERS ARE REMOVED, AMENDED, OR OBSCURED**.

(6) In this context, *the issue of* ***SORITES*** *as a whole* is meant to consist in: either (i) a single file (be it its official version as a WordPerfect 5.1 document or any unofficial version released by the **SORITES** team as an undivided file); or (ii) a collection of files produced by slicing one of the entire-file versions in order to facilitate handling, browsing or downloading. In the latter case, the conveyor is bound to distribute the whole collection. (In this context printed copies of this issue of ***SORITES*** are taken to be equivalent to electronic copies, their distribution being subject to the same conditions.)

(7) This issue of ***SORITES*** may be sold for profit or incorporated into any commercial material only with the previous explicit consent granted by the SORITES team. Otherwise, no fee may be charged for its circulation. An exception is granted to non-profit organizations, which are hereby authorized to charge a small fee for materials, handling, postage, and general overhead.

(8) Private copying of single papers by any lawful means is allowed only when done in good faith and for a fair use, namely for purposes of teaching, study, criticism or review; but no part of this issue of ***SORITES*** may be conveyed to another individual or to a gathering — whether in writing or through oral teaching or by any other means — unless the source is clearly and explicitly acknowledged.

(9) In particular, no part of this issue of ***SORITES*** or of any paper therein included may be conveyed to others by means of reproduction, quotation, copy or paraphrase, without a clear and explicit acknowledgement of the issue of ***SORITES*** and its date, the author's name and the paper's full title. Whenever the quotation occurs within a publication, it is also mandatory to mention the official pages (as shown within the Copyright box on top of the paper), the ISSN (1135-1349) and the official home site of electronic display, namely **http://sorites.net**.

(10) Any perpetration of, or complicity with, unfair use of copies or partial copies of this issue of ***SORITES***, or of papers therein included, especially forgery or plagiarism — being, as it is, an infringement of the authors' and the electronic publisher's rights — is in any case a civil tort, but may even be a crime under current legislation.

(11) This issue of ***SORITES*** is provided «as is», without any guarantee of any kind. The electronic publisher, the **SORITES** team, disclaims all warranties, whether expressed or implied, including, without limitation, the implied warranties of fitness for any particular purpose with respect to the papers included in this issue. By furnishing this document, the **SORITES** team does not grant any license or endorses any commitment except in so much as explicitly set forth in the present Copyright Notice.

(12) The electronic publisher, the **SORITES** team, does not necessarily agree with the authors' views or arguments. The electronic publisher cannot certify the accuracy of any quotations or references contained in the papers.

(13) Each author vouches alone for the originality of the papers they submit to ***SORITES*** and for their compliance with established Copyright laws. Acceptance of a manuscript is done in good faith under the assumption the originality claim is truthful. The electronic publisher — i.e. the **SORITES** team — does not pledge itself for the accuracy of such declarations.

(14) The **SORITES** team cannot be responsible for any real or imaginary damages suffered as a result of downloading, reading, using, quoting or circulating any materials included in this issue of ***SORITES***. The user assumes, at their own risk, full responsibility for the proper use of this issue of ***SORITES***.

(15) Downloading, reading or in any other way using this issue of ***SORITES*** or any part thereof entails full acceptance of the hereinabove stated terms and conditions. If, after downloading a file containing this issue of ***SORITES*** or a part thereof, a user fails to agree to the conditions and terms contained in this Notice, they must discontinue using the material and irrecoverably erase or destroy the downloaded file, so as not to occasion any third-part's unfair use thereof.

(16) Although, thanks to a permission kindly granted by the system's administrators, this electronic journal is displayed at the internet host <www.ifs.csic.es/sorites/> (hostname whose current official IP is 161.111.10.61), which belongs to the Spanish institution CSIC, the journal is not published or sponsored or endorsed by the CSIC, the only owner and publisher being the **SORITES** team.

(17) A specific licence is hereby granted for this issue of ***SORITES*** — and all future issues of the journal as well — to be freely displayed by any BBS and any Internet node or site, provided all conditions stated above are fully aheared to. No previous consent of the

SORITES team is required for such a display, which may be in the form of FTP, Gopher, http-WWW or any other electronic means.

Madrid. April 10, 1995

(Updated: December 31, 1996; April 27 1998; October 3 2002)

The **SORITES** Team

RELEASE NOTICE

This issue of ***SORITES*** is made available in several formats, but its only official version is that released with filename:

sorite14.wp

which is the only file within the archives: **sorite14.zip**. A PDF file (**sorite14.pdf**) is also released; it has been generated from the file **sorite14.wp**.

Although each version, whether official or not — as initially released today (2002-10-13) by the **SORITES** team — is an entire seamless file, it may be splitted down into chunks in order to facilitate downloading, browsing, transferring or e-mailing. In such cases, the unity of this issue of **SORITES** as a whole must be preserved by keeping the ensuing collection intact.