SORITES (Σωρίτης) ®

SORITES (Σωρίτης)

ISSN 1135-1349

ROLL OF REFEREES

_____

SORITES (Σωρίτης)

ISSN 1135-1349

Issue #18 — February 2007

TABLE OF CONTENTS

———————————————————

ABSTRACTS OF THE PAPERS

————————————————

SAMPLES AND LOGICAL FORM

by M. G. Yoes

This paper concerns a puzzle about the logical form of expressions involving repeated 'and's. Obvious attempts to solve the problem fail. It happens that a proper solution surprisingly requires the semantic notions of sampling and expression. I speculate that this may provide an alternative to conversational implicature.

————————————————

MAY JUDGES SOMETIMES LIE?

by Jason Glenn

Clearly, vagueness is an inevitable feature of natural language. The practice of law, which is necessarily done within the confines of language, must inevitably come to grips with the phenomenon of vagueness. What philosophers working with these concerns tend to debate about is the extent to which vagueness affects law and the role that vagueness plays therein. Roy Sorensen argues that vagueness in law is not functional and therefore to be avoided as much as possible by judges. Sorensen equates the decisions that judges make about borderline cases with lying. Sorensen bases his conclusions in part upon the idea that borderline cases do admit of truth-values; it is simply the case that we do not have the means to discover these truth-values. I find that Sorensen may be too quick on the draw in accusing judges of the moral indiscretion of lying, and that judges, though perhaps acting insincerely in a truly minimal sense, are in fact not lying.

————————————————

THE CONTEXTUAL NATURE OF COGNITION AND DANCY'S MORAL PARTICULARISM

by Mark Lovas

We can shed new light upon Jonathan Dancy's moral particularism if we frame it in terms of Daniel Andler's recent discussion of the *epistemological* problem of context. Andler helps us in two specific ways. First, we can see that Dancy's work is highlighting the problem context raises for moral knowledge. This makes some criticisms of Dancy seem off the mark. Secondly, Andler's approach also helps us understand why Dancy seems reticent to provide more epistemological details. Nonetheless, the paper closes with a suggestion about the

possibility of a particularist account of moral knowledge more detailed than anything Dancy has provided.

————————————————————

KUHN, METAPHYSICAL REALISM, AND REDUPLICATION

by Andrew M. Bailey

I offer in this paper an analysis of one troubling passage in Thomas Kuhn's *Structure of Scientific Revolutions*. First, I lay out a dilemma for Kuhn; that he is confronted with either an incoherent formal contradiction or a form of metaphysical anti-realism. Second, using reduplicative propositions, I attempt to carve out a third reading which avoids the dilemma. Finally, I answer two objections to the coherence of my reduplicative reading. In doing this, I dismantle one reason many have found compelling to view Kuhn as a metaphysical anti-realist.

————————————————————

NONSENSE AND THE PRIVACY OF SENSATION

by Juan Jose Acero

This paper explores the so-called Epistemic Privacy Way, one of Wittgenstein's lines of attack on the very possibility of a private language. The Epistemic Privacy Way has it that sentences like 'I know I am in pain' or 'But I must know whether I am in pain!', among many others, cannot be used as vehicles of a sort of knowledge incorrigible and immediate. To substantiate such a criticism, Cora Diamonds views of nonsense are spelled out and a constraint, i.e. the Meaning Restricted Exportability Requirement (MRE), is distilled out of them. Beside holding, in opposition to Diamond, that MRE is compatible with the Principle of Meaning Compositionality, it is argued that sentence constituency is subject to two kinds of conditions, namely, syntagmatic and paradigmatic, which contribute to a detailed explanation of why sentences like 'I know I am in pain', 'It is 5 o'clock p.m. on the Sun' or 'Spain is above New Zeland' could be nonsensical. Another result of making MRE to bear on this topic is a distinction between meaning and thought — a (declarative) sentence could be meaningful but express no thought at all — that echoes back Wittgenstein's distinction between depth and surface grammar.

————————————————————

AN ARGUMENT AGAINST EXTERNAL REASONS

by Jonathan Anomaly

In this article I first clarify and then defend Bernard Williams' claim that all practical reasons are internal. I argue that since external reasons are incompatible with a plausible version of the ought-implies-can principle, they are all false. Although some defend internalism by asserting that external reasons fail to explain rational action, a better defense appeals to the fact that only internal reasons are consistent with the ought-implies-can principle.

————————————————————

JUDGING LIFE AND ITS VALUE

by Brooke Alan Trisel

Some philosophers have noted that there is a difference between the questions «Is life meaningful or purposeful?» and «Is life worth living?» However, what it is that makes these questions different has not been explored. Although the famed question about the meaning of life has received most of the attention, I argue that the better and more fruitful question is «Is life worth living?» When addressing the question «Is life worth living?» one takes into account benefits and costs that are ignored when considering the question «Is life purposeful?» thereby explaining how life can be purposeless, but worth living or purposeful, but not worth living. After exploring the differences between these questions, I then evaluate various ways of determining whether life is worth living. I argue that the best method for making this determination is to ask oneself whether or not one would have chosen to live one's life.

————————————————————

NORMATIVITY WITHOUT EXCEPTION: DONALD DAVIDSON ON LANGUAGE AND COMMUNICATION

by Sonia Arribas

This paper deals with three texts by Donald Davidson's that discuss the issue of linguistic innovation (in metaphor and other devices) from different epistemological standpoints. I show that Davidson's theory of metaphor undergoes a crucial transformation: from an early stage in which metaphor is viewed as an unintelligible and exceptional element external to language, to a later stage in which it is conceived as intrinsic to language, and thus as potentially understandable. This tracing of Davidson's development leads me to formulate an understanding of the concepts of metaphor, literalness and linguistic renovation that follows the later Davidson. According to this formulation, both metaphor and literal meaning are comprehensible, and must therefore be said to exist in a relationship of mutuality. This relationship of mutuality is language's inherent normativity: the fact that language necessarily depends both on the speaker's and hearer's commitment to the literal meaning of the words employed, and on their endorsement of the intended meaning as that which is actually meant. Without committing themselves to the literal, speaker and hearer simply cannot understand the usual reference of words; without committing themselves to the intended meaning in its necessary relation to the literal, speaker and hearer cannot understand one another on each new occasion of utterance.

————————————————————

THE PRIVATE LANGUAGE ARGUMENT ISN'T AS DIFFICULT, NOR AS DUBIOUS AS SOME MAKE OUT

by Roger Harris

A simple argument against private rule-following and, hence, a private language, can be disentangled from Wittgenstein's discussion. It is based on three widely accepted premises concerning:

- the necessary privacy of experience you *own*,

- the subjectivity of judgements about those private matters, and

- the necessity, for a rule to be followed, that following or breaking that rule be equally possible.

1 & 2 prevent 3 being met within the necessarily private sphere. This saves the PL Argument from Brian Garrett's strictures, and the convolutions of Michael Ming Yang's exegesis in recent volumes of Sorites.

————————————————————

FACTS, TRUTH, AND REALISM: TOWARD A MULTI-LEVEL THEORY OF KNOWLEDGE

by Paul Bowyer

Although there are strong arguments in favor of Fodor's «language of thought hypothesis,» it creates the following difficulty concerning the relationship between mind and reality: if the mind is composed of mental sentences, which in turn consist of symbols arranged in sequences, whereas external reality consists of physical objects arranged in space, then how can there be a simple relationship between the two? I suggest that to answer that question, we must formulate a metatheory, in the language of possible-worlds semantics, which essentially views reality as composed of sets of possible worlds that correspond to the sentences of our scientific theories. These sets of possible worlds can then be seen as isomorphic with the mental sentences that constitute the mind. This paper pursues some of the consequences of that suggestion.

————————————————————

A DILEMMA FOR THE WEAK DEFLATIONIST ABOUT TRUTH

by Glen Hoffmann

The weak deflationist about truth is committed to two theses: one conceptual, the other ontological. On the conceptual thesis (what might be called a 'triviality thesis'), the content of the truth predicate is exhausted by its involvement in some version of the 'truth-schema'. On the ontological thesis, truth is a deflated property of truth bearers. In this paper, I focus on weak deflationism's ontological thesis, arguing that it generates an instability in its view of truth: the view threatens to collapse into either that of strong deflationism (i.e., truth is not a property) or that of some form of inflationism (i.e., truth is a substantial property). The instability objection to weak deflationism is sketched by way of a truth-property ascription dilemma, the two horns of which its proponent is at pains to circumvent.

SAMPLES AND LOGICAL FORM

M. G. Yoes

_____

Standard logic, the classical logic of truth-functions, quantifiers and identity, enjoys a measure of authority. It is not unchallenged, but by dint of its simplicity, power, and length of time in the curriculum its authority thrives. So examples of sentences whose logical form resists recasting in standard logic are often tantalizing. Who knows what revolution in logic lurks in an example?

This one has been folklore for several decades though the source is unknown to me:

(1)   Jones talked and he talked and he talked and he talked.

Let 'Jones talked' be an atomic sentence and represent it by 'A'. Why not let

(2)   A & A & A & A

do the work? And then

(1')   Jones talked and he talked and he talked and he talked and he talked

comes in for similar treatment as

(2')   A & A & A & A & A.

And so on. Obviously this is faulty, one hears, because (2) and (2') are logically equivalent to each other and each is logically equivalent to A. This would have standard logic saying not only that (1) and (2) are logically equivalent but also that (1) and (2) are logically equivalent to

(3)   Jones talked.

There is, then, a logical difference at a fundamental level to which standard logic is insensitive. So much for standard logic.

Logical form is deep structure. Those following Russell's principle that grammatical form is no sure guide to logical form might look below the surface in this case somewhat as follows. What appears to be a conjunction of singular sentences is actually no conjunction at all but at bottom a singular predication of some predicate of Jones: Jones was being loquacious (or some such)! With a single one-place predicate the form of (1) is represented by

(4)   Lj.

(3), then, being of the form Tj, is seen not to be logically equivalent to (4). Two atomic sentences each with different predicates are not logically equivalent. If (1) and (1') may be regarded as instances of the same predicate, they are logically equivalent; otherwise not.

Perhaps this situation indicates an underlying ambiguity. (1) and (1') say similar things. But do they say the same thing? Have it your way.

Whatever the predicate, one result is needed: both (1) and (1') should yield (3). He talked, but he was not loquacious. But he can't be loquacious, or whatever the predicate is, without talking. Jones was loquacious, so he talked. If (1) and (1') are at bottom singular sentences, then how are any implicants they may have established? For each such predicate, we could introduce a set of axioms governing it and designed to produce just the consequences wanted. $(\forall x)(Tx \rightarrow Lx)$ and Tj would give (4), for example. This would mean a new set of axioms for each new predicate. If one decides that the underlying predicate in (1) and (1') are different and that further sentences with more conjuncts are different, then more axioms would be needed. Alternatively, such implicates could be left to some sort of analytic implication. Is this the best that can be done? No doubt the problems of this example are merely illustrative. There must be many such sentences.

One important way a symbol may function is as a sample; it may exemplify something. On Goodman's account, a sample is a symbol of some property it has; it *exemplifies* that property. (See [1].) Of course a sample does not symbolize all its properties. The salesperson's sample stands for its color and texture, say, but not the property of having been carried around in this sample case all day or its specific gravity. Since sentences are symbols, they too can function as samples, as exemplifiers of some of their own properties. Such sentences as (1) give information, but some of that information is by way of its being a sample.

(1) makes a statement, but it also exemplifies a property of the scene. (1) does not merely *say* that Jones talked, but it also exemplifies how he talked and by doing so conveys information not conveyed in what is said. There is more to conveying, more to talking, than saying. Let us recognize, then, that there are two symbol systems here, that they overlap, and that they operate together on the same symbol. Logical form is one thing, conveying by exemplifying is another, not wholly distinct, thing.

The talked and talked case is just such a case. The illusion of nonlogical equivalence here is fostered by the fact that the sentences have different exemplifications! This also explains why the question of equivalence becomes intuitively uncertain as the number of clauses grows: two different such sentences each with many but a different number of conjuncts such as (2) and (2') can exemplify the same vague predicate. Intuitions may differ as the number of conjuncts increases. It is clear enough that these do not pose questions of *logical* equivalence.

Anything can be a symbol, and nothing is intrinsically one. That is, anything can function as a symbol, but nothing intrinsically so functions. At a minimum something functions as a symbol if it stands for something, refers to something, though of course this is at best a reasonable first approximation if not a truism. How a symbol functions depends on the symbol system of which it is a part and which is operative at the time. A sign reading «No Parking Here» may function as a command in one symbol system, a symbol for Mars in an impromptu explanation of the solar system, or as a sample of an imperative in an English lesson.

What a symbol system is, how it determines or contributes to the function of symbols within it, and how to tell whether a given symbol system is in play — these are questions which in one form or another have exercised philosophers. In recent years further important

questions have been raised. When does a symbol function as a sample? When does a symbol express something? When does it represent?

These and related questions lead, or perhaps should lead, to a symbolic turn off the linguistic turn. A fixation on language, natural or formal, carries with it a blindness to the way symbols function in general and to a lack of appreciation of language as one symbol system among many. A subtle effect of this fixation seems to be the assumption that given a symbol and its system in operation there can be no other symbol system overlapping it. A symbol plays in only one language game at a time. The simple case before us demonstrates otherwise.

What further consequence are there of this observation? Does it corrupt counterexample-making by making counterexamples radically uncertain? Is this a basis for explaining conversational implicature, a basis quite distinct from Grice's intentional account? Does it go any way toward clarifying the function of figurative symbols?

Reference

[1] Nelson Goodman, *Languages of Art*. Indianapolis: Hackett, 1976 (2nd edition).

————————————————————

M. G. Yoes
Professor of Philosophy Emeritus
University of Houston
Houston, Texas USA
<myoes@houston.rr.com>

MAY JUDGES SOMETIMES LIE?: REMARKS ON SORENSEN'S VIEWS OF VAGUENESS AND LAW

Jason Glenn

Philosophers concerned with the problem of vagueness agree on at least one thing: a predicate is vague if it breeds sorites paradoxes. A sorites paradox arises when a predicate (e.g., 'tall') allows for a range of borderline cases in which the truth-value of propositions containing the predicate is indeterminate. For example, there are cases in which 'tall' clearly applies (Michael Jordan) and cases in which it clearly does not (Danny Devito). When creeping along the spectrum between clearly 'tall' and clearly 'not tall', we must admit that there is no discernible difference in *tallness* (although there is in *height*) between someone who is 6' 10» and someone who is 6' 9». Likewise, we would not claim that 'tall' applies to someone who is 6' 9» and not to someone who is 6' 8». What this ultimately leads to, when the above type of inference is made successively enough times, is the claim that 'Danny Devito is tall,' which is clearly not true. Because vagueness in natural language seems to be an inevitability, it is not surprising that vagueness frequently finds its way into the law. An example of vagueness in law would be the notion of 'excessive bail'. The sorites paradox reveals itself in this case when one acknowledges that while it is not the case that $1 million is excessive bail and $999,999 is not excessive bail, it is also not the case that, after reasoning this way enough times, $1 is excessive bail.[1] It seems obvious that we cannot draw a distinct line between the zones of *legitimate* bail and *excessive* bail without being arbitrary.

Roy Sorensen argues that when judges make judgments about vague cases they are acting *insincerely*. Furthermore, Sorensen equates this insincerity with lying. It is Sorensen's view that because judges try to avoid, as much as possible, what Sorensen claims is 'lying', vagueness does not have a function in law. Sorensen's arguments, though indeed penetrative of the sometimes obscure role that vagueness plays in the operations of Western democratic legal systems, seem to be in need of some clarifying remarks. It is my purpose here to argue that not only may Sorensen's notion of 'lying' be too loose, given that judges would seem to be 'lying' *no matter what* judgment they might make in borderline cases, but also that Sorensen's epistemic view of vagueness, which holds that vagueness entails *ignorance* (ignorance being a phenomenon that cannot possibly, for Sorensen, be functional) may also need clarification. It seems that what view one takes towards vagueness and its occurrence in law affects the view one must take towards judicial discretion.

---

[1]. Sorensen, Roy. «Vagueness Has No Function in Law». *Legal Theory.* (Cambridge Univ. Press, Cambridge UK) Vol. 7. No. 4. December 2001. p. 394.

I should begin by describing precisely what view Sorensen holds towards vagueness in general. First of all, there are a number of views one could take. One could argue that borderline cases simply have no truth-value at all. Another could argue that borderline cases have partial truth-values, entailing that there are *degrees* of truth (which entails a departure from the notion of bivalence). Sorensen supports an *epistemic* view of vagueness, which holds that there exists a distinct line between what is, for instance, tall and not-tall; it is simply the case that we cannot know where this line of demarcation falls. This allows one to maintain faith in classical logic. The epistemicist grants the possibility that one can be mistaken when discussing borderline cases.

The first main claim in Sorensen's paper is that judges act insincerely when they make judgments about borderline cases. In making judgments about cases that are indeterminate, a judge is giving the appearance that she knows something about which no one can know anything. Such a judicial statement may be stipulative, but it certainly does not seem to constitute an expression of knowledge. The peculiar dilemma for a judge in this case is that she must say *something*, which seems to require, by moral imperative (perhaps), her not qualifying her statement by admitting that it is, in fact, *stipulative*.

If judges' decisions are, in the case of stipulation, never wrong, then, Sorensen admits, the problem of vagueness in law disappears because no such decision could be disputed in terms of truth-value. However, while admitting that judges do engage at times in stipulation, according to Sorensen, such «precisifications are regarded as fictions or technical uses. They are not regarded as acts which bring into existence an answer to the original question.»[2] Given that Sorensen supports the epistemic view, it is no wonder that Sorensen isn't satisfied with the view that verdicts may sometimes not be about discovering truth.

On behavioral grounds, Sorensen equates making statements about borderline cases with *lying*. Sorensen advocates using a definition of lying provided by Arnold Isenberg, who holds that a «lie is a statement made by one who does not believe it with the intention that someone else shall be led to believe it.»[3] Insincerity, Sorensen argues, results in internal feelings of guilt that can be verified externally (polygraph results being a case of such 'verification') and that are sufficiently similar to external indications of lying. Unfortunately, Sorensen's comments on such behavioral considerations are brief, leaving us to question how we should gauge 'sufficient' similarity. This is a classic metaphysical issue.

In any event, it seems prudent to explain Sorensen's distinction between *relative* and *absolute* borderline cases. A relative borderline case is one which may admit of a precise truth-value if certain «answering resources» (experts, precise measuring instruments, etc.) are available to clarify what is *prima facie* indeterminate.[4] An absolute borderline case, for Sorensen, is one that, although having a definite truth-value, will remain indeterminate forever. According to Sorensen, «only the possession of *absolute* borderline cases makes a term vague,» which entails that «(o)nly absolute borderline cases generate sorites paradoxes.»[5] It

---

[2]. Ibid., p. 390.

[3]. Ibid., p. 390.

[4]. Ibid., p. 392.

[5]. Ibid., p. 393-394.

seems fair enough to grant the uniqueness of the absolute borderline case that this definition allows, given that a relative borderline case is capable of, by referring to «answering resources», being clearly determined. Making the above distinction involves taking the kind of epistemicist stance (which Sorensen does) that logically explains away the prospect that certain borderline cases might ever become cases of the *relative* kind. That said, judges, when faced with adjudicating absolute borderline cases, are torn between two moral responsibilities: *decisiveness* and *sincerity*. According to Sorensen, «judges know that decisiveness takes precedence over sincerity … However, this moral certainty does not relieve the judge of a sense of guilt and regret.»[6] Because of this tension, according to Sorensen, judges seek to reduce the amount of absolute borderline cases with which they deal. This is one of the key reasons Sorensen gives for thinking that vagueness could not possibly be functional in law. We can assume that something is 'functional' if it is deliberately introduced and employed and not simply something that happens to exist in the law by virtue of its being, for example, an indelible 'defect' of natural language.

By referring to Joseph Raz's notion of 'creative indeterminacy', which is essentially the idea that it is pragmatically beneficial to introduce indeterminacies into law, leaving them to judicial discretion, Sorensen concedes that some borderline cases are functional in law, but it is simply those cases that are relative borderline cases that are so functional.[7] In other words, sometimes making legislation indeterminate saves lawmakers the trouble of having to make lists, for example, of what constitutes 'reckless driving', lists that undoubtedly will be inhumanly long (assuming it's even possible to finish such a list).[8] According to Sorensen, «(l)egislators deliberately create relative borderline cases to control the way a question is answered, not to make it entirely unanswerable.»[9] Although absolute borderline cases may sometimes play a role in law, as in legal loopholing, this ability of lawyers to throw a wrench into the legal machinery, according to Sorensen, does not constitute a *function,* one reason for this being shown by the fact that «custodians of the common good» usually try to combat such loopholing, at the very least by representing it to the public as being to some degree condemnable.[10]

It should seem clear that judges, when confronted with absolute borderline cases, simply have no choice but to make decisions. According to Sorensen, some philosophers of law have attempted to «exonerate the judges by diluting the notion of assertion.»[11] One way of doing this, attempted by A.D. Woozley, is to semantically water down the notion of assertion. Woozley claims that «(r)easonable, hardheaded lawyers can properly discuss the question (and disagree with each other in their answers to it) of what the right answer to a

---

[6]. Ibid., p. 391.

[7]. Raz, Joseph. «Legal Principles and the Limits of Law». *Ronald Dworkin and Contemporary Jurisprudence.* ed. Marshal Cohen. Totowa, NJ. Rowman & Allanheld, 1984. p. 83.

[8]. Endicott, Timothy. «Law is Necessarily Vague». *Legal Theory.* (Cambridge Univ. Press, Cambridge UK). Vol. 7. No. 4. December 2001. p. 382.

[9]. Sorensen. Ibid., p. 397.

[10]. Ibid., p. 399.

[11]. Ibid., p. 401.

question of law is, even though they agree that there is not a right answer to it — yet.»[12] What this implies is that we *can* make assertions without their having a truth-value. Sorensen finds fault with this view on the grounds that it would result in contradictions such as: «Patty Hearst robbed a bank voluntarily but it is neither true nor false that Patty Hearst robbed a bank voluntarily.»[13]

Another way in which one may attempt to dilute the notion of assertion is to claim that adjudication of absolute borderline cases relies upon the notion of verdicts being *performative*. Sorensen draws upon Austin's theory of performatives to make this view more clear. Austin claims that it is possible for *some* judicial acts to be 'exercitive', as opposed to 'verdictive', verdictive speech-acts «having obvious connexions with truth and falsity, soundness and unsoundness and fairness and unfairness.»[14] Exercitives (like 'appointing' or 'proclaiming'), which appear to be a matter of *dubbing*, don't seem to be constrained by evidence, and their existence in law simply points out that law involves there being a «double system of bookkeeping.»[15] That is, there is an official report of what occurs in a case (the exercitive) that exists alongside an understanding of what historically occurred, an understanding which instructs the exercitive. Even though there may be cases in which the judge does not know what actually occurred, she is still obliged to make a decision. Both strictly and simply adhering to a precautionary principle in this case is merely, according to Sorensen, a matter of shifting the location of indeterminacy and does not constitute an establishment of *truth*. Again, as mentioned above and according to Sorensen, predicaments like this are undesirable and should be avoided as much as possible. In addition, it seems that Sorensen's motivations for the above refutations might, to some degree, depend upon his inability to entertain the idea that sometimes the law may be simply stipulative and yet functional.

Concerns

Being an epistemicist about absolute borderline cases entails, for Sorensen, characterizing the dilemma posed by vagueness as a matter of *ignorance*. Again, for the epistemicist, there are truth-values of propositions containing vague predicates, but (alas!) we are ignorant of them. That said, Sorensen claims that ignorance, in the case of vagueness, cannot possibly be functional. But it seems as though Sorensen is assuming that truth-value lies 'out there'. This is another classic metaphysical issue. If there are certain predicates the truth-value of which we will *never* be able to ascertain, how does it benefit us to make assertions that they do have a truth-value at all? Another classic metaphysical dispute. To be ignorant of something requires that there be something (in this case, a truth-value) of which to be ignorant. It seems unclear, however, how one can assert that an absolute borderline case has a truth-value without at the same time *ever* being able to say what that truth-value is. Ignorance is ignorance *of*. Concerning absolute borderline cases, what is not clear is not simply what the object of the preposition in this case would be, but whether there is one. The problem I find with the epistemic view may be analogous to the problem of a color-blind man

---

[12].   Ibid., p. 402.

[13].   Ibid., p. 402.

[14].   Ibid., p. 402.

[15].   Ibid., p. 402-403.

who sees only shades of gray who asserts that an object he sees is a color other than gray, without being able to say what color it in fact is. The question here is: how does the color-blind man know that the object he's perceiving is not in fact *just gray*. In any case, if the truth-values of propositions constituting absolute borderline cases are *forever* undisclosed it seems at least pragmatically sensible to not make assertions about their existence at all.

Sorensen is what one might call a *strong* epistemicist. Some epistemicists claim that, although we may be currently ignorant of the truth-value of a vague proposition, it certainly may be the case that we can eventually gain access to such a truth-value. Sorensen claims that there are some truth-values that will never be discoverable, although it does not seem clear what criteria he's using to ascertain that the discovery of a certain truth-value is inconceivable. In any case, it seems that he's using a criterion based on what is *currently* inconceivable, which seems to leave him open to the same criticism he advances towards weaker epistemicists like James Cargile, who claim that although there may be no current legal thinkers who can discover (and not simply stipulate) the line that separates legitimate bail from excessive bail we have no way of knowing whether or not future legal thinkers may be capable of doing so. This uncertainty about whether or not future thinkers can discover the truth-values of vague propositions is a form of higher-order vagueness. Higher-order vagueness occurs when there exist borderline cases of borderline cases. The question that Sorensen asks Cargile is this: «Who are we to say what can and cannot be discovered?»[16] This seems, however, to be a question that Cargile could clearly (and equally) pose to Sorensen. In any event, Sorensen seems to give no clear argument for why one should be an epistemicist in the first place. It seems that if we were to take the view that borderline cases don't have truth-values then there would be nothing of which we can be ignorant when discussing them. This would entail that we could still grant, *with* Sorensen, that ignorance lacks a function, while granting that what absolute borderline cases breed is not ignorance in any strict sense.

That said, if we were to at least assume that the epistemic view of vagueness is the correct view (and that Sorensen's version of it is also the most viable version), it seems that there would still be some issues that Sorensen would need to clear up. Most of these issues directly concern Sorensen's claim that judges, when giving verdicts on absolute borderline cases are acting *insincerely*, and that this insincerity is a form of *lying*.

First of all, Sorensen's claim that judicial insincerity equates with lying, which is partially based on the argument that the behavioral clues exhibited when one makes a precise claim about something that is vague are the same behavioral clues discernible when one is lying, seems to be tenuous. Perhaps the feelings of 'guilt' a judge may have when giving a verdict on a vague case aren't the same feelings of guilt one has when lying. Perhaps the judge's feelings of guilt are based upon a recognition of the fact that the degree of her fallibility, in these cases, is stretched to an enormous (perhaps infinite) length (only assuming that the judge *is* an epistemicist like Sorensen). This doesn't seem to be the kind of internalization which may occur when one knows the facts but simply makes an assertion that is diatmetrically opposed to such facts. It seems plausible that the reason judges avoid absolute borderline cases may have less to do with avoiding 'lying' and more to do with the fact that judges wish to avoid having their decisions reversed (which would be, ultimately, a prudential concern, and perhaps not a moral one). Also, perhaps the type of insincerity that could be deemed sufficiently similar to lying occurs when one makes an assertion knowing that one

---

[16].   Ibid., p. 395.

doesn't know while still admitting that it's *possible* (however remotely) that one *could* know. This seems to be different from one's making an assertion when one not only knows that one does not know but also knows that *no one can* know, which seems to be the type of insincerity to which Sorensen is referring and perhaps should not be considered sufficiently similar to lying.

So where does this leave discretion? Certainly we can with little effort grant that judicial discretion is in place to allow for lawmakers to implement 'creative indeterminacy' (which in itself is in place for pragmatic reasons), and so serves a function. We could (perhaps) concede to Sorensen that absolute borderline cases do not serve a function in law, but it is at least not clearly the case that they fail to do so for the reasons that he gives: 1) because judges are trying to avoid 'lying', and 2) because judges are sometimes ignorant and ignorance has no place in law. That is, regarding this latter reason, it might be the case that ignorance has no place in law, but it is not clear that what Sorensen is talking about is ignorance. In any event, are judges lying or are they ignorant? Both? If they are lying, what are they lying about?

If it actually is the case that vagueness serves no function in law, perhaps to *say* that it doesn't is more commonplace than Sorensen might think. Perhaps to say that vagueness either does or does not serve a function in law could be, in some ways, akin to saying something like «Time either does or does not serve a function in law.» To refer to time in this manner seems rather strange given that time is just something that's *there*, a background feature if you will. Vagueness is a limitation of language, just as time is a limitation of experience. One could claim that lawmakers deliberately seek out vague expressions when they could choose more precise ones. It is not clear, however, that there are more precise ones, at least not without one's being arbitrary, which means not without one's being *stipulative*, which, in turn, is precisely (forgive me) what judges are there to do. Perhaps the spirit of why we *let* vagueness *be* involves our wanting to have individual persons do the stipulating individually, rather than writing the stipulations into the law. This is, presumably, one of the things that make judges special. Anyone could go by the book.

Lying seems to involve one's knowing that one is deceiving another, or at least one's *believing* that one is deceiving another. How can one know (or believe) that one is deceiving another if they themselves admit (assuming, again, that they are a strong epistemicist) that they cannot know the truth-value of the particular proposition being 'passed off'? If a judge happens to be wrong (which Sorensen, being the epistemicist that he is, allows for), it is not verifiable that he is wrong (not even to the judge himself).

Interestingly enough, it would seem that lying, in the case of vagueness, would have to involve specifically one's *not* being a strong epistemicist. That is, if a judge were to acknowledge that *no one can know* anything about which one makes statements, how could the judge be lying, since lying entails at least believing that one is making a statement that is incorrect? One can only assume that what Sorensen claims judges sometimes lie about is not the content of the proposition that admits of borderline cases («X constitutes reckless driving.»), but rather some internal state of their own such that they present themselves as being assured of the truth of something when in fact they can't be. This, it seems, would have to be the type of 'lying' that Sorensen would be talking about. Or perhaps we should just call it *bullshitting*, which is not the same as lying. Again, when one lies one deliberately states a negation of what they know (or think they know) to be the case. When one is bullshitting, the

only thing that one deceives another about is one's attitude towards the mode of presentation of one's claims about a certain subject matter, and not the subject matter itself.[17]

Of course, this is beside the point. It is not at all clear that what judges are doing is bullshitting, for it still remains that even judges faced with absolute borderline cases aren't so much saying «I know» but may very well be *performing* an act (not to be confused with *pretending*), an act that has moral *consequences* as well as moral *motivations.* If we were to consider a judge mistaken in this context, it seems that we either must mean that the judge is mistaken about incidentals *surrounding* the subject matter (and not about the subject matter itself, which is vague) or we must mean that the judge is *morally* mistaken, in which case we mean that the judge is not wrong about the truth-value of a certain proposition but is wrong in what he is *doing*. Sorensen seems bothered by the fact that sometimes judges may, in fact, just be stipulating. If vagueness is not functional in law, it is at least plausible that mere stipulation is. Occasional precisification in vague cases seems to be necessary, lest we halt the legal mechanism. Judges, after all, set precedents even if such precedents may be 'fictional' or 'technical'. The fact that they may not provide «an answer to the original question» may be because it is not always the case that there is an answer.[18] It's probably the case that judges, when making decisions on borderline cases, are *creating* law (which seems to be Hart's position).[19] This basically means that judges, when making judgments about borderline cases, are acting *only* stipulatively. Given our discussion here, this appears to be anathema to Sorensen. But, exactly what is wrong with this view seems unclear (even given the fact that Sorensen accepts the 'double bookkeeping' claim) and perhaps is in need of further elaboration by Sorensen.

———————————————————

Jason Glenn
Baton Rouge Community College
<cal_glenn@yahoo.com>

---

[17].   Frankfurt, Harry G. «On Bullshit». *The Importance of What We Care About.* Cambridge, UK. Cambridge Univ. Press. 2005. pp. 117-133. See this famous essay for Frankfurt's compelling distinction between lying and bullshit.

[18].   See pages 3-4 of this essay.

[19].   Hart, H.L.A. *The Concept of Law.* Oxford, UK. Oxford University Press, 1997. 2nd edition, paperback.

THE CONTEXTUAL NATURE OF COGNITION AND DANCY'S MORAL PARTICULARISM
by Mark Lovas

——————————————————

… a moral principle may draw attention to a feature that is always relevant, and relevant in the same way, under certain implicit conditions… We don't normally spell out those conditions … But the list of conditions is not open-ended, and it is knowable in advance. (D. McNaughton and P. Rawling, 2000, p. 269)

Principlism, or … the *Leibnizian theory of inquiry, has been regarded, and is in some circles still regarded, as a live option. Thus the context problem has long remained invisible, and is still hard to grasp from certain standpoints. (Andler, 2003, p. 364)

——————————————————

Introduction

Oh how post-Socialist countries have problems! I recently read about a Czech university student who returned home to find her apartment empty. She hadn't exactly been robbed. A court had ordered that a different woman's possessions be re-possessed; but, unfortunately, due to a bureaucratic mix-up the re-possessors went to the wrong address. The officials responsible for the mix-up were in no hurry to admit their mistake; they insisted that the student produce sales' receipts or reliable witnesses who could testify that the books, CD's and furniture really belonged to her. The whole affair was so upsetting that the student had to postpone graduation, but today she vows that she will save all of her receipts.[1]

One might wonder whether saving literally every sales receipt is a realistic policy. It is, perhaps, every bit as realistic as wondering about the epistemic credentials of each of one's beliefs, or habitually doubting one's first thought about what to do next. Skeptics can make us insecure about our evidence, and in the realm of morality, Dancy's particularist arguments for the holism of reasons could make us doubt our thoughts about what to do next. Dancy claims that any given reason which is, in a given context, moral justification for an action, might lose that weight or valence in a different context. A reason might count for an action in one context, but against it in a different context. The skeptical thought is: what if the context changed without my noticing it? Or, what if I failed to notice a crucial aspect of the

---

[1]. The editor points out that proof of ownership was already required in Roman times. There are, however, important differences between Ancient Rome and Central Europe of today. Friends of capitalism are fond of pointing out how industrialization brings us a much larger number of material possessions, more even than were previously available to the rich. Additionally, in the implicit background of my little anecdote lurks an especially detestable character: The Central European Bureaucrat, a person who begins with the assumption that you are guilty and need to prove your innocence, and expresses her superiority via a variety of gestures all with the subtle, indirect, but clear message that you are a disgusting inferior and that anything she might happen to do for you is a favor granted by a higher being to a lesser entity. The most perfect artistic expression of this cultural phenomenon is found in the writings of Kafka.

situation? If I am about to perform an action because I think it's just, should I scrutinize the situation to ensure that there are no hidden defeaters able to rob my reason of its usual weight?

Dancy suggests that there may be times when cruelty, rather than kindness is called for, and times when we should lie rather than tell the truth. There may even be cases where the just thing is the wrong thing to do on account of its being just.[2] More generally, so-called thick moral concepts such as «integrity, fidelity, gratitude, reparation, and so on..» (Dancy 2004, 121). can't be expected to provide reasons in every context-come-what-may, and we cannot suppose that a good reason will always be good in no-matter-what context.

Dancy is no skeptic (see, e.g., Dancy 2004, p. 1) and it would be a misunderstanding of his views if we went away doubting every reason that occurred to us, but his arguments should at least make us curious. How do we do it? If there's nothing necessary about a given reason's supporting an action, how do we recognize that it does support the action, when it does? How do we manage to change our minds when contexts change? If we have moral knowledge, shouldn't we be able to give some account of how we manage to employ the reasons which, in a particular context, are good ones?

Dancy has surprisingly little to say on these matters. In *Moral Reasons* (Dancy 1993, Chapter 7, Section 2) he appeals to the notion of salience. The agent doing the right thing perceives the situation as having certain saliencies. She can explain her choice by giving the right sort of description of the situation, one which makes salient what should be salient. She tells a revealing sort of story. In his most recent book-length treatment, (Dancy 2004, pp. 160-161) he tells us that our ability to grasp reasons is primitive. Without rejecting the holism of reasons, I think that it may yet be possible to say more, and I shall attempt to do so below.

In broad outline, I would like to borrow a suggestion made by Daniel Andler. (Andler 2003) We should not think about context as if it were another thing, as if it were to be dealt with by adding another variable into a principle, or as if it could be the object of a special sensitivity. What we call adjusting to the context is just the creativity and sensitivity which a person shows when rules-of-thumb or principles don't work. This insight allows us to deal with a peculiar response which Dancy's examples can provoke. They can make us wonder how it can be possible that we track good reasons. The insight here is that what we are keeping track of is, simply, the basic moral features of our lives.

The interpretation I wish to suggest is morally robust and natural. It surfaces when Dancy writes:

> … moral reasons are just ordinary considerations such as his distress or the loss to her self-image.
> (Dancy, 1993, p. 115)

The particularist thinks that there are times when such ordinary considerations should be given great weight, and cases where they might be out-weighed. The ability to know when to do

---

[2]. To the best of my knowledge, Dancy has not strenuously argued that justice can sometimes count against doing an action. He tends to argue in a very schematic way, saying the particularist must say this—on the basis of the general position. There may be something odd about that. In one place where there is a bit more substance (Dancy 2001), the claim is that justice sometimes might not be the strongest reason in favor, and a very particular theory of the nature of justice seems to be assumed—roughly one where property rights define justice. Arguably it is in virtue of that particular view that justice loses its weight. However, it isn't Dancy's own view that's being discussed in the relevant passage.

what is a matter of perceiving such reasons and being aware of any competitors. To speak of «context» here is a way of describing that very perception of the weight of his distress or the damage to her self-image. There is not some additional calculation of the weight-in-this-context; to perceive it as mattering is to already to have done any needed calculation.

I also propose to borrow from Kent Bach's discussions of reliabilism and default reasoning. (Bach 1984, 1985, 2005) Insofar as moral agents know when to think twice, and know when to simply trust their first thought about what to do next, their actions will tend to be morally right. This also enables us to accomodate an insight of Andler's. Andler (2003, p. 360) notices that one odd feature of a principlist view of inquiry is that when inquirers are creative, they seem to be «shooting in the dark». Bach (1984) used the pejorative phrase «jumping to conclusions» as part of his title in a paper about default reasoning, but he sketched a view which deprived the phrase of its pejorative tint. The general point is that we bend and break rules when we see fit, and that's not a failing. That's just what we do, and criticism of the process takes place when we implicitly compare it to what we don't do, invoking an ideal of inquiry which no inquirer instantiates. In the end, serious questions may remain about how we do it, but framing Dancy's particularism in this broader context should lead to a truer evaluation of its strengths. Dancy's particularism is a sustained attempt to recognize the contextual nature of moral cognition.

The Problem of Context

> … situations are not always unambiguously of a single type: there are emergencies in restaurants, and meals in emergency wards. (Andler, 2003, p. 368)

In a recent article, (Andler 2003), Daniel Andler provides the following characterization of the «contextual nature of cognitive processes»:

> What happens in a given, particular situation rarely only depends on the type of the situation: there is something about it which is not exhausted by its being a *token* of the type. (Andler, 2003, p. 352)

Suppose we have a principle or law which tells us that things of such-and-such a type have certain properties. Andler is suggesting that even after we've classified an object, we are, so to speak, not done. I may have blamelessly and correctly classified an object or situation as belonging to a certain category, intending thereby to understand it through the laws governing such entities, but my work is not over. There are other properties of the object or entity which I've not recognized, and those properties are not insignificant. By the same token, in the case of action, I may recognize my situation as being of a certain type, and thereby imagine a certain course of action is the right one. However, my classification does not exhaust the situation's possibilities, and there may be a «defeater» existing in the world outside my classification. It is illuminating to consider Andler's own further expansion on this remark:

> This is a problem because, or rather, to the extent that it does not yield to the straightforward cure, viz., a redesigning of the types. Situations seem to resist classifications, somewhat like persons, or again, some will think, like illnesses of living organisms….

(Andler 2003, p. 352)

Andler recognizes two broad reactions to the general problem of context. The deflationist thinks context is not a big problem, and is committed to using principles. The paradigmatic deflationist will be a *Leibnizian — with the asterisk indicating that Andler is not

engaged in a historical inquiry. The *Leibnizian thinks the principles can be modified, or tweaked just enough to capture the context. The key point is that inquiry is, above all, a matter of following principles. The other reaction is that of the inflationist. The paradigmatic inflationist turns out to be a kind of sceptic — not about our knowledge of the real world, but about our ability to describe how we do it. (One example mentioned by Andler is Fodor 2000)

Andler allows that *Leibizian methods can take us far, but they have inherent limitations. There is, on Andler's account, a deep incoherence in the *Leibnizian account because all situations represent a sort of information, and the view has not got the resources to mark a principled distinction between text and context. (Andler, 2003, pp. 359-60). This point is echoed in Dancy's complaint that critics ignore his distinction between *enablers* and *reasons*. (See, especially, his complaint about Crisp in this regard in Dancy 2004, pp. 96-7) For Dancy, my decision to keep a promise does not include as a reason the fact that my promising was un-coerced. On the contrary, the fact that I was not coerced is not a reason at all. By contrast, McNaughton and Rawling (2000, pp. 271-2) claim that the fact that I was not coerced is among my reasons for keeping a promise.[3] In making this move, McNaughton and Rawling are feeling the attraction of a *Leibnizian view.

The second problem for the *Leibnizian is novelty. No investigator simply gives up because his explicit principles won't capture a particular case. (Andler 2003, pp.364, 370) What people actually do is make a guess, or use their best judgment or insight. (369-70) And, that is where Andler, who is proposing an epistemological view, wants to get some support from recent discussions of moral particularism. Andler claims that recent discussions of moral particularism show that principles alone can't be the whole story:

> …on the epistemic side, the need for principles is not seriously in doubt, and the question is whether they can be consistently supplemented by another resource;while on the ethical side, at least in recent discussions, doubts are on the side of principles, while context-linked abilities have seemed pretty secure. (Andler 2003, p. 363)

Andler does not discuss the ethics literature in detail, but his general observation seems correct, so far as it goes. Even Dancy's critics admit a need for «context-linked abilities», however, I am suggesting that even within the literature which Andler refers to, *Leibnizianism continues to exert an influence.

Dancy and Context

Jonathan Dancy can be classed as a context inflationist. Dancy's argument for the holism of reasons — the claim that a supporting reason in one situation need not retain its supporting value, and might even switch value in a different situation — evinces a lively sensitivity to the role of context. Consider the following example:

> I borrow a book from you, and then discover that you have stolen it from the library. Normally the fact that I have borrowed the book from you would be a reason to return it to you, but in the

---

[3]. McNaughton and Rawling (2000, p. 271-2) claim that Dancy has confused what it would be odd to say with what is false. If we consider a comparison with discussions of perception where the problem is to specify what is intuitively the right object being seen, McNaughton and Rawling's view seems less plausible. Suppose that if I am to see my computer, some disjunctive set of neurons must be activated. It seems not merely odd to say that when I see, I see those neurons which are firing. The parallel is that in a given moral situation I perceive certain aspects of the situation. I do not perceive the causal background which makes that situation possible. (Cf. Recanati 2002, pp. 109-110 footnote and references therein.)

situation it is not. It isn't that I have some reason to return it to you and more reason to put it back in the library. I have no reason at all to return it to you. (Dancy 1993, p. 60)

We have two situations where you've loaned me a book. So far as that classification goes, they are both tokens of the same type. But, there is a moral difference between them which brings out the difficulty in attempting to derive an action-guiding rule applying to all situations in which someone has loaned me a book. And, that is Dancy's general point: due to the holism of reasons, we can't expand particular reasons into principles; so, principles have a very limited role to play within ethics. (For an account of why principles seem important, and the limited role which Dancy recognizes for them, see Dancy 1993, 66-71)

Someone might say that the two situations are *not* equally of the same type. One situation involves a masquerade. The friend who «loans» me the book is actually impersonating someone who has rightfully acquired a book: you cannot really loan something which you do not own in the first place. In raising this point, however, we are making Dancy's point. There is a moral difference between the agent who loans what he owns and the agent who loans what s/he stole, but the attempt to describe the situation, as it were, in advance of complete (or fuller) knowledge runs into the danger that we have missed the very facts that are morally relevant.

One might well ask how we are sensitive to the context. After all, since the weight of reasons depends upon context, insofar as we are aware of reasons, must we not be (indirectly at least) sensitive to the context? It may well appear that Dancy simply hasn't got an answer, even if he is no skeptic.

> … our account of the person on whom we can rely to make sound moral judgements is not very long. Such a person is someone who gets it right case by case. To be so consistently successful, we need to have a broad range of sensitivities, so that no relevant feature escapes us, and we do not mistake its relevance either. But that is all there is to say on the matter. (Dancy 1993, p. 64)[4]

Dancy returns to this general issue briefly in *Ethics Without Principles*. Part of his particularism is a denial that we can give anything like a general set of criteria which determine whether something is a reason: «… particularists tend to deny the existence of any substantial general theory of what it is to be a reason.» (2004, p. 160)

Dancy makes an important distinction between reasons and «enablers». Consider the question: Why should I keep a promise? Dancy suggests that I do what I promised because I promised. It's true that if I had been coerced into promising, my promise would not be binding, but that's not my reason: it's an *enabler* for my reason. (Dancy 2004, pp. 38-9) Similarly, in the example above, that s/he didn't steal it was an enabler for my friend's loaning me the book. Sometimes Dancy makes the claim that there are «too many» such potential defeaters for them to be reasons. (e.g., Dancy 1993, p. 53) To repeat what was

---

[4]. Cf. Dancy 1993, p. 50 where Dancy says that the virtuous agent brings to each new situation «a contentless ability to discern what matters where it matters» (p. 50) This is surely an over-statement; not all content is linguistic. I suggest that the most that Dancy could say would involve using Sperber and Wilson's notion of a «code» model of communication. We cannot encode our thoughts about moral reasons. If, as Carston (2002, Chapter 5) suggests, psychological concepts are especially prone to shifts of meaning to fit specific contextual needs, and if psychological concepts are essential to moral reasoning (as I suggested above), then we may have an independent path to particularism. The very concepts we operate with get modulated in specific contexts so that there is no guarantee that they retain their exact content across contexts. Whether the path is independent, and how exactly it is to be described is something I plan to discuss in future work.

mentioned above, this shows Dancy employing a distinction much like that between the «text» and the «context» — something which, according to Andler, the *Leibnizian cannot find room for. Dancy has consistently claimed we don't have to think about them, or know such background conditions in order to act morally, or to know what we are doing. (Trenchantly expressed in Dancy 2004, p. 142.)

Should we be satisfied by what Dancy has to say? Once we realize that any given reason can lose its weight, we've lost the ability to think of ourselves as following action-guiding principles. At best, a principle works in a restricted area, but there's no hope of spelling out what the relevant restrictions are. Commentators have wanted to argue that not all reasons are like that, but their assertion looks like a bold-faced and stubborn re-assertion of what Andler characterizes as *Leibnizianism. Dancy allows that there may be a few reasons with constant valence — justice may be one — but, in general, something more than following principles is needed.

However, a puzzle does arise once we notice that no reason need stand firm as circumstances change. How are we to know if we are in a situation where a reason stands firm or not? Somehow I notice the reason, and somehow the enablers are in place, and no defeaters are present. But, doesn't it begin to seem like a miracle?

One wants to say that context is by definition (to use a notorious phrase) *background*. It's something like the lights in a theater or the camera angle in a movie. It has an effect upon us which enhances meaning, but it's not generally part of what we are interested in. Yet, without the right lighting the performance loses something, and camera angle can contribute much to a film's content. The effect of Dancy's examples is to *foreground* context, and thereby highlight the actual subtlety of our abilities.

Perhaps I can make my point in another way. Typically, Dancy's examples have a clear basic format. We imagine one example: In a given situation, S1, a reason, R1, is a reason to perform an action, A1. Then, we imagine a slightly different situation, where R1 is also present, but R1 does not support A1. A1 is not right in the second situation. It's right to lie in a game, but not in real life. In general, one should avoid cruelty, but if someone is torturing you, perhaps you should refrain from being considerate. (Dancy, 2001) In all of these examples, Dancy's point is not to raise a sceptical «what if?» question, but to emphasize the subtlety of our sensitivities. When commentators react by stressing that some virtues (honesty) or some features (justice) always count for an action, they seem to fail to notice that Dancy is stressing the subtle nature of our very real capabilities. He is, in effect, suggesting that in cases where justice (or honesty) is not what a situation calls for, we are capable of recognizing it. When Dancy describes such examples, he is presupposing that morally sensitive individuals are able to notice the difference.

> Competent moral judges do not need to be aware of everything that just might make a difference in order to determine whether it does or does not… (2004, p. 142)

One might persist in asking, «But, how do we do it?» When Dancy endorses the possibility that our ability to recognize reasons is simply a primitive faculty, he rejects the question:

> … our sense that something is a reason is *absolutely underivative and immediate, and can be given no independent theoretical support*. (2004, pp. 160-161; added emphasis)

It seems to me that Dancy's view is, at this point, rather hasty. On the one hand, Dancy rightly does not wish to offer an analysis of, say, «reasons» or «good reasons» in terms of a set of necessary and sufficient conditions. On the other, he is quick — possibly, too quick — to employ the weapon of particularism.

This is clearest in his response to the challenge that particularism leads to a kind of scepticism because there are simply too many defeaters and enablers. The objector says that one can't ever know enough. Notice how the context is expanding. What we'd like to keep in the background has now occupied the foreground, and our knowledge, it seems has disappeared.

Dancy's solution is to look for a sort of «filter» — a principled way to limit the epistemic demands upon a moral agent. Ah, but notice; I have lapsed into speaking with the vulgar: what can it mean for a *particularist* to seek a «principled» limit? The particularist, it seems, must say — and Dancy does say — that there's no drawing a line:

> …as a particularist I am happy to maintain that there need be no general answer. (2004, p. 159)
> Nonetheless, Dancy, apparently, senses that this is not an adequate answer, and does say more:

> … all and only aspects that the agent is capable of recognizing and/or at fault for not recognizing or for not being able to recognize can pass the filter, and so affect the morality of the action. (2004, p. 159)

Without troubling ourselves too much just now over the exact content of this subtle disjunction of ingredients — whose complexity Dancy himself implicitly acknowledges when he later remarks upon «tortuous» epistemological problems (p. 159, 2004) — I would like to suggest that Dancy has gotten himself focused upon the morality of the *agent,* and *not* the morality of the action. A well-meaning, sincere, good agent might do moral harm to others. (Cf. Honderich 1996) But, then, perhaps, here I assume a consequentialism which Dancy would reject.

Andler's Solution

It is enlightening to consider how Andler deals with the problem of the expanding context. Andler shares with Dancy the recognition that the potential search space can expand to such an extent as to overwhelm our resources. Thus, the problem is to limit it — as Dancy says to find a «epistemic filter». We are able to solve problems insofar as the space of possibilites has been, in some way, cut down. If Dancy asserts (and his opponents deny) that the background conditions — the enablers — are too large a class to be an object of our apprehension, to be grasped as a reason, and then asserts our grasp of reasons is «primitive», how is that any better than saying that our ability to recognize good reasons and act morally is a miracle? Andler draws an analogy with treasure hunt games, textbook solutions, Relevance Theory's approach to communication (Sperber and Wilson 1995) and recent suggestions by Dan Sperber and others that many of our daily survival problems are solved for us by the software which comes as standard equipment in our brain — the thesis of the «massive modularity» of the mind. (Defended, e.g., in Sperber 2001) Common to this varied list is the way in which the search space is limited. A textbook problem is solvable because the relevant chapter has given the student the materials needed to solve it. So, too, a treasure hunt has clues. The apparent infinity of the search space is whittled down. In the case of Relevance Theory, a communicator exploits aspects of the environment which are mutually salient to the audience. Modules are psychological features which are supposed to allow us

to solve problems faced by our evolutionary ancestors: they are narrowly focused upon specific problems, and for that reason do not have a problem with context.

In the case of ethics, I would make a proposal not explicitly found in Dancy's writings — though it would involve resurrecting and modifying the notion of salience which appeared in *Moral Reasons*. The search in ethical space is, I suggest, limited in somewhat the same way as the search in communicative space — via mutual saliencies.[5] Recall Dancy's examples mentioned above: the damage to her reputation or his distress. Ethical reasoning, and the communication of ethical reasons is possible because, in general, we have such sensitivities. (The need for a qualification is suggested by research concerning psychopaths, who lack sensitivity to the suffering of others. (See Nichols 2002)

The trick is not to conceive of the moral agent's problem as knowing when a reason has been disabled. We don't perceive enablers and disablers. We directly perceive reasons. This follows if we take seriously Andler's suggestion that context is not a thing which can be the object of a sensitivity, not something we sense with a context-sensor. Context is not another thing, over and above our reasons for acting and the actions we perform. When Dancy takes enablers and disablers as though they were possible objects of knowledge, (2004 p. 158) he has taken context to be a thing. Consider Andler's words:

> We cannot assume, on pain of circularity, a separate «situation-sensor»: we are precisely asking how the inquirer is able to focus upon the potentially relevant features and facts…

(Andler, 2003, p. 368)

Conclusions

If it is theoretically misguided to try to describe a «situation-sensor», then what looks like the absence of an epistemology in Dancy's discussions of moral particularism may, in part be, only evidence of the rejection of a *Leibinizian view of inquiry. If we conceive of context as a thing and try to adjust our principles to capture it, then we wind up with the sort of view expressed by McNaughton and Rawling in the quotation on the first page.

The deflationist thinks that by complicating principles, s/he can tame context. The problem with that solution is that it misses the fact that there's a sort of insight and creativity which we call upon when we notice that a principle doesn't apply, or that a principle needs to be tweaked. (See Andler p. 360) If context is not a thing, then what, we might well ask, is it?

> contexts do not [exist], any more than the beauty of a piece of art, or the elegance of a parting gesture, or the nobleness of a sacrifice. Context is not to be discovered, but appreciated and debated, setting an inquiry in the proper context does not imply the identification of the context: it means getting about the job in a 'good', skillful, laudable way: it means, in a word, proceeding intelligently. (Andler, 2003, p. 368)

Dancy's moral particularism draws our attention to our ability to proceed creatively and intelligently in a moral situation. His critics have missed this, and underestimated the

---

5.   Levinson (1995) argues that communication requires mutual salience: «…a notion of 'natural salience', such that I can be sure for an indefinite range of phenomena or scenarios that what is salient for me is salient or you.» (1995, p. 230). The notion has it basis in research by Schelling discussed in his *Strategies of Conflict*. (Schelling 1960) Levinson remarks that both Schelling himself and commentators agreed that in order to explain the remarkable ability of people to coordinate their actions without explicit communication—e.g., to agree upon a meeting place or a specific number—«roughly nine times out of ten», it was necessary to appeal to a notion of mutual salience.

depth of his thought. Dancy's silence about epistemology is largely motivated by a rejection of an implausible view of inquiry which Andler terms *Leibnizianism. However, Dancy's silence on these issues may have other causes as well. Above I suggested that we can solve the epistemological problems raised by moral reasons if we suppose those reasons are apprehensions of the effects our actions have upon other human beings. Moreover, moral communication — the expression of and understanding of a moral agent's reasons — is possible because we share mutual saliencies. These suggestions, once developed, may well lead to a full-blown form of consequentialism. Dancy (1993) has argued that consequentialism fails as an account of moral reasons. In order to fully defend my suggestion, I would both need to provide more flesh on the bones that I've so far offered, and to deal with Dancy's objections to consequentialism.[6] Moreover, there is need to examine the proposal that among the modules of the mind is one for reasoning about the mental states of others. Do the concrete proposals which are available support my project? That is partly an empirical question, but also a question about the current state of theoretical development. I plan to carry out that project in a sequel.

Acknowledgments

References

Andler, Daniel, 2003, «Context: the case for a principled epistemic particularism», *Journal of Pragmatics* 35, pp. 349-371.

Bach, Kent, 2005, «The Emperor's New 'Knows'», to appear in *Contextualism in Philosophy: On Epistemology, Language and Truth*, Gerhard Preyer and Georg Peter (eds.), Oxford UP, available at Bach's homepage.

Bach, Kent, 1985, «A Rationale for Reliabilism», *The Monist* 62: 248-63

Bach, Kent, 1984, «Default Reasoning: Jumping to Conclusions and Knowing When to Think Twice», *Pacific Philosophical Quarterly* 65, pp. 37-58

Carston, Robyn, 2002, *Thoughts and Utterances: The Pragmatics of Explicit Communication,* Oxford: Blackwell.

Crisp, Roger, 2000, «Particularizing Particularism», in Brad Hooker and Margaret Olivia Little, eds., *Moral Particularism*, New York: Oxford UP.

---

[6]. Insofar as my speculation here is based on Dancy's examples involving harm to other individuals, it might be that a form of «sentimentialism» is required—minimally, a theory giving prominence to our capacity to feel pain at the thought of causing pain to others. One such recent theory is found in Nichols 2004. (This reference may seem surprising to anyone who has read that work, as Nichols dismisses Dancy's particularism very swiftly in a footnote (Nichols, 2004, p18, fn. 6). On the other hand, Nichols never considers the sort of arguments I discuss above.)

Dancy, Jonathan, 2004, *Ethics Without Principles*, Oxford: Clarendon Press,

Dancy, Jonathan, 1993, *Moral Reasons*, Cambridge., Mass: Blackwell.

Dancy, Jonathan, 2001, «Moral Particularism», Stanford Encyclopedia of Philosophy

Fodor, Jerry, 2000, *The Mind Doesn't Work That Way*, Cambridge, Mass: MIT University Press.

Honderich, Ted, 1996, «Consequentialism, Moralities of Concern, and Selfishness», available at Honderich's homepage: <http://www.ucl.ac.uk/~uctytho/ted9.htm>; originally published in *Philosophy*.

Hooker, Brad, and Little, Margaret, 2000, *Moral Particularism*, Oxford UP: Oxford.

Kirchin, Simon, 2003, «Particularism, Generalism and the Counting Argument», *European Journal of Philosophy*, 54-71.

Levinson, Stephen C., 1995, «Interactional Biases in Human Thinking», in *Social Intelligence and Interaction; Expressions and Implications of the Social Bias in Human Intelligence*, ed. Esther N. Good, Cambridge UP, 221-260.

McNaughton, David, and Rawling, Piers, 2000, «Unprincipled Ethics», in Hooker and Little, eds., pp. 256-275

Nichols, Shaun, 2004, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, Oxford UP: Oxford.

Nichols, Shaun, 2002, «Is it Irrational to Be Amoral?» *The Monist* 85, 285-304.

Recanati, François, 2002, «Does Linguistic Communication Rest on Inference?», *Mind and Language*, Vol. 17, 105-126.

Schelling, Thomas, 1960, *The Strategy of Conflict*, Harvard UP: Cambridge, Mass.

Sperber, Dan, and Wilson, Deirdre, 1995 *Relevance: Communication and Cognition*, Second Edition, Cambridge, MA: Blackwell.

Sperber, Dan, 2001, «In Defense of Massive Modularity», in Dupoux, E., (ed.), *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, Cambridge, Mass: MIT Press. Also available at Sperber's homepage: <http://www.dan.sperber.com>

Mark Lovas, Ph.D.
Bratislava, Slovakia
[formerly Slovak Academy of Sciences, Philosophical Institute]
<mlovas@hotmail.com>

KUHN, METAPHYSICAL REALISM, AND REDUPLICATION

Andrew M. Bailey

———————————————————

One orthodox interpretation of Thomas Kuhn's *The Structure of Scientific Revolutions* charges him with two follies. First, Kuhn's paradigms are supposed to have a constructive or constitutive function.[1] They do not merely reflect reality — they *are* reality. Second, Kuhn is said to disregard truth for something lesser (such as coherence). At the least, he is said here to have abandoned any common sense notion of correspondence. Consider Franklin's radical and well-phrased claim that, «The worst effect of Kuhn, and the one taken up both most unthinkingly and most forcefully across the whole range of disciplines he influenced, has been the frivolous discarding of the way things are as a constraint on theory about the way things are.»[2] The conjunction of these two prongs of attack amounts to a charge of creative anti-realism.[3] Kuhn's anti-realism is supposed to be more than a merely epistemic thesis (such as scientific anti-realism), furthermore — it is a metaphysical view about what is.

The alleged vices of creative anti-realism are well-known and oft-recited; I will not catalogue them here, merely noting that the view tends towards self-stultification. Instead, I argue that *Structure* does not commit Kuhn to such a view. I will contribute to this end this by offering an analysis of one troubling passage. It is not unique; many analogues can be found scattered throughout *Structure*. In providing an analysis of the one, then, I hope to shed light on the many.

The passage in question ends Kuhn's initial discussion of the nature of scientific theories and their relation to the paradigmatic commitments of relating scientific communities:

> An investigator who hoped to learn something about what scientists took the atomic theory to be asked a distinguished physicist and an eminent chemist whether a single atom of helium was or was not a molecule. Both answered without hesitation, but their answers were not the same. For the chemist the atom of helium was a molecule… for the physicist, on the other hand, the helium atom

———————————————————

[1]. See, for example, the (not uncommon) claim that, «Kuhn suggests that in one sense one's paradigms are partially constitutive of one's world. It isn't just that with different paradigms we see and mean differently (although he means at least that), but that in some sense which he has difficulty defining, the *worlds* to which are paradigms address themselves differ.» Del Ratzsch, *Philosophy of Science,* (Downers Grove, IL: InterVarsity Press, 1986), p.51.

[2]. James Franklin, «Thomas Kuhn's Irrationalism,» *The New Criterion* Vol. 18, No. 10, June 2000.

[3]. «[creative anti-realism is the thesis] that we are actually responsible for the basic lineaments, the fundamental structure and *framework of the world itself* [emphasis added].» Alvin Plantinga, in *The Analytic Theist,* ed. James F. Sennet. (Grand Rapids/Cambridge: Wm B. Eerdmans Publishing Co, 1998), p.330.

was not a molecule. Presumably both men were talking of the same particle, but they were viewing it through their own research training and practice.[4]

Kuhn may hold there to be exactly one object in question («the same particle»); he has thus thrown the realists a bone. But what are we to make of the other language he uses? I take his «For the chemist…» locution to mean more than merely «according to the chemist.» That is, he is not merely pointing out a place of disagreement, with one right and one wrong party. Instead, he points to a case of *difference in paradigmatic commitments*.[5] The question then arises: is Kuhn suggesting that an object of scientific study can actually have feature and not have it? Can one and the same object be a molecule and not be a molecule at the same time and in the same sense?

A tempting solution lies in the opposite side of side of the spectrum; to maintain that more than one object is in question — that the chemist and the physicist really are looking at two non-identical things. On this reading, Kuhn's famous claim that that scientists of different paradigms «study different worlds altogether» might be taken at face value. But this is to embrace a problematic notion of creative anti-realism, given what Kuhn has already claimed about the nature of scientific paradigms. Neither of these solutions, then, seems promising. The dilemma's two horns can be formally explicated as follows:

i. $\exists x(Fx \,\&\, \neg Fx)$ «There exists some object which both has and does not have feature F»

ii. $\exists x(Fx) \,\&\, \exists y((y=x) \,\&\, (\neg Fy))$ «There exists some object which has feature F and some distinct object which does not have F.»

(i) is a formal contradiction; if Kuhn is committed to this, his position seems reduced to incoherent babbling, hardly a promising prospect. (ii), in conjunction with other remarks by Kuhn entails a denial of realism (in some important sense), a first step towards a global creative anti-realism. Can a third reading be found to avoid these pitfalls?

One key to unraveling this potentially incoherent tapestry is in the locution «for S…»[6] I suggest that Kuhn's language hints at a solution involving reduplicative propositions of the form x as c is F. Armed with this conceptual resource, one can maintain that an object 'as *c*' has a feature while the exact same object 'as *d*' does not.[7] We can see how Kuhn might maintain this more precisely, with Fx^c signifying a reduplicative proposition attributing a feature F to a subject x, not as such, but *as something c*, by positing the following:

---

[4].   Thomas Kuhn, *The Structure of Scientific Revolutions*, 3rd Ed., (Chicago/London: University of Chicago Press, 1996), 50-51.

[5].   As others have noted, there are at least two important senses of «paradigm»: «(1) an example to be followed (in style and technique), or (2) a core set of assumptions (embedded in the example) to be used to ground research.» Mark R. Brawley, *Hegemonic Leadership: Is the Concept Still Useful?*, 19 Conn. J. Int'l L. 345, Spring 2004. In this paper, I focus strictly on the latter. For critical analysis of Kuhn's various senses of «paradigm,» see Dudley Shapere, «The Structure of Scientific Revolutions,» *Philosophical Review* 73, 1964, 383-94 and Gerd Buchdahl, «Revolution in the Historiography of Science,» *History of Science* 4. 1965, 55-69.

[6].   That is, «For the chemist…» or «For the physicist…»

[7].   Here I develop in more detail a passing remark in Israel Scheffler, *Science and Subjectivity* (New York: Bobbs-Merrill, 1967), 45-66. While Scheffler's approach is critical of Kuhn; I intend to show how a reduplicative solution can cohere with the text and with realism.

iii. $\exists x(Fx^c)$ & $\neg(Fx^c)$ «There exists some object which *as c* has feature F and *as d* does not have F.»

A direct formal contradiction can thus be avoided. While the Fx^c locution is perhaps ambiguous,[8] it is not obviously incoherent — just the sort of epistemic virtue needed in rescuing *Structure* from the dilemma explicated above.

One example (that Kuhn himself draws on)[9] may increase the plausibility of the reduplicative solution: the Gestalt Switch. A single drawing *as a rabbit* can be truthfully said to have ears, while *as a duck*, it cannot. The difference is not merely one of perspective, furthermore — it *really* is the case that the 'rabbit picture' has ears, and it *really* is the case that the 'duck picture' does not. These differences cannot be reduced to perceptual or epistemic differences — they are ontological, in some important sense.

Reduplicative propositions are common-place in English — I take this as further evidence of their cogency. Consider Morris' clearly coherent example of the shocking Jones who is an upstanding Sunday School teacher but a lecherous theology professor. Jones *as a professor* is a seducer of his students, yet *as a Sunday School teacher* he is not.[10] Both statements can be true at the same time without fear of contradiction. While success of this analysis will rest upon the ultimate coherence of reduplicative propositions, a subject beyond this paper, I here suggest that Kuhn's central thesis have more promise when seen 'as (iii)' than 'as (i) or (ii).'

A respondent may object that (iii) is no different than (i) or (ii), in that it faces an analogous dilemma. It must collapse into one of the two untenable positions already articulated. There are, after all, either one or more than one objects of study. x^c must be either identical to, or distinct from x^d. If x^c is identical to x^d, then there is just one object, and the formal contradiction of (i) can readily be derived — one object with an inconsistent set of properties. If x^c is *not* identical to x^d, two distinct (non-identical) objects are being studied, and the anti-realist folly of (ii) has been succumbed to.

Though this objection seem on its face to be a powerful one. I grant that the dilemma it presents has some intuitive force, but I believe it to be misguided nonetheless. The inference of the dilemma's first horn seems valid enough. Its seems, in fact, to merely rely upon an iteration of the universal Principle of the Indiscernability of Identicals, as innocent a premise as any. The inference in question can be formalized (using my reduplicative notation) as something like the following:

---

8.   For example, it could refer to a mode of being («x as an electron») or to a mode of perception («x seen as an electron»). The distinction between the two could prove to by an important one in attacking or defending the coherence of my reduplicative analysis.

9.   See Kuhn, 63, 111-114, 127, 150. At one point in *Structure*, Kuhn himself dismisses the Gestalt switch as a complete model to understand changes in paradigmatic commitments (p.85). In other passages, however, he suggests that something closer to the opposite is the case. I shall simply assume that this aberration is an oversight on Kuhn's part and will side with the majority of the textual data in giving a reduplicative reading of what happens in cross-paradigm study (scientists of differing paradigms studying the same world) and paradigm shifts.

10.   Thomas V. Morris, *The Logic of God Incarnate*. (Ithaca: Cornell University Press, 1986), 48-49.

iv.　　$\forall x, c, d\ (x^c = x^d \supset \forall F(Fx^c \leftrightarrow Fx^d))$.　　«For every x, every *c*, and every *d,* if x *as c* is identical to x *as d*, then for every feature F, x *as c* will have F iff x *as d* has F.»

This seems reasonable enough. If the Principle of the Indiscernability of Identicals is a law of logic, and reduplication is a coherent maneuver, furthermore, (iv) is not only true — it is necessarily so — a theorem of basic predicate logic with identity. Thus, the option of positing x^c and x^d as identicals seems precluded. To claim this option would be to affirm a contradiction of the most egregious sort: one subject with explicitly incompatible predications made of it. The first horn of the dilemma, then, is not a tenable option for a defender of the reduplicative reading of Kuhn.

Despite the apparent validity of (iv), and its admittedly intuitive force, it is not a sufficiently strong premise to make the second horn of the dilemma fatal to my project. Here's why: the second horn of the objection relies upon a *different* inference, which I hold to be invalid. Instead of merely positing that identical objects have identical sets of properties, it claims that if some object x is referenced via a reduplicative proposition in two distinct orthographic strings, and that the two reduplications are not identical, then 'x' being referenced in the first reduplicative proposition must be non-identical to 'y' in the second reduplicative proposition. The force behind this horn of the objection seems to rely upon an inference, then, of the following sort:

v.　　$\forall x, y, c, d\ ((x^c = y^d) \supset (x=y))$

«For every x, every y, every *c*, and every *d*, if x *as c* is not identical to y *as d*, then x as such is not identical to y as such.»

But what reason is there to believe that this follows? It seems consistent to posit x and y as identical (disregard for a moment their disparate designations 'x' and 'y' to see this), despite the distinction between their 'reduplicative counterparts.' Application of a prior example will illustrate my point. If (v) were a valid rule of inference, the following conjunction I've already considered *could not* be true, where 'Jones' is taken to rigidly designate the same person in both atomic propositions:

vi.　　Jones *as a professor* is a seducer of his students, yet *as a Sunday School teacher* Jones is not.

If (v) is a law of logic akin the Principle of the Indiscernability of Identicals (from which I take it to derive much of its intuitive force), it must be *necessarily* true. But what is at stake by claiming that (v) is a necessary theorem of some logic? I take this to in turn entail that there is no possible world W in which there are some objects x and y such that x is identical to y where x *as c* is distinct from y *as d*. And yet, it seems more obviously *possible* that (vi) than that □(v). This is so since (vi) seems to point to a genuinely possible world in which the very state of affairs precluded by (v) obtains. Since (vi), of course, can be possibly true only if □(v) is false, I suggest that one is justified in rejecting □(v), and, by extension, (v) as a law of logic.

I conclude that (v) may obtain in some worlds, but if it is not a law of logic, we have little reason to believe that it obtains in the actual world. If a cognate of (vi) actually obtains, we have good reason to believe that (v) does not obtain *in the actual world*. The evidence is thus stacked against (v), and for this reason, I suggest that it is false.

Taking on the second horn of the dilemma as I have done entails that that x^c and y^d are non-identical.[11] But this is not as problematic as the objection initially suggests. That two objects 'x *as c*' or 'y *as d*' are distinct, as I have argued above, does not preclude the identity of x as such and y as such. Furthermore, that x as such and y as such remain identical (despite differences which obtain between their reduplicative counterparts) can ground a robust metaphysical realism, even across paradigmatic divides, giving Kuhn just the sort of reality hook necessary to maintain the claim of a scientist experiencing a paradigm shift that, «Whatever he may then see, the scientist after a revolution is still looking at the same world.»[12] Hence, I conclude that the objection is not a defeater to my project or to Kuhn's. I finally note that this recognition makes room for the possibility of (iii), not just as a good reading of Kuhn, but as being within the realm of genuine possibility — and perhaps even something we have reason to believe actually obtains.

The same worry I have already considered can be stated somewhat differently. Let us suppose that there is some object x that *as c* is F and *as d* is not-F. Does the conjunction of these two atomic propositions yield the contradiction of (i) — that 'Fx & ¬Fx?' I see no grounds for this conclusion. To claim that (iii) collapses into (i), as the objector here does, seems to simply to beg the question against the defender of the reduplicative reading of Kuhn.[13] It is to suppose that there is nothing different about a reduplicative proposition and an atomic proposition of the normal variety — but this is the very matter under dispute. The move being made by my potential detractor, then, is simply a denial of my project's success — but absent independent reason to affirm this denial, it is a dialectically suspect maneuver, and persuasive only to one already convinced of the detractor's position.

I have considered a means of digging Kuhn out of a potentially deep pit. I have furthermore defended this analysis against a powerful objection. My solution has three advantages. First, as already articulated, it avoids the 'formal contradiction or anti-realism' dilemma. Kuhn and his defenders need not posit objects of scientific inquiry under differing paradigms as entirely distinct in all sense — nor need they claim that such objects have contradictory property sets. Second, it suggests an immediate research project which may prove fruitful; reduplicative propositions. While some quality work has been done in this area in the history of philosophy (eg, Aristotle's sketch of what it is to speak of 'x qua y'), there is no systematic, analytical body of work on the topic. Reduplication could be treated as a special sort of modal sentential operator, as a component of the subject of predication (as I have done above), or as a descriptive addition to a rigidly designating name, fixing reference thickly, so to speak.[14] Third, I note that my solution coheres well with Kuhn's overall project and self-identification:

---

[11].   This is my way of making sense of Kuhn's claims that the «data themselves change» subsequent to a paradigm shift. Kuhn, 134.

[12].   Ibid., 129.

[13].   For one enlightening (and amusing) take on presumption and begging the question, see Peter van Inwagen, *An Essay on Free Will*, (New York/Oxford: Clarendon Press/Oxford University Press, 1983), 18, 101-103.

[14].   Kuhn himself intriguingly suggests something similar to this in the descriptivist theory of reference he assumes when discussing the different worlds observed by scientists of differing paradigms: Kuhn, 148.

But is sensory experience fixed and neutral? Are theories simply man-made interpretations of given data? The epistemological viewpoint that has most often guided Western philosophy for three centuries dictates an immediate and unequivocal, Yes! In the absence of a developed alternative, I find it impossible to relinquish entirely that viewpoint. Yet it no longer functions effectively, and the attempts to make it do so through the introduction of neutral language of observations now seem to me hopeless.[15]

The above quotation, combined with statements like, «though the world does not change with a change of paradigm, the scientist afterward works in a different world,»[16] suggest that Kuhn sees himself at a mid-point on a spectrum between creative anti-realism and deep or naive realism. My solution carves out a bit of dialectical space for Kuhn to work within by articulating part of that third way — reduplication.[17] In doing this, I have dismantled one compelling reason to view Kuhn as a creative anti-realist.[18]

Andrew M. Bailey
Biola University
<wrathius@gmail.com>

---

[15].   Ibid., 126.

[16].   Ibid., 121.

[17].   It's not surprising that Kuhn has often been interpreted as a creative anti-realist, given certain features of the history of thought at his time. Morawetz notes one of them: «Kuhn's model of scientific and conceptual revolutions fell on receptive ears and had influence beyond its intended domain. As an intellectual ploy, it resonated with the mid-century currency of Wittgenstein's model of language games. Just as Wittgenstein urged us to think of all communicative efforts (and, by implication, all thought) as embedded within a bedrock way of proceeding that is not subject to reflection and questioning, a form of life, Kuhn seemed to apply this insight to science.» Thomas Morawetz, *Paradigms, Assumptions, and Strategies: Royce and Method,* 10 Conn. Ins. L.J. 123, 2003/2004.

[18].   I am indebted to Gregg Ten Elshof for his critical comments that contributed to the development of this paper. All errors are, of course, my own.

NONSENSE AND THE PRIVACY OF SENSATION

Juan José Acero

————————————————

       Among the many threads that make up the so-called Private Language Argument (PLA, in what follows) comes to the conclusion that the philosopher who gives credence to the possibility of a private language does not have an insight into the proper function of words like 'pain', 'I', 'know' and a few others. This is what I will call the Epistemic Privacy Way.[1] The specific argument that follows this route lays the blame on such a philosopher for thinking that there are linguistic vehicles — I will restrict my attention to certain sentences — that are apt to convey incorrigible and immediate knowledge of our own minds. It is alleged that 'I know I am in pain', 'But I must know if I am in pain!' or 'Only you can know what you feel', among many others, belong to the kind of sentences that can be used in this manner. In contrast to such a view, it has been adduced that to recruit those sentences for the purpose of serving the demands of an infallibilist view of inner perception is, as Wittgenstein showed many years ago, doomed to failure. There is no such a use, and philosophers that are not aware of it end up speaking nonsense, i.e. depriving those sentences of the function speakers require from them. The Epistemic Privacy Way thus leaves room enough for language analysis to play a role in critical philosophy and highlights the fact that not everything in PLA depends upon some or other subtlety, whether metaphysical or epistemological. As is well known, Wittgenstein considered that the philosopher's task consists of bringing words back from their metaphysical to their everyday use (*PhI*, § 116), and the Epistemic Privacy Way nicely fits in with this metaphilosophical advice.[2]

       On the other hand, if we aim at explaining why the private language advocate does not succeed in bending language — that is, the part resorted to — to his or her will, a detailed account is due that pinpoints where the philosophical mistake lies. Which amounts to spelling out a view of nonsense that squares with what PLA ends up. In other words, it is one thing to embrace the Epistemic Privacy Way and to agree that the believer in the possibility of a private language has not succeeded in endowing his or her claims with sense; it is another thing to found such a criticism on a systematic view of nonsense. This is what this paper aims at. Accordingly, in § 1 the view of nonsense expounded by Cora Diamond in her book *The Realistic Spirit* (Diamond 1990) will be spelled out. In § 2 I will examine this more in-depth

————————————

[1].   Here I follow Hacker (1990: 54&ff.)

[2].   Hacker explicitly traces out this connection when he writes that Wittgenstein «insisted that it is wrong, even *nonsense*, to say 'I know that I have a pain' […]. Such an utterance can be grammatical assertion, or merely an emphatic affirmation, but not, as philosophers typically take it to be, an epistemic claim» (Hacker 1990: 56). However, the most penetrating attempt I know of to explain why 'I know I am in pain' lacks sense is carried out in Kripke (1982: 117&ff.) I do like Kripke's analysis because of the connection he traces out between that sentence and 'It is 5 o'clock p.m. on the Sun' — an example I will be largely concerned with below.

in order to emphasize its affinity to a certain condition, the Meaning Restricted Exportability Requirement (MRE), which contains the essentials of Diamond's analysis of nonsense. I will also argue, in contrast to Diamond, that MRE is compatible with the Meaning Compositionality Principle, according to which the meaning of a complex expression is a function of the meaning of their parts and of the expression's constituent structure. In § 3 two aspects of MRE will be unfolded, its syntagmatic and its paradigmatic dimensions, in order to pave the way to acknowledging two kinds of relationships concerning the exportability of sentences and their constituents. In § 4 I will pause to take up an objection aimed at applying Diamond's view of nonsense to cases very similar to the ones that matter in ALP. Answering the objection hinges upon distinguishing meaning from thought, a distinction that echoes back to Wittgenstein's distinction between depth and surface grammar and at the same time sheds light on the sort of temptation to which philosophers are prone. Finally, in § 5, I reconstruct the steps to be followed by anyone adhering to the Epistemic Privacy Way and advance the main thesis in this paper, namely, that the possibility of a private language is an idea fuelled by the failure to comply with the paradigmatic exportability constraints to which certain constituents of sentences and thoughts are subjected.

1 Diamond on nonsense

In «What Nonsense Might Be» Cora Diamond offers a weighted answer to the question of what nonsense consists in. Nonsense has, in principle, two sources. Sentences lack sense either because of a categorical error or because of the presence of a constituent that bears no meaning or sense; and thoughts lack content because of either a categorical error or the presence of a constituent devoid of it. According to the first attempt at an explanation, nonsense is the outcome of a categorical incompatibility between the meanings of two constituents. Both sentences and thoughts are complex, articulated entities, and the composition of their constituents' meanings and contents must be made in accordance with rules that fix which combinations are permitted and which others are forbidden. Any failure to be in line with those rules results in a lack of sense. This is what Diamond calls the natural view of nonsense. According to the second source of nonsense, not having been successful in providing a sentential or thought constituent with a meaning or content, i.e. not having managed to give it a use in the context of the sentence or the thought where the constituents occurs, is what accounts for the lack of sense. This view of nonsense Diamond attributes to both Frege and Wittgenstein, and that is the view she herself adopts. Each view has its own banner. The first one waves in sentences like (C) (for Carnap); the second, in sentences like (M) (for Moore):

    (C)   Caesar is a prime number

    (D)   Scott kept a runcible at Abbotsford

While (C) illustrates the first kind of deficiency, the second one is discerned in (D). Of course, this is not the final analysis, because the advocates both of the natural and of the Frege-Wittgenstein views mean to give a general diagnosis of where nonsense originates, which forces them to justify why the other's banner could be waved in his or her own parade.

Diamond is on Frege and Wittgenstein's side. This view's very idea is that, regardless of the appearances, the lack of sense of sentences like (C) is due to the same principle that explains the senselessness of (D) and similar cases. The natural view, Diamond points out, confronts the difficult prospect of construing cases like (D) as examples of categorical mistakes — not an easy task since 'runcible' is not an English word and *a fortiori* has no

meaning or content that can contribute to any sentence it could occur in. The Frege-Wittgenstein view, on the other hand, is adequately equipped to construe (C) in harmony with what (D) is supposed to show. Diamond's argument for this depends on Frege's Context Principle, that is, the claim that a word has meaning only as a constituent of a sentence (alternatively, that a concept only contributes a content to a thought only as a constituent or a part of such a thought[3]).

> The expressions we see or hear can be identified with the items to which a definite meaning has been given only in a sentence which does make sense. In general, then, what the assignment of meaning to Logical Elements does is connect a sentence's being constructed out of these Elements in some definite way with its expressing some definite sense (Diamond 1990: 100).

According to this, (M) is nonsense because no rule has been stated that assigns 'runcible' a meaning, and this amounts to saying that no rule has been laid down that makes 'runcible' either a meaningful constituent or provides it with a syntactic function *vis-á-vis* the remaining constituents which make up what seems to be a sentence. Therefore, its occurrence in (M), alongside with 'Scott kept a ___ at Abbotsford', does not therefore result in a meaningful sentence. No thought could find expression in this combination of words. Now, what is remarkable, according to Diamond, is that the same diagnosis can be made of (C). Though (1) is perfectly in order,

(1)   Caesar crossed the Rubicon

it does not follow from that fact that any of its constituents, i.e. 'Caesar', has to bear the same meaning anywhere else, independently of its being a constituent of that sentence. The step that mediates between the premise that in (1) 'Caesar' names the history character, the Roman Julius Caesar, and the conclusion that 'Caesar' also names that Roman dictator in (C) is wrongly taken. When no rule has been enforced that adjusts the role played by 'Caesar' *vis-à-vis* '___ is a prime number', no meaning can be assigned to such a name, i.e. no sense can be made of it, in such a linguistic context. (C) lacks sense, not because any clash of semantic categories, not because arithmetical properties cannot be predicated of human beings — as the natural view of nonsense would suggest, but because the sort of imbalance between 'Caesar' and '___ is a prime number' that exists in (C) is the same you find in (M) between 'runcible' and 'Scott kept a ___ at Abbotsford'. As it happens, there is no need for postulating two sources of nonsense, one for (C) and another one for (M).

It is worth stressing that Diamond's diagnosis smoothly squares with an idea of meaning and content Wittgenstein made popular, namely, that the meaning of a word is its use in language, and that the content of a concept is its use in thought, i.e. in judgement,

---

[3].   In what follows, and in order to make the text more readable, I will abide by the following convention, namely, that any conclusion about sentences and their constituents I may arrive at will also apply, in what should be a plain parallelism, to thoughts and their constituents, i.e. concepts thoughts are made of. I will also follow the converse norm. I choose to follow these conventions because the kind of nonsense I want to discuss in this paper, and my reconstruction of PLA in particular, does not force anyone to adopt the more restrictive, albeit more realistic as well, starting point of accepting that thoughts are what sentences express *relatively to* extralinguistic contexts. In general, the sort of examples tI will bring up below does not play so significant a role that should oblige us to take into account extralinguistic roles separately. Put differently, the only kind of context for either a word or a concept that will matter to assess their contributions to sentence meanings and thought contents are further words and concepts. Nevertheless, in some of Wittgenstein's texts quoted below (in § 3), especially the one extracted from *OC*, § 333, extralinguistic or extraconceptual context makes a substantial contribution. If justice had to be made to this example, the exposition that follows would contain an excess of detail and would be much too involved. I expect, however, that by taking this shortcut I am not leaving out anything worth of consideration.

inference, deliberation and so on. Words are tools and should be understood as such. Despite the fact that the analogy is widely acknowledged nowadays, one of its consequences has to be highlighted one again. Only as long as they are properly connected to further words they can contribute a meaning to a sentence, in the same way as the 't' key on my computer keyboard can add a 't' token to a text on pressing it, if it is properly connected to rest of the system. Its role is played in so far as it is part of the sentence gear assembly. If we take the analogy seriously, then the Fregean Context Principle becomes a natural restriction to abide by in searching for word meaning and concept content. Words are like cogwheels or levers in a machine, and for them to have a meaning to contribute to a sentence requires them to be geared with the rest of sentence constituents.

> If we say 'A word only has meaning in the context of a proposition', then that means that it's only in a proposition that it functions as a word, and this is no more something that can be said that an armchair only serves its purpose when it is in space. Or perhaps better: that a cogwheel only functions as such when engaged with other cogs. (Wittgenstein 1975: 58.)

> The proposition, having multiplicity, is therefore a complex. Its constituents are words. Have words meaning apart from their occurrence in propositions? Words function only *in* propositions, like the levers in a machine. Apart from propositions they have no function, no meaning (King & Lee, eds., 1980: 2).

Therefore, Diamond's defence of the Frege-Wittgenstein view of nonsense exploits the virtues of the Context Principle to untie the knot of nonsense. It is now evident that her relying on the Context Principle to explain how nonsense emerges is in line with a use view of what meaning and content understanding consists in.

The next step is that of extending the Frege-Wittgenstein view to examples of nonsense much less apparent that either (C) or (M). They combine words in a manner that to decree their non-sentencehood would be to fly in the face of common sense — though, needless to say, common sense would be now informed by linguistic theory. They are sentences, but for a number of reasons it can be argued — in fact, it *has* been argued — that they might not express thoughts. That is, there is no guarantee that any of their utterances is the vehicle of a thought. Put briefly, their part-time nonsensical nature does not immediately strike us. Consider the following examples for discussion:

(2)   It is 5 o'clock p.m. on the Sun

(3)   Sapin is above New Zealand

(4)   I know I am in pain

(5)   He is in pain or he isn't

(6)   Milu believes that Tintin is in danger

To begin with, I am almost certain that *in today's* world we can think that it is 5 o'clock on the Sun. However, the same assurance cannot be credited anyone who believed in the same thought eighty or one hundred years ago, when the electronic or atomic clock technologies had not been developed and Earth rotation was measured according to the Sun's apparent movement. (Actually, this is not the whole truth, as the knowledge of how to calculate the time on the Sun by means of pendulum's movement dates from ancient times. But, again, the position of the Sun over the horizon played a role in making the usual estimates.) The reason for such an impossibility is that in order for us to think that it is 5 o'clock on the Sun, it is necessary to think that the Sun keeps with itself the kind of relation

that holds when it is 5 o'clock p.m.; that is, the kind of relation that would hold if the Sun could be seen from the Sun's very surface as occupying the 5 o'clock position! Even if someone were to assure us that he or she is able to think or to be thinking such a thought, we should insist that (1) expresses an apparent thought. To be in position to think that thought, one has to have acquired a rather sophisticated intellectual training. Somewhat less demanding are the conditions that control the expression of thoughts by utterances of (3). In this case, the difficulty is made apparent by the relational predicate 'is above'. No ordinary English speaker hesitates to tell whether the picture is above the chimney or the clouds are above the mountains, if the circumstances are the usual ones. On the other hand, to overstate the use of 'above', moving away from the norm designated by those central cases, to the point of demanding a yes-or-no answer to whether Spain is above New Zealand is to drive us over the edge of nonsense. If anyone utters (3) while thinking that New Zealand is on the antipodes of Spain, then to resort to the predicate 'above' can be interpreted as the choice of an austere and common mean instead of a more demanding one. Maybe the speaker is about to make a joke. However, 'to be on the antipodes of' is a complex relational predicate whose use is subject to rather strict conditions. If all one takes into account when uttering (3) is a quick glance at a map of the world globe that prompts in your mind an image of Spain's being over New Zealand, then nothing warrants that a real thought has been framed. I would say that, if the only means available to that speaker is an image together with the usual linguistic resources, such a speaker could well fail to articulate a thought. Clearly, this use of 'above' significantly differs from that of 'the picture is above the chimney'. Should it be true that the picture is above the chimney, then it follows that it has to be false that the chimney is above the picture. But if someone feels free to claim that Spain is above New Zealand, then he or she could add that New Zealand is above Spain as well. The fact that we have here different uses of 'above' and, correspondingly, different concepts of spatial relations, but only one word to cover them, gives rise to an illusion, namely, that it will suffice to entertain a clear image to think a thought. As for (2) and (3) I would add that the standard linguistic and conceptual resources available to people who lack a specialised training, and that are unable to see through them, contain blind spots, and this explains why they occasionally produce apparent thoughts. An unreflective experience of the Sun has created a blind spot on the conceptual system with which we estimate time; and a superficial understanding of the nature of closed, curved surfaces has given rise to another blind spot, this time on the conceptual system we use to map spatial relations.

Wittgenstein, who is the source of the problem (2) and (3) pose,[4] believed that the reason why they could express apparent thoughts makes clear why (4) and (5) could lack sense as well. In these cases, nevertheless, the question whether they make sense or not gets mixed up with the question of whose side you are on in some debates in philosophy of logic and mathematics — is the Law of Excluded Middle valid? — as well as of philosophy of mind — are incorrigible your judgements on your mental states and processes? This relation makes it reasonable for us to take precautions against focusing on such sentences if we seek to settle the question whether they are senseless or not. Accordingly, I will follow the strategy of putting them aside for the time being, aiming at a general analysis along elsewhere and then trying to see what might follow for the cases in point. But this in no way means that I would welcome to set up a principled distinction between, say, (6), and (2) or (3). I am not. Hergé and Goscigny repeatedly picture situations in which Milu thinks that Tintin is in danger. If we

---

[4]. Wittgenstein (1953/2001: §§ 350& f.) is also the source of (2) and (3).

did think that dogs cannot frame a thought with the conceptual and behavioural complexity required to have the content that a certain person is in danger — no such content is possible without concepts, and concepts either are or evince abilities dogs do not have, it would follow not only that (6) is false, but also that it is a nonsensical sentence.

So, what can be said about (2) and (3) from the Frege-Wittgenstgein view of nonsense? In case the final answer should be that none of them makes sense, the corresponding justification would have to mirror the kind of argument Diamond provides for (C) and (M) that we have put forward above. This task must confront the difficulty of bringing a dislocated item in (2) and (3) to view, as happens with 'runcible' in (M). This does not seem to be a minor point, as such an item seems to be absent in (2) and (3). However, Diamond's penetrating suggestion is not that nonsense originates in the presence of so-called runcible words and items. The point about these cases is — exploiting an analogy with jewellery craftmanship — whether each and every word is mounted on the sentence the way they should, not whether they are already acknowledged in language. (Of course, if it is not, they cannot be fitted with further words or constituents in a sentence frame.)[5] Thus, turning to (2), the question that has to be answered is whether we have managed to provide 'the Sun' with a use *vis-à-vis* the rest of constituents. As for (3) the question is: Does 'is above' have a meaning *vis-à-vis* the rest of the sentence constituents? Both questions must be answered in the negative. The design of a linguistic framework to state time estimates no matter where one might be situated on the Earth must entertain the possibility of combining a name of such a place ('Mexico, D.F.', 'Granada', 'Helsinki') with the predicable 'It is 5 o'clock on ___' while implicitly barring 'the Sun' from being taken as a legitimate choice. (The conscious inclusion of 'the Sun' in the list means that another point of reference to estimate solar time has been adopted.) Why not interpret this condition as a proof that 'the Sun' bears no meaning when inserted in the predicable blank, i.e. when it occurs in the sentence (2)? An analogous argument could be run that would justify why by uttering (3) someone did not succeed in expressing a thought. (3)'s lacking sense in some occasions of their use is due to the fact that, far from being true that (3) results from importing the relational constituent 'is above' from, let's say, 'The picture is above the chimney' to the predicable 'Spain ___New Zealand', 'is above' cannot be mounted on the latter. In the same way, (6) could be involved in nonsensical chatter, because the proper name 'Milu', unlike 'Captain Haddock' cannot be set off for the predicable '___ thinks that Tintin is in danger'. Following Frege and Wittgenstein, Diamond ought to opt for this way of analyzing (2), (3) and (6). Though I share this analysis, in the next section I will set out where my disagreement with her lies. Every word and every constituent is a runcible somewhere (sometime).

2 Restricted Exportability

One reason why I believe that Diamond's view on nonsense should be seriously taken into consideration is its flexibility. Though she does not seem to be interested in the kind of nonsense (2) — (6) illustrate, her analysis of nonsense can be extended to them in a natural way. Moreover, her analysis squares with the existence of blind spots in our conceptual system and promises to shed light on their nature. On the other hand, Diamond claims, finding inspiration in the work of both Frege and Wittgenstein, that nonsense is what one should

---

[5].   I prefer this analysis to arguing with Hertzberg (2001: § 1) that both (2) and (3) express mental contents that contain gaps to be filled, because by putting things in this way I feed the impression that these sentences are assimilated to (C) — what would be a mistake.

expect to obtain if the Functionality (or Compositionality) Principle is applied unrestrictedly. This principle holds that the meaning of a complex expression, i.e. a sentence, is a function of the meaning of their constituents and of the syntactical mode of combining them, i.e. of their syntagmatic relations. She interprets this condition as requiring that the meaning of a complex expression is a complex extralinguistic entity made up of the meaning of the expression's constituents and their mode of syntactic combination. Therefore, nonsense is what results if the meanings of the constituents do not fit into each other in the right way, something that takes place when rules of semantic combination do not operate on the required arguments. The Compositionality Principle, argues Diamond, is tied to the natural view of nonsense:

> […] on the natural view there is and on the Frege-Wittgenstein there is not what you might call a functional account of nonsense. Now, on any plausible view you can say that the sense of a sentence depends on the meanings of the words of which it is composed. […] What the natural view holds, though, is something beyond this: that whether a sentence makes sense or not is functionally dependent on its parts, on their logical category. […] The Frege-Wittgenstein view does not take any kind of nonsense to be functionally dependent on the categories of the terms combined in a sentence. If we are not talking of the category of the thing you psychologically associate with the word, then to give the category of a word in a sentence is to give the kind of work it is doing there. The word does not have a category assigned to it which it brings with it into *whatever* context (Diamond 1990: 104).

'México D.F.', 'Granada' and 'Tombuctú' are names that can be combined with the predicable 'It is 5 o'clock p.m. on ___'. 'The Sun' cannot. And an analogous remark is valid of 'is above' and 'Spain ___ New Zealand'. The fact that these constituents make a semantic contribution to any sentence they might occur in does not support the conclusion that they can do so in any other sentences as long as syntax and morphology requirements are complied with. As I see it, the most suitable way of making this point[6] has it that words and concepts acquire an use in more or less specific or paradigmatic circumstances, and that such a use develops into new ones as words get anchored to new situations, which to a certain extent may be similar to the former ones. I take it that we first learn to say and think that it is 11 a.m. 5 in Granada, 5 p.m. in Helsinki and so on; to learn and think that the picture is above the chimney, that Los Angeles is above San Andres fault, etc. However, as soon as we move beyond the circumstances in which the target expressions receive a use, the risk of giving rise to an apparent thought becomes real. And this is what in fact transpires in both (2) and (3). No word or expression carries with itself the meaning it has in a sentence or phrase to any other linguistic context.

It is worth dwelling on this point, because Diamond seems to go into a dead end in making (M) the measure of nonsense, i.e. in claiming that nonsense results from having failed or forgotten to mount an expression on a sentence with the rest of its constituents. The Compositionality Principle certifies that the meaning of a sentence is a function of the meaning of its constituents and of their syntagmatic relations. Once these meanings and relations are fixed, the meaning of the sentence is univocally determined. By invoking the principle we gain the right to assert that (M) lacks sense because one of its constituents either is senseless or its syntagmatic relations to other constituents in the sentence have not been settled. If this explanation has to guide us, it would follow either that (2) and (3) are perfectly in order — an uncomfortable conclusion — or that they lack sense for the same reason that (C)

---

6. Due to Wittgenstein. See his *PhI*, § 347&ff.

does. Thus, the Compositionality Principle traces the analysis back to the natural view while denying the Frege-Wittgenstein view. Because of it, she is doomed to rejecting a compositional approach to meaning and content, and particularly the condition, built into the compositional approach, that semantic combinatory rules apply across the board, that is, no matter what constituents are acknowledged in the lexicon. Put alternatively, Diamond opposes the following

Meaning Free Exportability Requirement (MFE)

There is an autonomous repertoire of meanings, which the language lexicon captures, ready to bee freely combined by the compositional semantic rules.

Diamond openly rejects that words should be mapped onto grammatical categories, and thus that their meanings should fit into modes of compositions, independently of the contexts they might occur in. Neither word's meanings nor concept's content are freely exportable from contexts to contexts. Sentences (2) and (3) deserve close attention, since they reveal that MFE is all of a piece with a way of understanding language and thought that generates unexpected troubles. It is the so-called mechanical conception of meaning and thinking.[7] Diamond rejects the Compositionality Principle because she conceives of it as a mechanism, though described in abstract terms, which compounds meanings and thoughts out of simpler meanings and concepts. Those outcomes are the values of a complex function which take as its arguments the meanings of constituents and whatever syntactical categories they belong to. According to such a conception the Compositionality Principle does its best when taken as psychic mechanism, maybe endowed with a computational nature. Seen to this light, the principle plays an essential role within a representational conception of mind and language. Which thought we may think, which *content* it may have, will at last depend on what representations the thought is made up of, what syntactical patterns these representations fit into, and what meanings or contents they have. All these pieces are put together by the Compositionality Principle, which accordingly plays a centrally strategic function in depicting human mind as a sort of mechanism. In Baker and Hacker's opinion, it is an essential part of a calculus view of language as much as of a generative theory of meaning understanding. Regardless of whether sentence meanings are made up of the meaning of its constituents or simply are functionally determined by them,[8] subject to their mode of syntactical composition, either to understand a sentence or to grasp a thought consists of calculating a value out of the value of its parts and their syntactical relationships.[9]

---

[7]. Baker and Hacker (1980: 267&ff.), Heil (1981), McDonough (1989) and Wright (1989) contain criticisms of this conception. In another place Diamond asserts that «[i]t is just a phrase that we have put together by analogy with other phrases that we use» (Diamond 2000: 69). This suggests that she might welcome what Robbins has called the Minimal Compositionality Constraint: «The content of indefinitely many complex concepts is exhaustively determined by the concepts of their constituents and the rules governing the combination of those constituents» (Robbins 2002: 319). This constraint would allow excluding (2) and (3) as meaningful sentences. I gather that Robbins consciously articulates his constraint by substituting 'every' for 'indefinitely many'.

[8]. Baker and Kacker (1980: 263) and a few others do not make this distinction, which it is forced on us anyway. Meaning compositionality does not only requires 'composition' to be literally understood, but it makes it necessary that there be a functional relation between the meaning of the complex and the meaning of its constituents. Obviously, a composition relation is functional, but not the other way around. Unless a difference between functionality and compositionality is acknowledged, we should follow Hale (1997) in distinguishing Weak Compositionality from Strong Compositionality.

[9]. See Baker and Hacker (1980: 263&ff.).

To my mind, Diamond's rejects the Compositionality Principle because she makes an unnecessarily restricted reading of it. This becomes plain when she explains how she believes our knowledge of language rules, which governs not only the meaning of its expressions but their combinatorial possibilities as well, should *not* be understood. After making it clear that meanings and syntactical categories are not unconditionally exportable, she points out that language rules do not say what role expressions play in sentences, including their semantic contribution:

> The word does not have a category assigned to it which it brings with it into *whatever* context. This is not to say that words are not assigned to categories, but that the identification of a word in a particular sentence as playing a certain role, as meaning a certain kind of thing, cannot be read directly off the rules. (Diamond 1990: 104.)

Diamond's alternative to a conception of rules that views them as content containers is a conception that understands rules as means of analyzing the expressions they govern. The analysis carries out two tasks: rules split up sentences — complex expressions — into pieces or parts, i.e. names, predicates, relation terms, and so on, and fix their meanings as constituents of sentences. What really matters, she says, is that «neither kind of rule will apply unconditionally to a given sentence» (Diamond 1990: 109&f.). When she comes down to the last details, she insists on the idea that language rules capture what language users do with words and structures, instead of preceding those uses in the language constitution order. Thus, when tackling the sentence 'Venus is more massive than Mercury', she writes:

> We may know that the proper name 'Venus' stands for Venus; our knowledge may now be conditionally applied: the sentence is the proper name 'Venus' *standing for Venus*, in the left-hand place of the relational expression, with the proper name 'Mercury' in the right-hand place, only if the thought expressed by the whole sentence is that *Venus* has whatever relation 'more massive than' stands for whatever object 'Mercury' means. […] Taking the rules that fix the meaning of expressions in the language to apply to the particular sentence is not separable from making sense of the whole sentence. (Diamond 1990: 110.)

Nevertheless, I see nothing here that an advocate of the Compositionality Principle may be not ready to embrace. Such a theoretician would refuse to accept the view that, for example, the constituent 'got burnt' makes the same contribution both to (7a) and (7b):

(7a)   The forest got burnt (= Some burnt the forest)

(7b)   The forest got burnt (= The forest was consumed by the fire)

In order to mark and expose differences of meaning or thought that could matter, the believer in compositionality will look for syntactic categories, appropriate assignments of items to them and the corresponding semantic combinatorial rules that provide (7a) and (7b) with different meanings, i.e. truth-conditions. Therefore, the analysis put forward by the theoretician will keep an eye trained on (7a) and (7b) and the other one trained on the meanings they respectively express, aiming at the conditionality requirement, to which the application of language rules are subject, as Diamond insists. If follows that the Compositionality Principle can be read in a manner that it does not become part and parcel of the mechanical conception of thinking and language understanding. On the contrary, the Compositionality Principle leads to the acknowledgement of systematic relations, of a functional nature, between the meanings of complex expressions and the meanings of their constituents. It simply claims that the meaning of a sentence, the meaning of their constituents

and its structure are systematically related.[10] No more demanding requirements should be allowed to bear upon it. Moreover, when so interpreted, the conflict between the Compositionality Principle and the Context Principle, which some philosophers have pointed out, smoothly dissolves. On the one hand, the Compositionality Principle is addressed to the conditions that rule the links between the meaning of any complex expression, particularly any sentence, and the meanings of its constituents. On the other hand, in its most natural interpretation the Context Principle claims that words and expressions in general can contribute a meaning as long as they are sentence constituents. Thus, no unbearable tension results from these balanced readings of both principles.[11]

Once the Compositionality Principle is given an explanatory role to play independently of the mechanical conception of thinking and meaning, Diamond's objections to the natural view of nonsense become arguments against MFE. And there is nothing in meaning compositionality that stands in the way of replacing that constraint by a weaker one. The following one is a natural choice:

Meaning Restricted Exportability Constraint (MRE)

Any sentence constituent carries its meaning with itself to any other context which is appropriate to that constituent.

Of course, 'appropriate to its constituent' is the key clause in MRE. It alludes to nothing that contemporary grammarians ignore nowadays. Far from being true, it is a significant part of their business. By assigning words and expressions to syntactic categories in a systematic way, they attempt not only to capture, but also to anticipate, which contexts are appropriate to which items. Those assignments provide one sort of standard to catch those adjustments among expressions that have to be preserved and those that must be ruled out. Another standard lies in the subtleties involved in syntactic and semantic composition, because they become a powerful instrument to fix those contextual restrictions we deem necessary. Thus, the challenge that (7a) and (7b) pose to us can be met as soon as we make it sure that compositionality and contextuality requirements neatly square with each other. This calls for two arrangements. Firstly, the verb 'to burn' is assigned to two different syntactic categories, one reserved to medial verbs and the other one to non-medial verbs. This step is taken in accordance with the spirit of the Context Principle. Secondly, the rules that govern meaning composition and in particular the contribution of the verb 'to burn' must be redesigned to answer the fact that 'The forest got burnt' is ambiguous.[12]

---

[10]. Some have argued that this relation is supervenience. See Szabó (2000).

[11]. To quote only one author, Dummett (1981a; 1981b) is the classic reference to put on record the fact that philosophers have not overlooked the existence of some sort of incompatibility between meaning compositionality and contextuality. It must be added that the conciliatory reading of the Context Principle I have chosen above, no matter how much reasonable it may sound, is not unanimously embraced nowadays. See note 12 to catch a glimpse of Diamond's troubles to deal with both principles simultaneously.

[12]. Therefore, MRE is in line with the structured nature of thought, a view that Diamond does not wish to deny, because it contains the key piece to explaining language learning and thought productivity. It also guarantees the existence of systematic relations among thoughts, provided that the Compositionality Principle makes it clear why two thoughts differ that result from combining the same conceptual items in different ways; and why thoughts that do not combine the very same items are distinct. It is no coincidence that Diamond rejects meaning functionality while openly welcoming the idea that the meaning of a sentence depends on the meaning of their constituents. It is no coincidence as well that she forcefully opposes MRE while, immediately after doing so, adding that this does

3 Two dimensions of meaning exportability

To deal with the kind of nonsense that crops up along the Epistemic Privacy Way, it is necessary to have a closer look to MRE and figure out what makes a context an appropriate or inappropriate one to import either a word or a sentence constituent. I will claim that meaning exportability can be assessed along two dimensions, in order to be faithful to the fact that two kinds of valuations have be taken into account to understand a word's disposition to fit into a sentential context gear assembly. I will borrow a distinction, which has been repeatedly tried out in structural linguistics, to characterize those dimensions and give them a name and speak of *syntagmatic* and *paradigmatic* relations (Lyons 1968: 70&ff.). The justification for borrowing such a pair of terms lies in its utility to the analysis of (4) to be carried out below (in § 5).

Compare (8) to (9):

(8)   I don't know whether there is a red book on the shelf.

(9)   I do not know whether I have ever been in China

Wittgenstein rejects that (9) will make any sense in normal circumstances: «Normally Europeans do know whether they have been in China or not» (*OC*, § 333). While the sentential context (¢ 9) is not appropriate for the expression 'I do not whether' to be exported to it, context (ç 8) is:

(ç 8)   ___ there is a red book on the shelf

(ç 9)   ___ I have ever been in China

'I do not know whether' is exportable to (ç 8), but it is not exportable to (ç 9). I will say, accordingly, that by not loosing sight of its exportation behaviour, we gain an insight into how different the syntagmatic relation it maintains to (ç 8) is if compared to those it holds with (ç 9). Those differences help explaining why (8) may be used to express a thought while

---

not entail the idea that words have to be assigned to syntactic categories (cf. Diamond 1990: 104). In fact, to my mind, the main problem that Diamond's paper poses for any reader is how to harmonize all these claims. For example, she seems to find in the Fregean doctrine that «we cannot properly ask for the meaning of a word except in the context of a sentence» (Diamond 1990: 109) the counterweight she is looking for to decide what is prior over what, whether the sentence over the word or the other way around, the word over the sentence. All this leads the reader to conclude that Diamond values the Context Principle higher than the Compositionality Principle. However, it is difficult to know whether this choice is full of content or not. In effect, compositionality is a blunt scalpel if Frege's Context Principle is read, as Diamond does following Quine, as requiring that «it is through the sense of the whole that the parts get their meaning» (Diamond 1990: 109). Let's take seriously the advice that sentences have priority over the words they are made up of and, accordingly, to go on to segment sentences assigning words to syntactic categories and meanings to the bits so obtained. Once this task has been accomplished, it follows that the meaning of the sentence is a function of the meaning of its parts and of its syntactic mode of combination. True enough; but the compositionality requirement is trivially satisfied if the sense of the whole comes first. It cannot be put to service to explain how language is learnt and thought is understood, unless we allow the domain to which the Context Principle is applied to be a proper subset of the whole set of sentences that define language. Sentence meanings and thought contents must be out there, many of them anyway, before we start cutting them into pieces. Therefore, Diamond denies compositionality, because she takes it to be incompatible with sentence priority, while on the other hand she needs it as a condition to carry on those explanatory tasks. That her position is unstable is also shown by her brief comments on language understanding. A sentence is grasped by applying rules that analyze it into words and map meanings onto them, but to follow these rules does *not* consist of using the competence of recognizing its constituents, their meanings, and putting them together. To understand a sentence is to make sense of it; in other words, «to make the sentence *his*, but using the rules» (Diamond 1990: 111). This remark is hardly helpful, to say the least, because Diamond has not solved the tension described above.

(9) cannot. The exportability of 'I do not know whether' to (ç 8) makes it possible to express a thought; the failure to exporting 'I do not know whether' from (8) to (ç 9) explain why (9) is nonsense, i.e. why it usually cannot be used to express a real thought. Why certain words or expressions are not liable to be exported to specific contexts, while others are, is an important and difficult question. The Frege-Wittgenstein view of nonsense has it that exportability is restricted to those cases in which the exportable item has not been provided with a use *vis-à-vis* the target context; in other words, when language and thought use admit of mounting the moving item on the target frame. On this dimension what counts is whether these linguistic and conceptual gears successfully work or not.[13]

In caring about paradigmatic relations we pay attention to quite a different sort of concern. We do not consider a moving item E_j relatively to a target context C, but how does E_j behave relatively to a number of further exportable *items E_1,... E_{j-1}, E_{j+1},... E_n* with respect to C. Any difference between two members of this class with respect to C — one of them can be exportable to C, while another one cannot — is a difference in their paradigmatic relations. Alternatively, we could say that they have different paradigmatic properties. Wittgenstein made the most of paradigmatic relations in his sustained battles against the bewitchment of our intelligence by means of language:

> I ask someone 'Have you ever been in China?' He replies 'I don't know'. Here one would surely say 'You don't *know*? Have you any reason to believe you might have been there at some time? Were you for example ever near the Chinese border? Or were your parents there at the time when you were going to be born? (*OC*, § 333).

> That is to say: only in such-and-such circumstances does a reasonable person doubt *that*. (*OC*, § 334).

> The procedure in a court of law rests on the fact that circumstances give statements a certain probability. The statement that, for example, someone came into the world without parents wouldn't ever been taken into consideration here. (OC, § 335).

Here he is suggesting that the use of 'I do not know whether' is appropriate when, i.e. in those contexts in which, it is also appropriate to use 'I have reasons to doubt that'. It follows from this constraint that both 'I do not know whether' and use 'I have reasons to doubt that' are not exportable to either (ç 10) and (ç 11):

(ç 10) ___ I came into the world without parents

(ç 11) ___ cats do not grow on trees

If we are not sensible to these nuances, we run the risk of mistaking apparent thoughts for real ones, that is, mistaking (10) and (11) for meaningful sentences:[14]

(10) I know I came into the world without parents

---

[13]. Things are not so straightforward, as a more thorough consideration of (9) leads us to recognize. Wittgenstein points out that (9) makes no sense for Europeans, i.e. Western people: (9) does not express a thought if uttered in a conversation among them. However, as soon as we widen the language context by including in it the place where people are born and grown, then syntagmatic relations become quite distinct, and the possibility of exporting a word or an expression to a context now makes a difference for the speaker's CV. (See note 3 above again.)

[14]. Wittgenstein immediately adds that what human beings think is reasonable changes not only among persons but from one culture to another, and within one and the same culture, and from one time to another. See *OC*, § 336. (ç 11) is dealt with in *OC*, § 282.

(11)   I know cats grow on trees

And conversely: that any context to which 'I do not know whether' may be exported is also a context to which 'I have reasons to doubt' may be exported. In the structural linguist's jargon, whether 'I know' and 'I have reasons to doubt' belong to the same paradigm.

The fact that words and expressions share paradigms is something philosophically worthy of attention, because it offers clues as to the range of application of our concepts, as well as to the range of our words. It is not possible to export 'I do not know whether' either to (ç 9), (ç 10) or (ç 11), because ignorance and doubt are somewhat correlative: one is a natural move to make in language use just where the other one is a possible choice. Thus, if there is room for doubting whether such-and-such is the case or not, it must be possible to ignore that such-and-such. Doubt can take root only where the possibility of ignorance is built into. Conversely, nobody can ignore that such-and-such unless the possibility of doubting whether such-and-such is built into the very situation in which our ignoring takes shape. It follows that to determine the scope and limits of meaning and concept exportability, we are forced not only to detect what compatibilities and incompatibilities happen to take place between exportable resources and target contexts, but to polish off the ability to identify when those resources belong to a certain paradigm and when they do not. We do not grasp the function of a concept unless we somewhat map its place in the thought system. Put it in less abstract terms, we must be acquainted with its syntagmatic and paradigmatic properties to be able to pinpoint what difference it means to the possible meaning and thinking.

It is evident, then, that what lies behind MRE's opaque clause 'appropriate to that constituent' is the class of syntagmatic and paradigmatic relations any constituent enters into. It is also clear what is involved in looking at sentences like (2) and (3) through the emblematic (M). Those sentences make no sense because some of their constituents, i.e. 'the Sun' and 'is above', have not been provided with a use relatively to the rest of their respective constituents. Their syntagmatic and paradigmatic relations fix their functions' scope and limits. The fact that 'the Sun' holds such a specific paradigmatic relation to names like 'Mexico, D.F.', 'Granada', 'Helsinki', and so on speaks for itself of how delicate its place in the time estimation conceptual system is. And the fact that 'is above' has got the very particular syntagmatic properties it exhibits speaks for itself of how much subtlety there is in the system we have designed to locate places on the globe.

4 Meaning and thought

I have argued why MRE states what is essential to the Frege-Wittgenstein-Diamond view of nonsense. The time has arrived to face up to a difficulty which has been looming large since the moment we addressed the question whether or not it possible for (M) to model (2) and (3)'s oddity. Up to now, I have tried to explain why (2) makes no sense, i.e. why an utterance of it may express no more than an apparent thought. However, it might be adduced not only that the possibility of there being a thought with the content that it is 5 o'clock p.m. on the Sun is an illusory one, but also that the sentence 'It is 5 o'clock p.m. on the Sun' is in order according to both syntactic and semantic criteria. There might be no such a thought, but the sentence meaning is out there, ready to be translated to other languages ('Son las 5 de la tarde en el Sol' in Spanish, 'On kello auringossa viideltä iltapäivällä' in Finnish). Translation preserves meanings, at least to a certain point, something that it would not be possible to achieve, if there were no such meanings to capture in different languages, at least to a certain point. Translation, when conceived of in a certain way and performed well, aspires

to leave intact some sort of content or semantic value. However, this practice demands intellectual abilities that are, partially or wholly, independent of those required to frame a thought out of a concept repertoire. This remark can be made extensive to (3) — (6). In conclusion, there is something to arguing that (2) does have a meaning, although it expresses no thought. Since the terms 'meaning' and 'sense' have been chosen to mark the difference between (1), on the other hand, and (C), (M) and (2) — (6), on the other, are we not in the grip of an more or less explicit contradiction?

Throughout the discussion in §§ 1 — 3 no distinction was made between a sentence's capacity to have a meaning and its capacity to express a thought. Moreover, a serious commitment was advanced — see note 4 above — namely, that whatever would be claimed on words and sentences when dealing with nonsensical sentences would also have a straight counterpart within the domain of concepts and thoughts, so that it would be superfluous to break up the symmetry between the realms of language and thought. To claim that (2) could lack any sense in one or another of theirs utterances was another way of taking it as expressing no thought in those very occasions. The meaning that a sentence has in one occasion is the thought the sentence expresses in it. By adopting this equation we have managed to maintain the complexity of our discussion within reasonable clear bounds. Or so I hoped. In spite of it, the threat of contradiction is strong enough to modify our previous commitment without giving up the thesis that MRE is the cornerstone of nonsense analysis. The way-out consists of distinguishing thoughts from other kinds of semantic values. It will help alleviating the tension our first commitment gave rise to if we reserve the term 'meaning' to predicate semantic properties and relations of words and sentences. (« — What does 'Son las 5 de la tarde en el Sol' *mean*? — It *means* that it is 5 o'clock p.m. on the Sun.»)

Before hastening to welcome this reaction, it could be argued that the objection is based on a premise that concedes too much. It could be alleged that translation only makes sense if it is of a bit of speech or of a text that contains or expresses thoughts; in other words, that thought is what makes translation a serious business. When we discuss what people and texts say, the objects of everybody's assertions are the same as the objects of everybody's thinkings. It cannot be asserted that *p* unless it can also be thought that *p*, and conversely. Since there could be no thinking that it is 5 o'clock p.m. on the Sun, there could be no saying it too. Any language that created the room for the expression of such a thought would be an illogical language, as would one which allowed to say that the execution of Charles I is in the crater of Vesuvius:

> An «illogical language» would be one in which such a sentence would say of an event that it was in the crater. But there is no such saying as that, and no such thought as that: there is only a string of words imitating the expression of thought; […] there is nothing at all but a confusion of words which has the appearance of expressing something or trying to express something that we then say cannot be. (Diamond 1990: 105).

I take Diamond to be claiming not only that there is no saying that the execution of Charles I is in the crater of Vesuvius or that it is 5 o'clock p.m. on the Sun, but that those things cannot be thought as well. Therefore, (2) and (3) are no more strings of words than they are vehicles of thought. Taken at first value, they look like sentences, but they are not really. If we remember that Diamond thinks that (2) and (3) are like (M), we can infer from her words that the argument which adduced that there is something in (2) that makes it suitable for translation does not touch her at all. Not being sentences, why worry about it? (It would be tantamount to playing the game of translating into Spanish Chomsky's famous

example 'Colourless green ideas furiously sleep', at most an extreme and degenerate case of translation.)

I part from this reply, because there is a sense of 'saying' according to which someone can say that it is 5 o'clock p.m. though it is overlooked that there could be no thought behind his or her saying it. It seems to me that it is perfectly possible for anyone to be blind to nuances built into the estimation of time conventions and think it perfectly normal to ask what time is it on the Sun. If he or she does it and the speech occasion is not a scientifically sophisticated one, it becomes appropriate to object that his or her words lack any sense. It is natural to conclude, then, that assertions and thoughts do not have the same objects, though these objects may often coincide. This departure from the argument line developed until now does not forces us to give up the conclusion that nonsense results from a failure to mount an item on an appropriate conceptual frame. Grammar constraints do not have to be sensitive to each and every condition which features in the conceptual systems we employ. In spite of the fact that our sentences are usually adequate guides to our thought' contents and that, because of it, in setting up the limits of meaning we are more or less mapping the limits of our thought, it does not follow that sentence meanings must strictly keep track of thought's junctures. (2) and (3) are worth taking into consideration since they show that such a kind of gap is not merely a theoretical possibility. Explaining (2)'s oddity is laying out in what respect language only partially mirrors thought.

It is not necessary to enter now into sa ystematic reflection that justifies the need for distinguishing thought from sentence meaning. From the point of view adopted here, the border between those two domains is passed over on replacing MFE by MRE, i.e. on restricting meaning exportabiliby from certain sentential contexts to further sentential contexts. (These are also the limits beyond which the validity of the Compositionality Principle is in danger.) The reason on which MRE is founded — one that has repeatedly emerged above and that will reappear in the next section, where we will be concerned with some Wittgenstein's arguments against the very possibility of a private language — can be briefly stated in the following way: the limits to constituency exportation in cases like (2) and (3) are the limits to which the use of the resultant sentences are subject to. Unless there are use conditions for those sentences, they are vehicles of thought. Since (2) is useless in the circumstances described above, in expresses no thought *in them*, and nobody could think it is 5 o'clock p.m. in the Sun *there*. On the other hand, this sentence means the same as 'Son las 5 de la tarde en el Sol'.[15]

Wittgenstein could have welcomed a distinction close to the one I have put forward in separating meaning from thought. Those cases of nonsense we have been focusing our attention on typically reproduce an scheme in which words have a use as long as they go together in accordance with patterns that language reproduces again and again. That aspect of word use that becomes easily recognizable through sight and ear is something Wittgenstein refers to when he talks of word and sentence surface grammar:

---

[15]. In good faith it should not be alleged that exportability relationships are squarely understood by taking them to be computational relations. In a sense there is no denying that they are relations between sentential contexts, and that these contexts are complex symbols of a computational system. However, this is only part of the truth, because exportability relationships supervene on what use sentences involved have, i.e. in what circumstances they can be used.

> In the use of words one might distinguish 'surface grammar' from 'depth grammar'. What immediately impresses itself upon us about the use of a word is the way it is used in the construction of the sentence, the part of its use — one might say — that can be taken by the ear. (*PhI*, § 664)

However, there are aspects of language use that cannot be perceived either through sight or ear, aspects we must be acquainted with if words and sentences have to make sense to us, namely, those circumstances in which it is correct to use words and sentences, the consequences that follow from using them here and there, and so on. Mere sentence form is not enough to turn a string of words into a useful tool; put briefly, to have learnt the right technique of using it. Therefore, having access to a sentence syntactic structure, even to its logical form, and to the meaning of its constituents falls short of knowing its depth grammar.

> It is not every sentence-like formation that we know how to do something with, not every technique has an application in our life; and when we are tempted in philosophy to count some quite useless thing as a proposition, that is often because we have not considered its application sufficiently. (*IF*, § 521)

Sentences (2) and (3), as well as 'that the execution of Charles I is in the crater of Vesuvius' are word formations devoid of the necessary technique of use. We do not know what they are good for, what to do with them.[16] Their being somewhat meaningful does not meet their being able to express a thought.

If the gap between meaning and thought must be acknowledged, then the theory of meaning and the theory of thought aim at different targets.[17],[18] The discussion carried out in §§ 3 — 4 belongs to the theory of thought and can now be rephrased in a manner — backed by MRE — that makes it clear what endeavour it is part and parcel of. Theory of meaning, and grammar generally, we might say as a short summary, lays down on phrase and sentence constituents less demanding and more subtle constraints than those the theory of thought enforces on any concept to be exportable to a predicable.

---

[16]. This is the apt way Diamond describes the situation. See her Diamond (2000: 69).

[17]. Deflationist theories of truth, among which I include Disquotationalist theories of truth, deal with the sort of linguistic abilities Wittgenstgein would comprise under the label of 'surface grammar'. In a typical statement of Deflationism, as that of Field (1994), semantic theory aims at giving the truth-conditions of any sentence belonging to the language the semantic theory is a theory of. What confers Deflationism such an extremely superficial character, in the Wittgensteinian sense, to meaning analysis is that, according to such a semantic approach, meaning theory aims at theorems of the form «'p' is true if, and only if, π», where «'p' is true» and «π» are cognitively synonyms. Since it can be argued that, let's say, (2) does not have substantial truth-conditions — because it expresses no thought, Deflationism is forced to distinguish surface truth-conditions from deep (or substantial) truth-conditions. Field's Deflationism could only admit of surface truth-conditions.

[18]. Nowadays philosophy of language hardly leaves room for distinguishing the theory of meaning from the theory of thought — a situation to which it is not alien the fact that the question of nonsense has disappeared from both philosophy and grammatical theory. In my opinion, such a distinction echoes another one, this time between nonsense that has a sense from nonsense without any sense, to which Carnap gave his approval, but which Wittgenstein rejected (cf. Whitherspoon 2000). However, to those who favour the Wittgensteinian attitude we can reproach their not being sensitive enough towards the fact that sentences (2) — (6) as well as many others that matter to philosophy (like 'colours are mind sensations' or 'this piece of marble has extension') are meaningful, in some sense of the word 'meaning'. A possible exception to this situation is Glock, who is aware that the demands of formal logic openly differ from what philosophical grammar requires (see Glock 1996: 222). Whitherspoon (2000: § 5) contains the most elaborated defence I know of the significance that sentences like those discussed in this paper have, a defence in which the notions of quasi-interpretation, quasi-understanding and quasi-meaning, among others, play a central role.

5 Epistemic privacy and conceptual exportability

The casual comments made on sentences (4) and (5) have got round what there is in them that matters. I have restricted myself to saying that they serve to reiterate those conclusions obtained from the analysis of (2) and (3). However, both (4) and (5) belong to the class of sentences the debate on the Private Language Argument (PLA) has centred around. In particular, (4) has been at the heart of attention. My purpose is to make plausible the thesis that one of Wittgenstein's argumentative lines in his PLA is to argue that (4) expresses no thought, if understood in the way the advocate of epistemic privacy intends. Thus, the attack on the Epistemic Privacy Way Wittgenstein mounts claims that thinking that only I can know whether I am in pain is an illusion due to overlooking those specific paradigmatic restrictions the expression 'I know' is subject to. As in the case of (2) and (3), MRE lies at the bottom of the illusion. It follows from it that to understand the meaning of an expression as its use is the central finding we become aware of at the end of the Epistemic Privacy Way. I will elaborate on this by deploying exegetical evidence. Anyone who embraces Wittgenstein's criticism will see why sentences (4) and (5) are on the list in § 1, as vehicles of apparent thoughts. Since I have just distinguished meaning from thought, there is no alternative left but to accept that (4) and (5) are meaningful sentences, thus feeding the philosophical illusion that sensations are private objects on a parade with no more than one watcher. I hope to have shown in detail how this illusion's roots can be pulled up.

Wittgenstgein's PLA is a group of arguments that from different stances discuss and display the baselessness of maintaining that private languages are possible. By the private language of a subject *S* it is meant a language whose words denote *S*'s private experiences, experiences nobody except *S* can be acquainted with in principle. As against this conceptual possibility — firmly entrenched in the philosophical modern and contemporary tradition — Wittgenstein essayed a number of arguments. It has been pointed out that the main argumentative line in PLA elaborates on the remark that to name anything, i.e. a sensation of pain, requires to competently follow complex and demanding techniques of name using; and that though use techniques share a family air, there being no essence of the naming relation, to name an experience is not a *sui generis* kind of naming. It seems to follow from this that the possibility of providing a private something, i.e. a sensation, with a name cannot be discounted. This amounts to acknowledging that in 'naming a private object', the adjective 'private' has to mean 'incidentally hidden to other people', not 'logical or conceptually impossible to be known by anyone else'.[19] Among the rest of PLA's argumentative lines there is one, the Epistemic Privacy Way, that concludes that sentences like (4) and (5) do not express the kind of thought the philosophers Wittgenstein confronts would like to speak aloud. My contention is that Wittgenstein's run along Epistemic Privacy Way can be reconstructed by making use of the kind of analysis I have put forward in §§ 2-3 and the subsequent interpretation of the distinction between surface and depth grammar. The way PLA relates to the Frege-Wittgenstein-Diamond view of nonsense is as I have just stated.[20]

---

[19]. Concerning the criss-crossing of arguments in PLA, I follow Hacker (1990: 22&ff.). On what PLA's central argumentative line is, I follow Stern (1994).

[20]. As far as this relation is concerned, I only know two precedents. The most direct one is Groenendijk and Stockhoff (2004), where the Epistemic Privacy Way is approached to in outline from the perspective of the Compositionality Principle. On the other hand, in the interesting discussion on Moore's Paradox in Schulte (1993: ch. 9) I appreciate the presence of the Meaning Free Exportability Requirement (see § 2 above) under the guise of a Uniformity Principle supposedly governing the use of the verb 'to believe'. «The advocate of uniformity», Schulte

This is how Wittgenstein, in *PhI*, § 246, states the argument I am about to reconstruct:

> In what sense are my sensations *private*? — Well, only I can know whether I am really in pain; another person can only surmise it. — In one way this is wrong, and in another nonsense. If we are using the word 'to know' as it is normally used (and how else are we to use it?), then other people very often know when I am in pain. — Yes, but all the same not with the certainty with which I know it myself! — It can't be said of me at all (except perhaps as a joke) that I know I am in pain. What is it supposed to mean — except perhaps that I *am* in pain?
>
> Other people cannot be said to learn of my sensations *only* from my behaviour, — for *I* cannot be said to learn them. I *have* them.
>
> The truth is: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself.

The observations made in this remarkable passage close off a possible escape for the believer in the possibility of a private language might wish to follow. As is manifest, Wittgenstein faces the doctrine that our sensations are private, i.e. *sui generis* entities we have an experience of, everyone their own, *in foro interno*, entities whose existence is certain for the one who perceives them and to whom he or she can refer through acts of inner reference, but which no one else can know what they are like. I have already pointed out that PLA's main line objects to such a possibility that private sensations — in the sense just fixed — are not suitable for bringing the techniques of using name expressions to bear. The partisan of the Epistemic Privacy Way might now reply by insisting on the fact that each and every of us maintains an epistemic link, an immediate and incorrigible one, to our own experiences. It is this link that warrants the truth of the thought expressed by our uttering (4). While only you can know whether you are in pain, the rest of us can guess at it, conjecture it or whatever, but not being certain of it. Well, in the above passage Wittgenstein rejects two things about how (4) could be used. On the one hand, he rejects the idea that in using 'to know' as people commonly do, any utterance of (4) expresses a false thought, since other people know whether I am in pain or not. On the other hand, the one that truly matters here, he insists that if a different use of 'to know' is intended, nonsense is what one gets. Not only (4) do not thus express any thought, but (12) not (13), unlike (15) and (16), are in the same bag:

(4)   I know that I am in pain (alternatively, Only I can know whether I am in pain)

(12)   I doubt whether I am in pain

(13)   I learn that I am in pain

(14)   I am in pain

(15)   I know that she is in pain

(16)   I doubt whether she is in pain

These judgements found on reasons, provided in the last part of *PhI*, § 246(a) as well as in *PhI*, § 246(b,c), that can be resumed in the point that where no room is left for

---

writes, «makes the mistake of expecting all expressions to behave in the same uniform manner and, if he encounters unexpected forms of behaviour, to avoid facing up to undesirable characteristic of our use of linguistic expressions by side-stepping and appealing to misleading similarities. He is tempted by surface analogies to assume that there are corresponding analogies of function and meaning. And he tends to overlook that surface analogies may hide completely different ways of functioning» (Schulte 1993: 150). As for the case I am concerned with in this paper, the privacy of sensation is but an outcome of yielding to one of those misleading similarities. My aim here is to bring out the details of the confusion.

distinguishing between to know that *p* and to have reasons either to have reasons to think that *p*, to doubt whether *p* or to learn that *p*, to resort to 'to know' gives rise to nonsense. Moreover, it is not just the contrast between 'to know', on the one hand, and 'to have reasons to think', 'to doubt whether' and 'to learn' that matters. To the very same contrast class belong 'to believe (in some of its uses)', 'to suspect' and 'to find out', among others.

> One says 'I know' where one can also say 'I believe' or 'I suspect'; where one can find out. (If you bring up against me the case of people's saying 'But I must know if I am in pain!', 'Only you can know what you feel', and similar things, you should consider the occasion and purpose of these phrases. 'War is war' is not an example of the law of identity, either. (*PhI* II, p. 221.)

Of course, this is not all. Wittgenstein openly admits that (4) has, or might have, another use, namely, the same (14) could be put to. Nonsense that results from uttering (4) in intending to claim that my sensations belong to the realm of my own acquaintance originates in ignoring that there is no movement in the language game that acknowledges the existence of a cognitive relationship between our pain experiences and ourselves. The basic error, which Wittgenstein focuses on in *PhI*, §§ 247 — 255, is due to mistaking the use of sentence (4): we believe we are in the presence of a cognitive thought, one that plays a central role in articulating the metaphysics of subjectivity, while the truth is that (4) does nothing but expressing a grammatical rule. The metaphysical emphasis laid on 'I know' fades away.[21] Put it in more systematic terms, that (4) is an apparent thought is the conclusion of the following argument:

> [P₁]   For every proposition *p*: I express/conceive a thought on uttering/saying silently to myself that I know that *p*, if I express/conceive a thought on uttering/saying silently to myself that I can learn *p*

> [P₂]   I do not express/conceive any thought on uttering/saying silently to myself that I can learn that I am in pain

> [C]   I do not express/conceive a thought on uttering/saying silently to myself that I know that I am in pain

Premise [P₁] captures the fact that 'I know' belongs to a paradigm or contrast class which 'I believe', 'I have got reasons to think', 'I surmise' and 'I can find out' also belong to. The paradigm member Wittgenstein explicitly mentions in *PhI*, § 246(c) is 'I can learn'. Of course, the remark that 'I know' and 'I can learn' belong to the same paradigm far from bending the usual meaning or use of 'I know' it is fully faithful to it.[22] The existence of paradigmatic relations between different expressions or conceptual resources — relations referred to in [P₁] — is an aspect of thought that Wittgenstein thoroughly exploits in arguing that having a meaning and expressing a thought are not equivalent properties.

On the other hand, premise [P₂] completes the argument by pointing out the impossibility of importing 'I can learn' to the following sentential context:

> (ç 13)   ___ that I am in pain

In other words, (ç 13) is not an appropriate context for 'I can learn'. It would then follow that 'I know' cannot be exported to (ç 13) as well. This is why (4) is nonsense in one of its uses. This amounts to arriving at the end of PLA's Epistemic Privacy Way. However, there remains the need for answering a couple of questions to finish reconstructing the argument. The first question is why Wittgenstein specifically focuses on that particular paradigm item, i.e. 'I can learn', compares it with 'I know' and concedes it such an emblematic role. The answer: Other items, like 'I have found out' and 'I surmise', would have suited equally well. What really matters is that, drawn by philosophical prejudices, it may cost us dear to appreciate the particular relation between 'I know' and (ç 13). The second question is more difficult to answer: Why cannot 'I can learn' be exported to (ç 13). Without trying to exhaustively answer this question — since many philosophers are familiar with at least a significant part of this realm, three different argumentative lines can be adduced to explain why MRE rules out this possibility:

a) There is no way of mounting 'I know' on (ç 13), thus succeeding to make up the vehicle of a thought, because (14) does not provide the traditional philosopher with the needed propositional object such a thought should have. This is due to the fact that (14) does not work as a piece of representational or intentional system, but as a piece of an expressive system, i.e. (14) is used to moan. It is an avowal, like 'It hurts'. «[W]ords» — suggests Wittgenstein as a possibility — «are connected with the primitive, the natural, expressions of the sensation and used in their place» (*PhI*, § 244(a)). Now, what focusing on the use of 'I know' brings out is that these words cannot have a role to play both in the expressive and in the intentional dimensions. Unlike other expressions that belong to the same paradigm, i.e. 'I believe', 'I know' is devoid of such a potential. (See *OC*, § 180.)

b) Leaving now aside what has just been put forward in *a*), let's seriously consider whether 'I know' has an intentional or representational function — not only an expressive one — which lends itself to being used to express a thought. A language game in which it would be proper to use 'I can learn', one that also allows 'I know' to be applied, requires from the language user the possibility of taking on initiatives to improve his or her epistemic backing with respect to what the language user might get to know. To improve one's epistemic guarantees about *p* will often consist of either stocking up with reasons for *p* or correcting and completing those reasons for *p* already gathered. It seems clear, then, that the possibility of exporting 'I can learn' to context (ç 13) has to be dismissed, since the result of that move is the improvement of my epistemic backing by finding out reasons that support the conclusion that I am in pain — an apparent thought to all effects. Yet no such a possibility is in the offing. In what direction should one look for ways of improving one's epistemic stance? The most natural option seems to be the observation of our own behaviour, provided that keeping watch on other people's behaviour is what we resort to in trying to know whether they feel pain. Since this is an absurd choice, Wittgenstein concludes that «*I* cannot be said to learn of them» (*PhI*, § 246(c)). This, of course, does not bring the answer to an end, because the possibility remains open for any of us of knowing whether we feel pain without having to examine our own behaviour, namely, by checking inside oneself. However, Wittgenstein also blocks this way. What one should now adduce is that (14), either uttered in the open or silently said to oneself, expresses a thought whose truth none of us would know how to question. Does not pain afford the sort of *foundation* on which our certainty sustains? Do not we *know* then that we are in pain? No — and this is the final answer. The exportability of 'I can learn' to context (ç 13) is not underwritten by any successful research work. At the apparent truth of (14) «[w]e do not arrive […] as a result of investigation» (*OC*, § 138). The conjuring trick lies in

acknowledging that any reason or foundation for its truth would be less compelling than the very thought's truth whose support we could be looking for. 'I can learn' was chosen because 'I know' is a right movement to make in a language game in which the possibility of articulating, correcting and accumulating reasons must be built into from the very beginning. If the shaping, correcting and gathering of reasons do not provide us with the desired support, then the only conclusion left is that we are not in the knowledge business. Thus, the use of 'I know' is not a possible movement in such a language game. 'I know' announces that reason supplies are on offer:

> One says 'I know' when one is ready to give compelling grounds. 'I know' relates to a possibility of demonstrating the truth. Whatever someone knows something can come to light, assuming that he is convinced of it.

> But if what he believes is of such a kind that the grounds that he can give are not surer than his assertion, then he cannot say that he knows what he believes. (*OC*, § 243.)

c) One more option before closing. The one who utters (14), 'I am in pain', either speaking loudly or talking silently for him or herself, really wishes to say that he or she is aware of a feeling of pain, but awareness is now conceived as the phenomenal modality of consciousness. In such a case, (14) would be used to expresses the same thought as 'I feel I am in pain'. Nevertheless, no real, non-apparent thought would have thus been expressed. In this occasion I won't take any time over the ensuing argument (see Wittgenstein 1980: § 913).

6 Summing up

In this essay I have put forward and analyzed the view on nonsense Cora Diamond has carried out starting from the work of Frege and Wittgenstein. From Diamond's views I have derived a condition, the Meaning Restricted Exportability Constraint (MRE), and subsequently applied it to a kind of nonsense — the blind spot thought variety — which is found in one of the arguing lines that constitute Wittgenstein's so-called Private Language Argument. According to the reconstruction of such a line, the Epistemic Privacy Way, Wittgenstein would have made MRE a strategic function by pointing out that the advocate of epistemic privacy overlooks what constraints govern the paradigmatic exportability relations of expressions like 'I know' and 'I can learn', among others.[23]

REFERENCES

Baker, G. P. y Hacker, P. M. S. (1980): *Wittgenstein: Understanding and Meaning*. Oxford: Basil Blackwell.

Diamond, C. (1990): *The Realistic Spirit*. Cambridge, MA: The M.I.T. Press.

Diamond, C. (2000): ««What Time Is It on the Sun?»» *Harvard Review of Philosophy*, **VIII**: 69-81.

M. Dummett (1981a): *Frege. Philosophy of Language*, 2nd. edition. London: Duckworth.

M. Dummett (1981b): «The Context Principle». En *The Interpretation of Frege's Philosophy*. Cambridge, MA: Harvard University Press.

Field, H. (1994): «Deflationist Views of Meaning and Content». *Mind*, **CIII**: 249-85.

Glock, H. (1996): «Abusing Use». *Dialectica*, **50**: 205-23.

Groenendijk, J. y Stokhof, M. (2004): «Why Compositionality?» In G. N. Carlson y J. F. Pelletier, eds.: *The Partee Effect*. Stanford, CA: CSLI Publications.

Hale, B. (1997): «*Grundlagen* § 64». *Proceedings of the Aristotelian Society*, **IIIC**: 243-61.

Hacker, P. M. S. (1990): *Wittgenstein: Meaning and Mind*. Oxford: Basil Blackwell.

Heil, J. (1981): «Does Cognitive Psychology Rest on a Mistake?», *Mind*, **XC**: 321-342.

Hertzberg, L. (2001): «The Sense is Where You Find It». In T. McCarthy & S. Stidd, eds.: *Wittgenstein in America*, Oxford: Clarendon Press.

King, J. y Lee, D., eds. (1980): *Wittgenstein's Lectures. Cambridge, 1930-1932*. Oxford: Basil Blackwell.

Kripke, S. (1982): *Wittgenstein on Rules and Private Language*. Oxford: Basil Blackwell.

Lewis, D. (1992): «Meaning Without Use: Reply to Hawthorne», *Australasian Journal of Philosophy*, **70**, 106-110.

Lyons, J. (1968): *Introduction to Theoretical Linguistics*. Cambridge University Press.

McDonough, R. (1989): «Towards a Non-Mechanical View of Meaning», *Mind*, **XCVIII**: 1-23.

Robbins, P. (2002): «How to Blunt the Sword of Compositionality». *Noûs*, **36**: 313-34.

Schulte, J. (1993): *Experience and Expression.* Oxford: Basil Blackwell.

Stern, D. G. (1994): «A New Exposition of the 'Private Language Argument': Wittgenstein's Notes for the «Philosophical Lecture»». *Philosophical Investigations*, **17**: 203-26.

Szabó, Z. Gendler (2000): «Compositionality as Supervenience». Linguistics and Philosophy, **53**: 475-505.

Witherspoon, E. (2000): «Conceptions of Nonsense in Carnap and Wittgenstein». In A. Crary y R. Read, eds.: *The New Wittgenstein*. London: Routledge.

Wittgenstein, L. (1953/2001): *Philosophical Investigations [PhI]*. Revised edition. Oxford: Blackwell Publishing.

Wittgenstein, L. (1969): *On Certainty [OC]*. Oxford: Basil Blackwell.

Wittgenstein, L. (1975): *Philosophical Remarks*. Oxford: Basil Blackwell.

Wittgenstein, L. (1980): *Remarks on the Philosophy of Psychology*, vol. I. Oxford: Basil Blackwell.

Wright, C. (1989). «Wittgenstein's Rule Considerations and the Central Project of Theoretical Linguistics». In *Reflections on Chomsky*, ed. by A. George. Oxford: Basil Blackwell.

———————————————————

Juan José Acero
Department of Philosophy
University of Granada
E-18011 Granada, Spain
<acero@ugr.es>

AN ARGUMENT AGAINST EXTERNAL REASONS

Jonathan Anomaly

_____

With the publication of «Internal and External Reasons» (1981) Bernard Williams breathed new life into an old debate about normative practical reasons. The question to which internalists answer 'yes' and externalists answer 'no' is deceptively simple: do all reasons for action depend on desires?[1] As stated, the question is doubly ambiguous. First, according to some, the question at issue is whether the *judgment* by agent 'A' that she has a reason to do action 'φ' entails a desire to act on that perceived reason.[2] Others argue, following Williams, that the issue concerns not the relation between desires and *judgments* about reasons, but the relation between desires and the *existence* of reasons.

The second ambiguity concerns the entailment relation in the generic internalist slogan that 'reasons entail desires'.[3] The entailment relation is ambiguous between whether, necessarily, reasons give rise to desires or, necessarily, reasons owe their existence to antecedent desires. In the first case the desire is explained by the existence of a reason; in the second case the reason is explained by the existence of a desire. *Existence* internalists such as Williams lay claim to the second view.[4]

On this interpretation, which I call constitutive existence internalism (CEI), practical reasons are constitutively tied to the contents of an agent's current desire set. Although he is often misinterpreted, Williams' endorsement of CEI is clear: «If A has a reason to Φ, then … there must be a sound deliberative route to Φ-ing which starts from A's existing motivations» (1995: 187).[5]

Externalism is the denial of this conditional. Consider an example. According to externalists, an instrumentally rational psychopath may have a reason to treat all human beings

---

[1]. 'Desire' is a philosophical term of art. Williams uses 'desire' and 'motivation' interchangeably, and includes under their rubric «dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects…embodying commitments of the agent» (1981: 105).

[2]. Korsgaard's 'internalism requirement' (1986) can be interpreted as a form of judgment internalism.

[3]. On this ambiguity see Hurley (2001).

[4]. I borrow the distinction between 'judgment' internalism and 'existence' internalism from Darwall (1992).

[5]. Williams clearly presupposes a 'procedural' rather than 'substantive' conception of rational deliberation. On this distinction see Hooker (1987) and Parfit (1997).

with respect even though he lacks the capacity to be moved by moral considerations.[6] In other words, some externalists claim that from the fact that a psychopath lacks the capacity to see other human beings as more than mere objects, it does not follow that he has no reason to treat others with respect.

I shall now argue that all such externalist claims are incompatible with a plausible version of the ought-implies-can principle (OIC). The intuitive case for OIC is straightforward: if we reject OIC, we must accept the existence of obligations that cannot be fulfilled. Although such obligations may be found in ordinary moral thought and discourse,[7] they make little sense in an account of normative practical reasons. To suggest, for example, that a quadriplegic has a reason to run the Boston Marathon — that he *ought* to run the race despite his inability to do so — strikes us as absurd. Such a suggestion seems, at best, an expression of a wish that things were different than they actually are and, at worst, evidence of an incoherent belief — namely, the belief that someone might act in a way that is physically impossible.

Before deploying OIC in an argument against externalism, its constituent parts must be parsed. For the purposes of the following argument, I take the *ought* component of *ought-implies-can* to apply to normative practical reasons[8], the *implication* relation to be one of conceptual entailment[9], and the *can* component to indicate motivational possibility (since CEI primarily concerns the relation between motivations and practical reasons).[10]

Many externalists would claim that a psychopath — such as Jeffrey Dahmer — has a reason to treat people with respect, despite his inability to be motivated by moral considerations. Externalists are thus committed to an argument of the following sort:

1) J.D. ought to φ.

2) J.D. cannot φ.

3) It is impossible for one to do what one cannot do.

4) It is impossible for J.D. to φ.

——— ∴ ———

5) J.D. ought to do something that he cannot do.

As the conclusion suggests, externalists are committed to the claim that there are cases in which people ought to act in ways that are impossible, given their motivational makeup or

---

[6]. Whether we think of the capacity the rational psychopath lacks as *psychological* or *motivational*, the point remains the same. Externalists deny that any such idiosyncratic facts limit either the scope of reasons or the class of people to whom reasons apply. See Scanlon (1998: 363-73).

[7]. On this, see Sinnott-Armstrong (1984: 260) and Fischer (2003: 248).

[8]. As Parfit argues (1997: 121), to say that we *ought* rationally to φ is just another way of saying that we have a *reason* to φ.

[9]. Although an excellent case can also be made for logical or linguistic entailment, I ignore these issues here.

[10]. Internalists about theoretical reasons, in contrast, would focus primarily on epistemic possibility, as well as the cost and availability of information relevant to belief formation.

psychological profile. Internalists are skeptical, and their tacit endorsement of OIC is a plausible explanation of their skepticism. If this is right, internalists might revise the above argument (using OIC as a premise) as follows:

1*) J.D. ought to φ.

2*) J.D. cannot φ.

3*) Ought implies can.

——— ∴ ———

4*) Either 1* or 2* is false.

5*) By assumption, 2* is true.

——— ∴ ———

6*) 1* is false (It is not the case that J.D. ought to φ).

This argument may lead us to reject all *strong* externalist claims — claims that fail to take into account the different capacities of people to be motivated by considerations that underlie reasons claims.[11] Perhaps, though, we can still accept *weak* externalist claims which apply only to people who can, in principle, be motivated by such considerations.[12]

However, if weak externalists are characterized as those who both deny internalism and affirm OIC, they are compelled by theoretical consistency to acknowledge other factors that further limit the ability of individual deliberators to be moved by considerations that underlie reasons claims. These include contingent limitations in the availability of information, the potentially prohibitive cost of gathering information, and cognitive limitations in one's ability to imagine alternatives or acquire beliefs that would help one solve a practical problem. But once weak externalists accept such factors as conceptual constraints on practical reasons, the battle is won. Externalism collapses into some form of internalism.[13]

Some internalists have defended their position by arguing that externalism fails to *explain* rational action, since external reasons claims do not refer to anything distinctive about an agent, such as her current beliefs, desires, or other psychological states.[14] But if the foregoing argument is correct, the decisive case against external reasons ultimately rests on the plausibility of the ought-implies-can principle. Unless we can find an argument that vitiates OIC as a theoretical constraint on practical reasons we are forced to reject externalism.

References

Darwall, S. 1992. «Internalism and Agency». *Philosophical Perspectives* 6: 155-74.

Fischer, J. 2003. «'Ought-implies-can', causal determinism and moral responsibility». *Analysis* 63.3: 244-50.

---

[11]. Considerations such as the fact that an action will harm X or help Y.

[12]. McDowell (1995) seems to endorse this position.

[13]. Scanlon (1998: 371-3) comes close to acknowledging this.

[14]. See especially Williams (1993).

Hooker, B. 1987. «Williams' argument against external reasons». *Analysis* 47: 42-44.

Hurley, S. 2001. «Reason and Motivation: the Wrong Distinction?» *Analysis* 61: 151-55.

Korsgaard, C. 1986. «Skepticism about Practical Reason». *The Journal of Philosophy*. LXXXIII (1): 5-25.

McDowell, John. «Might there be External Reasons?» Reprinted in *World, Mind, and Ethics*: *Essays for Bernard Williams*. Cambridge: CUP, 1995.

Parfit, D. 1997. «Reasons and Motivation». *Proceedings of the Aristotelian Society* Supp (71): 98-146.

Scanlon, T.M. 1998. «Williams on Internal and External Reasons». *What We Owe to Each Other*. Cambridge: HUP.

Sinnott-Armstrong, W. 1984. «'Ought' conversationally implies 'can'». *The Philosophical Review* XCIII: 249-59.

Williams, B. 1981. «Internal and External Reasons». Reprinted in *Moral Luck*. Cambridge University Press.

——— 1993. «Internal Reasons and the Obscurity of Blame». Reprinted in *Making Sense of Humanity*. Cambridge University Press.

——— 1995. «Replies». *World, Mind and Ethics*. Cambridge University Press.

————————————————————

Jonathan Anomaly
Tulane University
<Anomaly@cal.berkeley.edu>

JUDGING LIFE AND ITS VALUE

Brooke Alan Trisel

————————————————

We could live life — perhaps as animals do — without judging whether it is good or bad, meaningful or meaningless, or worthwhile or not. However, we do not. We make judgements about the value[1] of life, concluding that life has no value or great value, and then may attempt to convince each other that we have placed the proper value on life.

It is not just philosophers who perform these evaluations. Many people, whether they realize it or not, compare their own lives, or life in general, to their expectations or desires, and then render a judgement about whether life, as they have experienced and view it, measures up to the standard they have adopted to judge it. There is a wide range of sophistication with these evaluations and judgements. At one extreme are simple, commonly uttered statements such as «life is good» or «life is terrible,» where it may be unclear how the person reached the conclusion they did. At the other extreme are the differing judgements of philosophers — often carefully thought out and based on elaborate arguments. Between these two extremes are the evaluations and judgements of poets and playwrights, including Shakespeare's famous words that life is «full of sound and fury, Signifying nothing.»[2]

Different methods and standards have been and can be used to judge the value of life. Furthermore, our judgements about whether or not life has value depend, in large part, on which method is used, as will be shown. Therefore, it is important that attention be given to examining the methods used to make this judgement.

There has been a great deal of work done by normative theorists in evaluating various methods, such as utilitarianism, for helping us to decide among different courses of action. In contrast, there has been little attention paid to examining the methods used to make judgements about the value of life. Rather, attention has been concentrated, for many years, on assessing whether life can be meaningful without God and immortality. In this essay, I will explore and compare the various ways of judging whether life has value with the ultimate aim of identifying the best method for making this judgement.

Purposefulness Versus Worthwhileness

When people first begin to reflect on philosophic questions, they often start with purpose related questions such as «What is the purpose of life?» and «Why are we here?» After reflecting on these questions for a while, some people eventually realize that these questions presuppose that life was created for a reason and so they then strip the question of

————————————————

[1]. Value will be used to mean worth.

[2]. William Shakespeare, *Macbeth*, 5.5.27-28 (Arden edition), ed. Kenneth Muir (London: Methuen, 1951), 154.

its assumption and ask themselves «Is there a purpose of life?» Although people may ponder the famed question about the meaning of life, many of those who do never ask themselves the distinct question «Is life worth living?» which is unfortunate for reasons I will explain.

Based on the amount of literature on the question «Is there a meaning or purpose of life?»[3] relative to the question «Is life worth living?»[4] one might be led to believe that the purpose related question is more important. However, as I will attempt to demonstrate, I believe that the reverse is true.

In disputing the claim made by some that life is not worth living if it has no meaning, Kurt Baier convincingly argues that one can find life worth living even if one's life is not «meaningful,» regardless of whether meaningfulness is thought of as a «hidden,» universal meaning or simply as valuable or important. Referring to the second sense of meaningfulness, he writes: «Now, many lives are not valuable or important and are not, in this sense, meaningful. But, again, this need not imply that they are not worth living.»[5] The opposite is also true. A person may lead a successful and purposeful life and yet may find life not worth living and may even commit suicide.

Albert Camus had earlier argued a similar point to Baier when he writes: «Hitherto … people have played on words and pretended to believe that refusing to grant a meaning to life necessarily leads to declaring that it is not worth living. In truth, there is no necessary common measure between these two judgements.»[6] However, Camus, earlier in the same essay, makes a common mistake of failing to distinguish between these questions, which leads him to give undue importance to the question about the meaning of life. In a famous statement, he indicates:

> 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. … I have never seen anyone die for the ontological argument. … On the other hand, I see many people die because they judge that life is not worth living. … I therefore conclude that the meaning of life is the most urgent of questions.[7]

I believe that Camus was correct in asserting that judging whether life is or is not worth living is the most important philosophic question. However, he is incorrect to conclude that the meaning of life is the most urgent of questions.

---

3. For a discussion and bibliography of recent (since 1980) works on the meaning of life, see Thaddeus Metz, «Recent Work on the Meaning of Life,» *Ethics* 112 (July 2002): 781-814. Earlier works on the meaning of life can be found in E.D. Klemke, *The Meaning of Life*, ed. E.D. Klemke (Oxford: Oxford University Press, 2000).

4. For discussion on the question of whether life is worth living, see Kurt Baier, «Threats of Futility: Is Life Worth Living?» *Free Inquiry*, 8 (Summer 1988): 47-52; Kurt Baier, *Problems of Life and Death: A Humanist Perspective* (Amherst, N.Y.: Prometheus Books, 1997), 67-70; Richard Wollheim, *The Thread of Life* (London: Yale University Press, 1984), 244-48; John McDermott, «Why Bother: Is Life Worth Living?» *Journal of Philosophy* 88 (1991): 677-83; and Karsten Harries, «Questioning the Question of the Worth of Life,» *Journal of Philosophy* 88 (1991): 684-90.

5. Kurt Baier, «Threats of Futility,» 49.

6. Albert Camus, *The Myth of Sisyphus and Other Essays*, trans. Justin O'Brien (New York: Vintage Books, 1955), 6.

7. Camus, 3-4.

Regarding the questions «Is there a purpose of life?» and «Is life worth living?» the fact that a person may answer one question negatively, and the other question affirmatively, demonstrates that these two questions are different. Some have explicitly noted that these are distinct concepts. For example, Thaddeus Metz writes: «The concept of what makes a life worth living is distinct from the concept of what makes a life meaningful.»[8]

Although some philosophers have noted that there is a difference between a «meaningful life» and a «life worth living,» what it is that makes these concepts different has not been explored. How do these concepts differ? The principal difference between these concepts is that the latter concept has a much *broader* scope than the former concept. One who asks about the purpose of life generally seeks to know whether we were created for a purpose or have some role to play in carrying out a divine or cosmic purpose or plan. It is a narrow question about goal-directed activity. In contrast, the question regarding whether life is worth living is a broad question. Addressing this question has often been thought of as a process of weighing benefits against costs, good against bad, or pleasure against pain[9] to determine if, in one's life, the former outweighs the latter. Hereinafter, the terminology of benefits and costs will be used.

Granted, one benefit of being alive is the satisfaction obtained from having and pursuing goals or purposes. However, there are other benefits associated with life, including interacting with family and friends, music and art appreciation, and adventure, that may have little or nothing to do with goal-directed or achievement-oriented activity. If we ask ourselves the question «Is life worth living?» these other benefits would be taken into account when judging life. But this will not occur if we only reflect on the question «Is there a purpose of life?» and never ask ourselves the more encompassing and important question «Is life worth living?»

An evaluation of whether life is worth living takes into account, not only whether one has a «purpose» or «purposes,» in whatever way one interprets these words, but other experiences that do not involve pursuing goals. Because an evaluation of worthwhileness is more comprehensive than an evaluation of purposefulness, this explains how it is possible that one could live a life that is not goal-driven or purposeful and yet find life worth living or could have a purposeful life and yet find life not worth living.

In a cross cultural study, the anthropologist Gordon Mathews asked various people in Japan and the United States the question of what makes life worth living.[10] Mathews indicates that there is a term used by Japanese called «ikigai» which means «that which most makes one's life seem worth living.»[11] When he asked the Japanese participants the question of what it is that most makes them feel that life is worth living, they responded with answers such as family, work, and leisure activities. Because there are few references in the print media in the United States to the phrase «life worth living,» but many references to the phrase

---

[8]. Metz, «Recent Work,» from footnote on p. 788.

[9]. See, for example, Wollheim, «The Thread of Life,» 244.

[10]. Gordon Mathews, *What Makes Life Worth Living: How Japanese and Americans Make Sense of Their Worlds* (Berkeley: University of California Press, 1996).

[11]. Ibid., 5.

«meaning of life,» and because there is no cultural equivalent to the term «ikigai» in the United States, he initially asked American participants the following question: «What, for you, is the meaning of your life?» Mathews indicates that the participants in the study were perplexed by the question and were unable to answer it. However, when Mathews later asked them the question «What most makes your life seem worth living?» he indicates that the «Americans responded to these questions almost exactly as the Japanese responded to «What's your *ikigai*?»[12] This study supports two points that are being made in this essay. First, questions about whether life is worth living are intelligible whereas questions about the meaning of life are not. Second, when people are asked whether life is worth living, they will take into account benefits that are disregarded when they reflect on whether there is a purpose of life.

«Life,» as it is used in the questions «Is there a purpose of life?» and «Is life worth living?» could refer to life in general or to our individual lives. Therefore, when discussing these questions, it is necessary to define «life» to avoid misunderstanding. In this essay the word «life,» in the question «Is life worth living?» will refer to our individual lives. This is the important question. The question of whether life in general is worth living is peculiar and almost as unclear as the question «Is there a purpose of life?» — which is notorious for being obscure.

Of course, life in general does not live. It is individuals who live and so it is individuals, not life in general, who may or may not find life worth living. Thus, the question «Is life (in general) worth living?» is unclear. With the question, one could be asking whether the life of every individual who makes up the human species is worth living. One might also be asking whether the life of a typical or average person is worth living, even though no one individual may actually have the characteristics of this imaginary individual.

Judging whether life in general is worth living has no practical value. If one decides that one's own life is not worth living, one can act upon this judgement by attempting to engage in certain activities or form relationships with other people that will make life worthwhile.[13] One may also act upon this judgement by ending one's life. However, if a person concludes that life in general is not worth living, there is not much that this person can do with this judgement, except perhaps for engaging in a futile attempt to try and convince the entire population of the world to commit mass suicide.

The question «Is there a purpose of life?» has two possible responses: yes and no. Neither of these responses says anything about the value of our lives. However, some venture beyond addressing the question and claim that if there is no purpose of life that life has little or no value. For example, William Lane Craig asserts: «Without God the universe is the result of a cosmic accident, a chance explosion. There is no reason for which it exists. As for man, he is a freak of nature — a blind product of matter plus time plus chance. Man is just a lump

---

[12].   Ibid., 28.

[13].   I do not believe that there is a difference between finding life worth living and finding life worthwhile. Wollheim («The Thread of Life,» 244) uses «worthwhile» to mean satisfying one's desires and plans, whereas most people would use the term «meaningful» or «purposeful» instead. Because he uses «worthwhile» synonymously with «meaningful,» he concludes that there is a distinction between finding life worth living and finding life worthwhile. Although there is a difference between finding life meaningful and finding life worth living, if he had used «worthwhile» in the conventional sense to mean worth the effort, then he likely would not have concluded that there is a difference between finding life worth living and finding life worthwhile.

of slime that evolved into rationality.»[14] Craig and others who make similar arguments assume that there is a relation between whether there is a purpose of life and whether our lives have value. In claiming that life has no «real» significance or value if it has no purpose, they have, in effect, adopted a purpose-based standard to judge whether life has value.

What criterion or standard should be used to judge whether our lives have value? Should it be a purpose-based standard or a worthwhileness standard? The two standards are as follows:

(a) My life has value only if human life has a purpose.

(b) My life has value only if the benefits of being alive outweigh the costs.

The first standard will be referred to as the «cosmic standard.»

If we are going to judge life, then the standard that we use for making this judgement should be clear and take into account a broad range, if not all, of our experiences. As I have argued, the purpose-based standard fails to take into account experiences that do not involve pursuing goals. Furthermore, the notion of «purpose,» as reflected in the claim that life has value only if it has a purpose, is unclear as many have argued. Because there are different types of purposes, this is one source of confusion with the claim. Having a purpose could mean that human life was created for a reason (to serve a function) by a god or an intelligent cosmos. Alternatively, having a purpose could mean that we were not created by a superior being to serve a function, but that we have self-chosen goals or that we some role, either self-chosen or externally imposed, in carrying out the goal(s) of a superior being.

Because the notion of having a purpose is obscure and the purpose-based standard disregards many of our experiences — some of which we may cherish the most — it is an inadequate standard to use to judge whether our lives have value. Furthermore, this standard is based on the assumption, as mentioned, that there is a direct relation between whether there is a divine or cosmic purpose and whether our lives have value, but is there such a relation? If the cosmos were intelligent and had a self-chosen purpose, such as to expand forever, but human beings were unintended and simply a byproduct of this expansion, then the fact that the cosmos has a purpose in no way adds value to our lives. Some have asked the question of whether we would be satisfied if we learned that we had a purpose, but that this purpose was to serve as a food source for another entity.[15] Of course the answer would be «no,» which provides further support that there is no direct relation between whether human life was created for a reason and whether our lives have value.

A now commonly used way of responding to those who maintain that life is purposeless without God and immortality, is to point out, as Baier[16] did in 1957, that there are two different senses of the word «purpose» and that one can create purpose in one's life, by pursuing certain goals, regardless of whether life was created for a purpose. In pointing out

---

[14]. William Lane Craig, «The Absurdity of Life Without God,» in *The Meaning of Life*, ed. E.D. Klemke (New York: Oxford University Press, 2000), 45. Originally published in *Reasonable Faith: Christian Truth and Apologies*, 1994.

[15]. See, for example, Robert Nozick, *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), 586.

[16]. Kurt Baier, «The Meaning of Life,» in *The Meaning of Life*, ed. E.D. Klemke (New York: Oxford University Press, 2000), 101-132. Inaugural Lecture delivered at Canberra University College, 1957.

this other sense of the word «purpose,» and by arguing that it is sufficient to make life meaningful, these philosophers indirectly propose an alternative, lower standard for judging whether life has value. For one's life to have value, there does not have to be a grand cosmic or divine purpose of which we play a central part. Instead, they suggest that one can create value in one's life if one has nontrivial, achievable goals and pursues them with passion. Although this latter standard is clear and much more reasonable than the cosmic standard, it fails, as does the cosmic standard, to take into account experiences that do not involve goal-directed activity — experiences that would be factored into a worthwhileness evaluation.

When the separate concepts of «worth» and «while» are combined into the word «worthwhile,» it creates a powerful and elegant concept. In using this concept, one takes into account benefits and effort, placing them in a side-by-side comparison so that one can determine whether the benefits justify the expenditure of effort. Because we arose in an imperfect natural world, where there is competition among the creatures of this world for limited resources, which can result in hunger, disease, and suffering, it is unrealistic to expect that there would only be benefits associated with living. The worthwhileness standard acknowledges that there are, and will always be, costs associated with living whereas the purpose-based standards do not. Therefore, the worthwhileness standard is a more realistic way of judging whether life has value than are the purpose-based, benefit-focused standards.

Methods for Determining Whether Life is Worth Living

Method 1: Weighing Benefits and Costs

Although the worthwhileness standard has significant advantages over the purpose-based standards for judging whether life has value, how can one determine whether one's life is worth living? One can attempt to weigh the benefits of being alive against the costs, as has been discussed. However, there are well-known limitations with a cost-benefit analysis, including it can be difficult to determine what counts as a benefit or a cost. It also can be difficult to quantify benefits and costs. Consequently, one may be unable to determine whether there is a net benefit associated with living one's life. Because of the limitations of a cost-benefit analysis, it is important to explore alternative methods for judging whether life is worth living.

Baier contends that addressing the question of whether life is worth living has generally been misconceived.[17] Instead of being the outcome of an investigation such as finding out whether a thing is worth its price, he asserts that one can determine whether one's life is worth living without having first determined whether the good things outweigh the bad. He suggests that one can infer whether one's life is worth living based on the following method: «I assume, without argument, that we can say a person really is finding his life worth living if, and only if, supposing it were up to him to live his life over again, exactly as it was, he would be prepared, or even glad or eager, to do so.»[18]

Others have used different inferential methods to determine if life is worth living. I will evaluate the various methods that have or could be used to make this determination. Because one might question whether we really can infer whether life is worth living, as suggested by Baier, I will explain why I believe this is possible before evaluating the methods.

---

[17]. Baier, «Threats of Futility,» 49.

[18]. Ibid., 49.

We are continually faced with decisions about how best to use the limited time that we have available as finite creatures. We could perform activities impulsively, starting and stopping them without reflecting on whether the activities are worthwhile, but generally we do not. We evaluate the worth of activities — before, during, and after we do them.

Before we commit ourselves to doing one activity of the many that we could be doing, we consider whether it will be worth our effort. And if we choose to begin the activity, we do not just keep doing it without giving it any further thought. Rather, we periodically question whether the activity continues to be worth the effort, and if it is found not to be worthwhile, then we cease performing the activity, unless we are forced to continue or fear what might occur if we stop. Finally, before we repeat an activity, we consider whether we found the original activity worthwhile. If we found that the original activity was a «waste of time,» then it would not be repeated. However, if we found the original activity worthwhile, then it may be repeated.

As explained above, we do not generally make decisions about whether to begin, continue, or repeat an activity impulsively or without any rational thought. Rather, we evaluate whether the activity is worth *beginning, continuing, or repeating*. Therefore, if we begin or continue an activity, it may be possible to infer from this that we found the activity worthwhile. And if we repeat an activity, then this suggests that we found the original activity worthwhile. But how can we tell if we find our individual lives, as a whole, worthwhile? Can we adopt the methods discussed above for inferring whether an activity is worthwhile and use them to infer whether our lives are worth living? As will be shown, this has been the approach that some people have taken.

We did not decide whether to be born into this world (a point emphasized by many existentialists),[19] nor can we repeat our lives. Consequently, we cannot infer whether we find life worth living based on these decisions. However, we can imagine whether we would have chosen to *begin* life and live our lives, if this choice had been ours to make. In addition, one can imagine whether one would, if it were possible, *repeat* one's life. Finally, it might be thought that we can infer whether life is worth living based on whether or not we choose to *continue* living, but can we? These methods will be considered below, in an attempt to identify the best method for judging whether life is worth living.

In a previous section, the two purpose-based methods and the worthwhileness method, and the standards derived from these methods, were evaluated. They were judged based on whether they were clear and took into account a broad range of our experiences. The five inferential methods that will be discussed (methods two through six) are all relatively clear and broad in scope. Two additional criteria will be considered in judging them, including: (1) whether, using the method, one can validly infer whether one's life is worth living and (2) whether the standard associated with the method is reasonable. The best method for judging if life is worth living will be revealed in the concluding section.

---

[19]. Martin Heidegger is perhaps the best known example, referring to our facticity, which would include the fact that we did not choose to be born, as has having been «thrown» into existence. See Heidegger, *Being and Time*, trans. John Macquarrie and Edward Robinson (New York: Harper & Row, 1962), 174, 321.

Method 2: Continuing to Live

The fact that pessimists continue to live their lives and engage in projects such as writing suggests, it might be argued, that they really do find life worth living despite their claims otherwise. But does the fact that one has not ended one's life demonstrate that one finds life worth living? Human beings do not die easily, as evidenced by the drastic measures that people use to commit suicide, such as leaping off tall buildings. People may continue living at a miserable level rather than ending their lives because of a fear of death, as Baier points out.[20] Schopenhauer makes a similar observation when he writes: «It will generally be found that where the terrors of life come to outweigh the terrors of death a man will put an end to his life. But the terrors of death offer considerable resistance: they stand like a sentinel at the exit gate.»[21] In other words, people may continue living, not because they find life worth living, but for other reasons — because they fear death and the pain that would come from killing themselves or perhaps because they hold out hope that, at some point, their lives may become worth living. If true, then we cannot validly infer that we find life worth living just because we continue to live our lives.

Suppose, however, that we were equipped with an inborn switch, whereby we could painlessly and irrevocably turn off life when we wished and we somehow knew for certain that death was nothing to fear. Could we then validly infer that we find our lives worth living if we do not flip the switch to end our lives? Even under these circumstances, we could not validly make this inference. Some may go on living in misery, rather than committing suicide, because they do not want to inflict suffering on people who care about them. Due to the interpersonal attachments they have formed, the only way that they may ever flip the switch to end their lives is if the switch had the power to make it as if they had never been born, which would be the only way of assuring that their nonexistence would hurt no one.

Method 3: Choosing to Live

Although we did not choose whether we would be born into this world, to determine whether life is worth living, one can imagine whether one would have chosen to live the life that one is living, taking into account everything that one has experienced. If one would have chosen to live one's life, then this suggests that one finds life worth living.

If we do not believe that an activity, such as attending a conference, will be worth the effort or if we think that undertaking the activity will make us miserable, then we avoid it. But we had no control over whether we would be born into this world and some may feel resentment or dismay over this. This method under discussion places one in a position to decide whether to live one's life. It allows one to judge whether living would be worthwhile and then to decide for oneself whether to remain nonexistent or to come into existence and live the life that one has known.

This method requires that one imagine that one is a nonexistent thinker who knows in advance what one's life will be like if one chooses to live life, and who has control over the decision whether to remain nonexistent or to be born into this world. Stretching one's imagination to this extent, and accepting the contradiction of a nonexistent thinker, may be

---

[20].   Baier, «Threats of Futility,» 49.

[21].   Arthur Schopenhauer, «On Suicide,» in *Essays and Aphorisms*, trans. R.J. Hollingdale (London: Penguin Books, 1970), 78-79.

difficult for some people to do. There is, however, a similar method of evaluating whether life is worth living that does not require that one engage in a thought experiment and imagine that one is a nonexistent thinker. With this method, you ask yourself whether you «wish that you had never been born.» When these words are uttered, it expresses a feeling of regret about being alive. For example, in the Old Testament, Job, who is apparently suffering greatly, professes: «Let the day perish in which I was born ….»[22] He continues: «Or why was I not buried like a stillborn child, like an infant that never sees the light? There the wicked cease from troubling, and there the weary are at rest.»[23] To give another example, in Goethe's *Faust*, Faust proclaims «I wish I had never been born!»[24]

With this method, in contrast to the prior method, one does not advance to the step of putting oneself in the imaginary position of deciding whether or not to be born into this world and live the life that one has experienced. However, the judgements reached using the two methods would likely be consistent. People who regret that they were born would also choose not to live life, but to remain nonexistent, if this decision had been theirs to make. And people, who are satisfied or grateful that they were born into this world, would also choose to live the life they have experienced rather than remaining nonexistent, if the choice had been theirs to make.

Although the judgements will generally be the same using either of the methods discussed above, putting oneself in the imaginary position of deciding whether to be born into this world will give a better indication of whether one finds life worth living than simply asking oneself whether one regrets that one was born. Because there is no easy way to cease living this life that we did not choose to live without hurting others, one may feel trapped in this world. To cope with this, one may try to convince oneself, in an act of self-deception, that one does not regret that one was born into this world, though deep down one may feel otherwise. By giving oneself control over whether or not to be born into this world, by forcing oneself to make this potentially difficult decision, and by providing one with a painless way to escape life (deciding to remain nonexistent) that will not make others suffer, this method better reveals whether one truly finds life worth living than the method of surveying one's life from the time of entry into this world to the present, but not going the extra step of deciding whether one would have chosen to enter the world in the first place.

Method 4: Choosing to Bring Others into the World

Although we had no control over whether we would be born into this world, we do decide whether to bring future generations into this world. It could be argued that, as creatures who are capable of feeling empathy for others, we would not bring others into this world unless we found our own lives worth living. Stated differently, if we are miserable, then, because of empathy, we would spare potential persons from having potentially miserable lives by not bringing them into existence. For these reasons, it might be thought that one could infer whether one's life is worth living based on whether one has brought or wants to bring children into the world. Is this a valid inference to make?

---

[22]. Job 3.3. New Revised Standard Version (Nashville: Thomas Nelson Publishers, 1989), 307.

[23]. Job., 3.16-17, 307.

[24]. Johann Wolfgang von Goethe, *Faust,* trans. Louis MacNeice (New York: Oxford University Press, 1961), 152.

The inference discussed above is based on the false premise that people make decisions about whether to have children in a rational manner. Granted, some people do take into account the state of the world and reflect on the value of life when deciding whether to have children. For example, occasionally someone will make a statement such as «I would not want to bring a child into this world.» However, nearly half of all pregnancies are unintended[25] and so, quite often, people give no thought at all to whether life is worth living before bringing a newborn into this world.

Another problem with this method is that some people have children to make their lives worth living. Thus, in this situation, the fact that they want to have children does not indicate that they find life worth living. On the contrary, it suggests that they currently do not find life worth living and are searching for something that will give meaning to their lives. For the preceding reasons, we cannot validly infer whether we find our lives worth living based on whether we have brought or want to bring others into this world.

Method 5: Wanting to Live Again and Again

Nietzsche attached great importance to eternal recurrence, the ancient idea that the history of the world, including each human life, recurs again and again without end. In introducing this idea, he writes:

> What, if some day or night a demon were to steal after you … and say to you: 'This life as you now live it and have lived it, you will have to live once more and innumerable times more; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything unutterably small or great in your life will have to return to you, all in the same succession and sequence ….' … Would you not throw yourself down and gnash your teeth … Or … would [you] have answered him: 'You are a god and never have I heard anything more divine.'[26]

Various interpretations have been given of Nietzsche's doctrine of eternal recurrence. Maudemarie Clark contends that Nietzsche was not advancing a cosmological theory or proposing the following practical directive about how to live our lives: live them in such a way that we would want to live them innumerable times. Rather, she suggests, as follows, that Nietzsche was proposing a test to judge our attitude toward our lives:

> I interpret Nietzsche as taking our hypothetical reaction to the demon's message — how we would react if we accepted the message uncritically — to reflect our actual attitude towards ourselves and our lives. A joyful reaction would indicate a fully affirmative attitude towards one's (presumably, nonrecurring) life, whereas gnashing of teeth … would indicate a negative attitude.[27]

To use eternal recurrence as a test of affirmation, Clark maintains we must «imagine eternal recurrence in an uncritical or preanalytical manner, suspending all doubts concerning its truth or conceivability.»[28]

---

[25]. Sarah Brown and Leon Eisenberg, ed., *The Best Intentions: Unintended Pregnancy and the Well-Being of Children and Families* (Washington: National Academy Press, 1995), 2. According to the study, 44% of all births in the United States in 1990 were unintended.

[26]. Friedrich Nietzsche, *The Gay Science*, trans. Walter Kaufmann (New York: Vintage Books), 273-74.

[27]. Maudemarie Clark, *Nietzsche on Truth and Philosophy* (Cambridge: Cambridge University Press, 1990), 251.

[28]. Ibid., 270.

If eternal recurrence is thought of as a possible method and standard for judging whether one finds life worth living, regardless of how Nietzsche thought of it, would it be a good method to use? With this method, in contrast to a cost-benefit analysis, there is no direct weighing of benefits and costs. Rather, Clark's interpretation suggests that one can infer whether one finds life worth living based on whether or not one would want to repeat one's same life an infinite number of times.

In the literature debating Nietzsche's idea of eternal recurrence, one main point of contention is whether there would be a continuity of consciousness and an accumulation of what we remember from one life to the next.[29] If one had no memory of one's prior[30] lives, and thus each life was independent of each other, then there would be no reason to be horrified with this idea as Nietzsche thought some people might be. Therefore, with this idea, Nietzsche must have envisioned that one would remember one's prior lives. Clark writes: «I can only imagine recurrence as it would appear to a continuing consciousness ….»[31]

For the sake of argument, let us suppose, not that we have lived innumerable times before, but that each of us has a first life and that it is the life we are currently living. Let us also suppose that our lives will repeat forever after our first life and that what we remember would accumulate from one life to the next. The first time we live life, we experience the novelty, adventure, and excitement associated with being alive. We look around and see brilliantly colored flowers, flowing rivers, and majestic mountains. We also experience costs associated with living such as suffering, pain, and having loved ones die.

The second time we live life, we have the same experiences of our first life. However, this time there is less novelty and excitement because we have experienced everything, in the exact same way, once before. The benefits have lessened, but there has not been an offsetting reduction in costs. If our mother and father died before us in our first life, then we experience their deaths for the second time. We may not suffer as much emotional distress with their second death, as with their first, if we have reason to believe that they will live again. However, we experience the same physical pain as we did in our first life. Indeed, the pain may be intensified if we realize that we will experience this pain, not just during one life, but each time that we live one of our countless lives. The third time we live life, there is even less novelty and excitement than the second time. As we continue to repeat life, there is less and

---

[29]. Like Clark, Julian Young interprets eternal recurrence as a test. However, Young's view of the test differs from Clark's in two ways. First, Young conceives of it, not as a test to judge one's attitude toward one's life, but to determine «whether one is living as successfully as possible.» Second, Young envisions that one would be unaware of one's prior lives with eternal recurrence. See Young, *The Death of God and the Meaning of Life*, (London, Routledge, 2003), 89-90.

[30]. For explanatory purposes the word «prior» will be used, but it is a point of contention whether one's various lives would occur at different times. If a person has an infinite number of lives, and this person's lives are the «same,» then one could argue that they must occur at the same time and thus there would be no «prior» lives. If there are an infinite number of lives, then in what respect do they differ? According to Clark («Nietzsche on Truth,» 267): «Only a difference in temporal position fits the bill — the recurrence of my life must be part of an earlier or later cycle of cosmic history than my present life ….»

[31]. Clark, «Nietzsche on Truth,» 269.

less benefit derived from living until we reach a point where all the novelty and excitement have been drained from life and we are left with nothing but drudgery, pain, and suffering.[32]

The following standards can be derived from methods three and five, respectively.

Standard 3: My life is worth living only if I would have chosen to live my life, if the choice had been mine to make.

Standard 5: My life is worth living only if I would want to live my same life over and over again without end.

Eternal recurrence is a much higher standard for judging whether life is worth living than the other worthwhileness standards that have been and will be discussed. People may have no regrets that they were born into this world, and if the choice had been theirs to make, may have chosen to live their lives. However, they may have no desire to live a second time, let alone an infinite number of times. Thus, if they judged life using standard three, they would find life worth living, but they would not find life worth living if they used the eternal recurrence standard to make this judgement.

The eternal recurrence standard for judging life is unreasonable. The following example shows how ridiculous this standard is: Suppose that you have just watched a movie with a friend and then ask your friend how they liked the movie. The friend replies that the movie «was not worthwhile.» You ask why and your friend replies «it was not worthwhile because I would not want to watch it an infinite number of times.» If this were the standard that this person uses to judge movies, then no movie would ever be considered worthwhile. We recognize from experience that we derive less and less benefit each additional time that we watch the same movie — an example of the well-known utilitarian and economic principle of diminishing marginal utility. We eventually reach a point where we are tired of the movie and may never want to watch it again. Because human beings abhor unending repetition, as reflected by the principle of diminishing marginal utility, it would be unreasonable to judge whether life is worth living based on whether or not one would want to repeat one's same life an infinite number of times.

To return to the movie example, if a movie is good we may want to watch it again. However, we do not judge a movie based on whether we would want to watch it repeatedly. Rather, we use other criteria such as whether the movie was exciting, funny, innovative, or moving. Alternatively, instead of judging the movie based on specific criteria, we may indirectly judge the movie based on an inferential standard, but this standard does not involve unending repetition. Rather, it typically involves reflecting on whether we felt satisfied that we decided to watch the movie in the first place. Because we never employ the idea of unending repetition to judge the worth of the activities in our lives, this should serve to caution us about using eternal recurrence to judge whether our lives, as a whole, are worth living.

The eternal recurrence standard for judging life has been found unreasonable. The second criterion that will be considered in evaluating this method is whether, using the method, one can validly infer whether one's life is worth living. As I argued earlier, if a

---

[32]. Arthur Danto advances a similar argument against Nietzsche's version of eternal recurrence. He argues that it would involve «eternal monotony.» See Danto, *Nietzsche as Philosopher* (New York: Columbia University Press, 1965), 210.

person repeats an activity, then we may be able to infer from this that the person found the original activity worthwhile. It is important, however, to understand what it means for us to «repeat» an activity. When an activity is repeated, it is performed a second time, but the second time is never the «same» as the first. It occurs at a different time and perhaps at a different place as, for example, when people go to see their favorite entertainer repeatedly. Also, the second activity may be performed with a different group of friends than the first activity.

If we want to repeat an activity, realizing that the second activity will always be at least slightly different from the original, we can infer from this that the original activity was worthwhile. It would provide an even stronger indication that we found the original activity worthwhile if it were possible to repeat the «same» activity and if we would want to do so despite knowing that, because it is the same activity, we will not experience as much novelty and excitement as with the original activity. Analogously, if one would want to repeat one's same life, even if one's subsequent life or lives would lack some of the features (excitement, for example) of one's current life, then this strongly suggests that one finds one's current life worth living.

Because the eternal recurrence test is so demanding, hardly anyone would pass it. However, for that rare person who would want to live his or her same life innumerable times, in spite of having diminishing benefits with each successive life, this person can feel confident that he or she does truly find life worth living.

As noted earlier, people may continue living, not because they find life worth living, but for other reasons, such as not wanting to hurt others by committing suicide. Thus, the continuation standard, where it is presumed that people find life worth living just because they have not ended their lives, would in many cases falsely indicate that people find life worth living. In other words, it would yield false positives. This is also a significant problem for method four (inferring that life is worth living based on whether one has brought or wants to bring others into the world).

There is a relation between the level and the validity of the various inferential standards. If the standard for judging whether life is worth living is very low, such as the continuation standard, then everyone who uses this standard would find life worth living, but there would be many false positives. On the other hand, if the standard is extremely high, such as eternal recurrence, then the test — because it is so demanding — will not falsely indicate that people find life worth living. Although this is an advantage of the eternal recurrence test, very few people, if any, would find life worth living using this test. Furthermore, many of the people who would gnash their teeth at the thought of eternal recurrence would affirm life using a more reasonable standard, such as asking themselves if they would have chosen to be born into this world and live their lives.

Method 6: Wanting to Live One More Time

As previously noted, Baier asserts that we can say a person finds his life worth living, if, and only if, this person would want to live his «same» life over again. The only difference between this standard and the eternal recurrence standard is the number of times that we are to imagine repeating life. With the standard proposed by Baier (hereinafter called «unirecurrence»), one imagines that one's same life recurs once whereas, with eternal recurrence, one imagines that one's same life recurs innumerable times. One may be willing to live one's same life over again once, but not willing to live one's same life an infinite

number of times. Thus, unirecurrence is not as demanding of a standard as is eternal recurrence. Nonetheless, because it involves repetition without variation, it is too high of a standard to use to judge whether life is worth living.

There are two ways to lower the unirecurrence standard, both of which involve preserving novelty or, in other words, of counteracting the boredom that would arise from living one's same life over again. The first way is to imagine that we live our second life anew and have no recollection of the events in our first life. To return to the movie example, a person would not likely want to watch the same movie two days in a row. However, if a period of ten years has elapsed from the time the person first watched the movie, such that the person will have forgotten many details of the movie, then this person may want to watch the movie a second time.

As evidenced by the earlier quote from Clark, it may be difficult for one to imagine recurrence without also imagining that one would have a continuing consciousness. Therefore, in developing a standard to judge whether life is worth living based on the idea of repetition, it will be useful to explore if there are other ways that novelty could be preserved. One such approach would be to allow some variation in our second life. Suppose, for example, that you would be born a second time. You would be born at the same time and place as your first birth, would have the same parents, and would remember your first life. However, your second life could unfold differently from your first life. Would you want to live this second life?

If a person would want to live a second life that is similar to his or her first life, then this suggests that this person finds life worth living. However, the second life cannot be too different from one's current life. If it is, then one could no longer validly infer whether one's current life is worth living. For example, suppose that a patient in a hospital who has suffered with lifelong pain and disability is in a state of despair and yet observes others who do find life worth living. If we ask him whether he would want to live again, and he imagines himself living as a joyful and healthy physician, then he may very well say that he would want to live again. The reverse situation could also occur. If a joyful physician is asked this question and imagines living a second life as a person who has severe functional limitations, he may decide that he would not want to live a second life. Consequently, if we imagine a second life that is too different from our own, then we can no longer validly infer whether our current life is worth living based on whether we would want to live again.

Conclusion

At the beginning of this essay, I noted that we could live our lives without evaluating whether they are good or bad, but that we do not. Evaluations are often performed to help us make decisions. For example, a committee may do a cost-benefit analysis to help determine whether to construct a new facility. To give another example, clinicians or medical ethics committees may perform an evaluation as a way to help guide medical decision making. They may try to determine, using various methods, whether a patient, who is unable to judge for himself, finds life worth living.[33] If they conclude that the patient, under the circumstances, would not find life worth living, then a decision may be made on economic or moral grounds to withdraw treatment from the patient.

---

[33]. For further discussion, see Bobbie Farsides and Robert Dunlop, «Measuring Quality of Life: Is there Such a Thing as a Life Not Worth Living?» *British Medical Journal* 322 (June 16, 2001): 1481-1483.

Because evaluations provide us with a basis for making decisions, it is understandable why they would be used to judge whether *someone else*, such as an incompetent patient, finds life worth living. But why do we judge the worth of our *own* lives, instead of living them without judging them? Are these evaluations undertaken for the purpose of helping us to decide whether or not to continue living, as might be thought? If this was the reason that we judged the worth of our lives, then some difficult questions would arise, including when should the evaluation be performed. If people perform the evaluation when they are young, then their judgement about life would be based on incomplete information. It is possible that based on their short experience they may conclude that life is not worth living, yet if they waited until they have had more experiences before judging life, then they may conclude otherwise, and vice versa. If one's judges one's life when one is very young, for example 20 years old, then one's evaluation will be incomplete. But if one waits too long before doing an evaluation, such as performing the evaluation when one's life is almost over, then the evaluation would no longer be relevant if the purpose of the evaluation is to help us decide whether to continue living beyond adolescence.

Some people judge that life is not worth living, but continue to live their lives. Therefore, if it is true that the reason that we perform these evaluations is to help us decide whether to continue living, then of what practical value are these evaluations if some people will continue to live despite concluding that life is not worth living? However, I do not believe that we undertake these evaluations to help us decide whether to continue living. Why, then, do we judge the worth of our lives?

Oftentimes, the evaluations are prompted by a question, such as when people ask themselves «Is there a meaning of life?» They may long to know whether life has value and may assume that there is a relation between whether life has a purpose and whether our lives have value. As a result, they attempt to address the question of whether life has a purpose. In some cases, they keep searching for an answer and never reach a conclusion. In other cases, they go beyond the question and adopt a standard or ideal — many times unwittingly — to judge life. They then compare life to their adopted standard and make a judgement about whether life measures up to the standard.

The evaluations and subsequent judgements are also prompted by unhappiness, pain, and suffering as, for example, when Job suffers and then begins to question whether life is worth living. The persons who are suffering compare their lives to a prior time when they were not suffering, to other people's lives, or to what they consider an ideal life, which, in turn, leads them to make a judgement about the value of their lives.

Because we make these comparisons between life as we are experiencing it, and life as we would like it to be, we will make judgements about the value of our lives even if we do not deliberately set out to do so. Therefore, it is important to understand the methods used to make these judgements and, if we find ourselves judging life, to select the best method for doing so.

A number of methods and standards for judging whether life has value have been evaluated to identify the best method and standard for making this judgement. The two purpose-based standards have been cast aside because they fail to take into account experiences that do not involve pursuing goals. Because one's life can be worth living even if one's life is not purposeful, the important question is, not whether there is a purpose(s) of or in life, but whether one's life is worth living. In addressing this question, it can be difficult

to determine if the benefits of being alive outweigh the costs. It is possible, though, for one to infer whether one finds life worth living.

Of the inferential methods that were evaluated, methods two (continuing to live) and four (choosing to bring others into the world) were rejected because, in many cases, they would falsely indicate that people find life worth living. Although the eternal recurrence standard and unirecurrence without variation (the standard proposed by Baier) would not yield false positives, they are unreasonable. To avoid the problems inherent with using standards that are too low or too high, a moderate standard is needed. There are two moderate standards to choose between: reflecting on whether we would have chosen to be born and live our lives, if the choice had been ours to make, and reflecting on whether we would want to live a second life that is similar to the one we are living (unirecurrence with variation).

Many people, I suspect, if the choice had been theirs to make, would have chosen to live their lives instead of remaining nonexistent. However, they may have no desire to live again, even if their second life were different from their first such that they could experience novelty and excitement. Thus, the repetition-based method reflects a higher standard for judging whether life is worth living. For life to be worth living, according to this higher standard, it is not enough to be satisfied with life, one must also want to live again. But why should one have to want to live again for one's life to be considered worth living? There is no advantage in adopting this higher standard. If one would have chosen to enter the world and live one's life knowing in advance what one would experience, and one feels grateful that one has had the opportunity to experience the sights, sounds, and wonders of this world, then this should be sufficient to find life worth living.

In conclusion, the best standard for judging whether one's life has value is to ask oneself whether one would have chosen to be born into this world and live one's life. If the answer is «no,» then the question becomes, what, if anything, should be done with this judgement. Should one ignore the evaluation, commit suicide, or do something to try to make one's life worth living? In deciding this question, one point worth considering is that these evaluations can fluctuate over time such that a person can find life not worth living at one point and find life worth living at a later point in life. If one answers «yes» to the question of whether one would have chosen to be born into this world, then one can feel confident that one finds life worth living.

Brooke Alan Trisel
<triselba@cs.com>

NORMATIVITY WITHOUT EXCEPTION: DONALD DAVIDSON ON LANGUAGE AND COMMUNICATION

Sonia Arribas

---

1. 1 Metaphor: The Exception to the Norm?

Language is ordinarily called «metaphorical» when it is replete with expressions that contain words with meanings different to the usual ones. It is commonly stated, for example, that in poetic language we find terms that have been removed from their usual context and inserted into a new one. According to this common understanding, it is this movement of words from one context to another that evokes in the reader or listener an aesthetic response. We have also become accustomed to hearing that, in the language of science, terms are used that originally mean one thing, but which, when applied to a new situation and directed toward a «specialized» community, end up meaning something completely different.[1] And, in more everyday terms, our sense of shock has long become anaesthetized to the bizarre twists given to words by all variety of self-appointed political orators. All of these phenomena — those praised for their beauty or wit and those considered manipulative — are *popularly* placed under the category of the «metaphorical». What this placement expresses is the belief that metaphor enacts a deviation with regard to familiar, typical or, more precisely, literal meanings; metaphor is a different — either enlightening or distorting — use of the meaning that we understand as valid. Given the existence of this popular conception — which, somewhat contradictorily, tends to both extol and denigrate metaphor — it should come as no surprise that the study of so-called metaphorical language has occupied a prominent role in philosophy, psychology, linguistics and literary criticism. In this paper, I want to use the terms «metaphor» and «metaphorical» in the popular sense of that which is «different linguistically speaking» in order to bring to light the mistaken philosophical and political presuppositions of this popular conception of metaphoricity. I want to call into question, that is, the presupposition that metaphor constitutes the exception to the norm. I want to claim that Donald Davidson's essay «What Metaphors Mean» (WMM)[2] provides the theoretical basis for this mistaken conception — the supposed «exceptionality» of metaphor.

The idea of metaphor's «exceptionality» finds its most extreme formulation in Davidson's already classic 1978 article «What Metaphors Mean». In accordance with

---

[1]. Mary Hesse, *Revolutions and Reconstructions in the Philosophy of Science*, Bloomington: Indiana University Press, 1990.

[2]. Donald Davidson, «What Metaphors Mean», in *Inquiries into Truth and Representation*, Oxford: Clarendon Press, 1984, pp. 245-264.

Davidson's famous thesis, metaphor possesses no metaphorical meaning whatsoever; it consists, rather, of a purely literal employment of words that give rise to unexpected «effects» in the reader or listener. I want to reconstruct in this opening section the main lines of argument of Davidson's article — an article which, once again, provides the best exemplification of the thesis that metaphor produces a use of language radically opposed to the usual or literal meaning of words. In WMM Davidson formulates a compelling critique of those semantic theories of metaphor which comprehend metaphorical meaning as a meaning either *added to* or *derived from* the metaphor's literal meaning. My reconstruction of WMM will therefore allow me to briefly examine these semantic theories and to express my agreement with Davidson's critique of them. In the following section (1. 2), however, I will attempt to show how Davidson's understanding of metaphor falls prey to the same error committed by the very semantic theories he is criticizing: both undervalue metaphor by strictly separating it from literalness. In Davidson's view, literalness and metaphorical effects belong to two wholly independent spheres of communication: the sphere of linguistic meaning (semantics) and the sphere of linguistic use (pragmatics). In the remaining sections (1. 3 and 1. 4), I will investigate two of his later texts: «The Idea of a Conceptual Scheme» and «A Nice Derangement of Epitaphs». This will allow me to sketch the outline of a conception of metaphor that conceives it neither as a case of literalness (consistently with Davidson's critique of traditional semantic theories of metaphorical meaning) nor as a mere effect of the use of language (consistently with my critique of Davidson). I will understand metaphor as both meaning and use — as something different from literal meaning but, at the same time, as something intrinsic to language.

In «What Metaphors Mean», Davidson condenses his thesis, provocative in its simplicity, in the following way: «metaphors mean what the words in their most literal interpretation mean, and nothing more.» (WMM, p. 245) Metaphor possesses no other meaning than the literal meaning of its words. In effect, however, we can distinguish metaphor from its linguistic meaning. Metaphoricity has to be conceived, therefore, not as a linguistic meaning, but as an «effect» or emotion evoked in the receiver. Davidson's argument in favor of this thesis proceeds for the most part negatively: as a critique of traditional semantic theories of metaphor and of the very notion of a discrete metaphorical meaning. It is only on the basis of these criticisms that Davidson articulates his own positive account of metaphor. I will here reconstruct this analysis of metaphor according to the same sequence of negative and positive claims. Davidson's critique of the notion of metaphorical meaning will be seen to be based on two presuppositions: firstly, on the idea that the literal meaning of the words that form part of a metaphor has to be visible in order for the metaphor to exist as such; and secondly, on the stipulation that literal meaning and metaphorical use are strictly separated. Davidson's positive account of metaphor is articulated as a «causal» or pragmatic theory — one that conceives metaphor as a producer of effects.

Davidson's first presupposition can be stated as follows: if we are even to comprehend a metaphorical expression, then the literal meaning of the terms that make it up must remain active. Davidson defends this presupposition by criticizing two semantic theories of metaphor: the theory of «extended» literal meaning and the theory of elliptical simile.[3] According to the first theory of «extended» literal meaning, in a metaphorical utterance the literal meaning of

---

[3]. There are numerous versions of both theories. For a contemporary version of the theory of «extended» literal meaning, see Nelson Goodman, *Languages of Art*, Indianapolis: Bobbs Merrill, 1968. The theory of elliptical simile dates back to Aristotle, *Rhetoric*, III, 1406 b, and Cicero, *De Oratore*, III, 38.

a word is extended so as to include a new object under its scope of reference. In order to demonstrate the failure of this explanation, Davidson carries out a *reductio ad absurdum* by means of an example: the Biblical metaphor «the Spirit of God moved upon the face of the waters.» To explain the metaphorical use of the noun «face» here, the extended literal meaning theory states, as its name suggests, that the class of objects literally referred to by the concept «face» has been extended so as to include the object «waters». In logical terms, metaphor increases the extension of objects referred to by the word «face». In its metaphorical employment, that is, the word «face» literally means «the front part of the head» *and* «waters». If this account were true, Davidson contends, that is, if «face» literally referred to waters, then we would not possess an explanation of the metaphor we originally intended to explain, but an explanation of a new literal meaning for the word «face». Instead of elucidating metaphoricity, the theory of «extended» meaning provides us with an account of what is usually referred to as «dead metaphor». A dead metaphor is one in which a previously metaphorical expression has come to be accepted as literal — as in the expression the «mouth of a river», in which the original literal meaning of the word «mouth» (the opening of the digestive tube), at least at first sight, goes unnoticed, and is instead understood as a literal designation for the point at which a river meets the sea. In a dead metaphor, the original literal meaning of the words that form the expression disappears. The theory of extended literal meaning only provides an account, therefore, of a kind of metaphor that has lost its metaphoricity. If metaphor is to be explained in a way that preserves its metaphoricity, Davidson concludes, then the original meaning of the words that make it up must remain active and be immediately perceived.

In his critique of the theory of elliptical simile, Davidson once again argues for the necessary persistence of the literal meaning of the terms that constitute a metaphor. The elliptical simile theory tries to provide an account of metaphor in terms of its corresponding simile: the literal meaning of a metaphor is seen to be equivalent to the literal meaning of the related simile, and its metaphorical meaning to the simile's figurative meaning. The literal and metaphorical meanings of the expression «Tolstoy was an infant», for instance, are equivalent to those of the simile «Tolstoy was like an infant». Davidson subjects this account to criticism for two reasons. Firstly, because in equating the metaphorical meaning of metaphor with the figurative meaning of simile it explains nothing: a simile «no more *tells* us what similarities we are to notice than the metaphor does.» (WMM, pp. 254-5) Even when we accept that the metaphor and the corresponding simile could have the same figurative meaning, what simile figuratively conveys «is not a feature of the word that the word has prior to and independent of the context of use.» (WMM, p. 255) We cannot determine the figurative meaning supposedly shared by the metaphor and the simile simply by affirming that it is the same in both cases. Davidson's second criticism draws attention to the misapprehensions that necessarily follow from an identification of the literal meanings of metaphor and simile. The proposition articulated by a literal interpretation of the simile «Tolstoy was like an infant» explicitly affirms, for example, a similarity. And, on its part, the proposition expressed by the literal interpretation of this metaphor is either a falsehood (since Tolstoy did not write novels when he was a child) or an obvious truth (since Tolstoy, like everybody else, had a childhood). Whatever the truth content of the literal meaning of the words that form part of metaphor might be then, Davidson concludes, it has to remain active for metaphor to exist at all. When «we take the literal meaning of the metaphor to be the literal meaning of the matching simile, we deny access to what we originally took to be the literal meaning of the metaphor.» (WMM, p. 254)

The second presupposition underlying Davidson's understanding of metaphor comes to the fore in his critique of two strictly semantic theories of metaphor: one of which is derived from the philosophy of Frege, the other of which attempts to translate metaphor into paraphrase. This second presupposition can be summarized as follows: the capacity of *learning* the meaning of words and the capacity of *using* these words belong to two radically distinct spheres. Or, with more specific regard to the question of metaphoricity, whereas the literal functioning of words pertains exclusively to the learning of language, metaphor emerges exclusively through the use of language. Firstly, then, the Fregean theories criticized by Davidson conceive of metaphor as possessing both a literal and a metaphorical meaning. These two meanings are connected by a rule that specifies the situation or context in which each obtains. This explanation of metaphor is derived from Frege's account of the meaning of sentences that express propositional attitudes such as belief or desire («I wish I were healthy», for example).[4] The meaning of these kinds of expression can be ascertained, Frege claims, by means of a rule that derives it from the meaning of the words of the corresponding modal sentence («I am healthy», for example). Davidson refutes this characterization of metaphor — as the result of the application of a rule — by devising a somewhat curious example. If we were to teach a Saturnian the word «floor» for the first time by indicating floors and simultaneously uttering the word «floor», the Saturnian would be, Davidson claims, *learning* the meaning of the word. If the Saturnian were to subsequently want to test her grasp of this newly acquired knowledge, she would utter the word «floor», point at a floor, and await our confirmation. This second instance would also, in Davidson's view, be one of *learning* a new use of the literal meaning of the word «floor». But suppose the Saturnian were then to fly back to Saturn, point at the earth, and exclaim «floor». How could we then discern whether she is learning a new literal meaning of the word «floor», that is, whether she thinks that the earth literally is a floor, or whether she is using a metaphor, as when Dante refers to the earth as «the small and round floor that makes us passionate»? We have no rule, Davidson claims, to help us in the ascertainment of the Saturnian's intention and, in order to remedy this confusion, we need to make a clear distinction between the *learning* of the meanings and uses of words and the effective *use* of these words. Metaphors, to cut a long story short, only emerge in use. Three things follow from this example: (1) when the Saturnian attentively listens to our lesson on floors, she is *learning* the literal meaning of the word «floor» — or, as Davidson claims, «what is learned is that a bit of language refers to a bit of the world» (WMM, p. 251). (2) When the Saturnian points to something else, exclaims «floor» and waits to see if she has got it right, she is *learning* a new literal use of the word «floor». And (3) when the Saturnian finally looks at the Earth from Saturn and says «floor», she is metaphorically *using* a word that she already knows. In accordance with Davidson's analysis, the radical difference between cases (1) and (2), on the one hand, and case (3), on the other, is that whereas the former deal solely with the mastery and use of literal meanings through a process of learning, the latter refers to the active usage of these meanings. When we learn a language, we remain within the confines of the literal; it is only when we use language in the world that we can enter into the domain of the metaphorical (this does not mean, of course, that every use of language is metaphorical). If literal meaning is entirely governed by the conventions that must be mastered in order to understand and speak, the use of metaphor arises impulsively and only occurs outside of the boundaries of these conventions.

---

[4].   Gottlob Frege, «On Sense and Reference», in *Philosophical Writings,* ed. by M. Black and P. T. Geach, Oxford: Blackwell, 1962, pp. 56-78.

Davidson's critique of the Fregean theory of metaphor is grounded, then, upon a strict separation between the literal interpretation of words (understood as a purely linguistic function) and the metaphorical usage of expressions (conceived as a worldly, extra-linguistic activity). It is the same distinction that centrally informs Davidson's critique of the paraphrasability theory of metaphor — a theory which maintains that we can grasp the full meaning of a metaphor only when we paraphrase or restate it in words that exactly convey the meaning of the original expression. Although he never says so explicitly, Davidson's critique of this theory is based upon the notion of «cognitive content». In Davidson's theory, this notion functions as the guarantee of a fundamental unity of language — of a definite, fixed and limited set of thoughts or things referred to by locutions and determined by conventions. Because the cognitive content of each word is stipulated by convention, it follows that we can always find or create another term (or set of terms) equivalent to it: the *paraphrase*. In addition and opposition to this notion of cognitive content, however, Davidson also wants to maintain that the production of non-cognitive effects is essential to communication. Non-cognitive effects — in contradistinction to cognitive content — are never fully determined or governed by convention. They belong instead to the realm of emotions and can thus be aroused involuntarily in the reader or listener. The paraphrasability theory of metaphor mistakenly presupposes, in Davidson's view, that metaphor possesses a cognitive content over and above that of the purely literal interpretation of its terms: «If what the metaphor makes us notice were finite in scope and propositional in nature, this would not in itself make trouble; we would simply project the content the metaphor brought to mind on to the metaphor. But in fact there is no limit to what a metaphor calls to our attention, and much of what we are caused to notice is not propositional in character. When we try to say what a metaphor 'means', we soon realize there is no end to what we want to mention.» (WMM, pp. 262-3). What Davidson is proposing here is a distinction between literalness — conceived as the univocal articulation of a finite set of ideas — and metaphoricity — conceived as the expression or evocation of an inexhaustible number of suggestions and emotions which are, by definition, non-paraphraseable. According to Davidson's separation of spheres, literal meaning can be completely and precisely transferred from speaker to addressee; and the addressee can in turn grasp and even restate this meaning by means of an exactly equivalent expression. Metaphoricity, on the contrary, is an extra-linguistic phenomenon, one which cannot be substituted by any other term; it takes place in the world, as an external cause, a «bump on the head» (WMM, p. 262).

It is this distinction that allows us to summarize Davidson's causal theory of metaphor. If meaning is comprehended as a strictly propositional cognitive content capable of being paraphrased, then metaphor can only possess a literal meaning. The unparaphraseable and unsubstitutable character of the effects that a metaphor produces upon its addressee are the means to gain access to other cognitive contents. Metaphor assumes in accordance with this scheme a role similar to that played by speech acts. The effect of a metaphor — like the effect of a promise or a lie — does not reside in the literal meaning of the words that it enunciates, but exclusively in the use that we give to these words: «What makes the difference between a lie and a metaphor is not a difference in the words used or what they mean (in any strict sense of meaning) but in how the words are used.» (WMM, p. 259) Because metaphor lacks cognitive content, Davidson claims, it tells us nothing about reality; it possesses no meaning that could be deciphered or spelled out: «what we attempt in 'paraphrasing' a metaphor cannot be to give its meaning, for that lies on the surface: rather we attempt to evoke what the metaphor brings to our attention.» (WMM, p. 262) Whereas literal linguistic meaning can be

articulated and repeated independently of the context of its utterance, metaphor can only be identified and interpreted in the light of a given context. Metaphor produces an inexhaustible series of emotions, it invites us to look at reality in another way, it incites us to perceive connections between things that ordinarily go unnoticed: «When we try to say what a metaphor 'means,' we soon realize there is no end to what we want to mention.» (WMM, p. 263) The cognitive power of metaphor has to be clearly differentiated, however, from literal knowledge; as Davidson cautions: «[s]eeing as is not seeing that. Metaphor makes us see one thing as another by making some literal statement that inspires or prompts the insight. Since in most cases what the metaphor prompts is not entirely, or even at all, a recognition of some truth or fact, the attempt to give literal expression to the content of the metaphor is simply misguided.» (WMM, p. 263). When confronted by a metaphor, the interpreter attains a new perspective on reality, not a claim about truth or falsity.

1. 2 Language as Relation

In summarizing Davidson's critique of the various attempted reductions of metaphor — to a literal extended meaning, to a simile, to the application of a rule, to a paraphrase — I have brought to light the two basic presuppositions of his theory; both of which refute the idea that metaphor possesses a metaphorical meaning. According to Davidson's first presupposition — that of the necessary activity of the literal meaning of the metaphor's terms — if a metaphor were to possess any other meaning in addition to the literal meaning of the words that compose it, it would immediately cease to be a metaphor. According to his second presupposition — that of a strict separation between the meaning and the use of words — metaphoricity is characterized as a contextual and extra-linguistic phenomenon that triggers effects irreducible to a determinate meaning. These effects can be considered as «useful» but not as meaningful. In this section, I want to argue that Davidson's first presupposition (concerning the necessary persistence of the literal) is broadly correct — if a metaphorical utterance is to be understood, then the literal meaning of the terms must, to some extent, remain active. I also want to claim, however, that Davidson's second presupposition (concerning the strict separation of the meaning and use of words) contradicts his first: the strict separation of meaning and use precludes the assumption that the literal meaning of a metaphor's terms plays a role in its metaphoricity. In order to maintain my defense of Davidson's first presupposition, I will argue that literal meaning and metaphoricity must not be understood dualistically, but instead in terms of a relation. On the basis of this understanding, I will also try to show that metaphor is comprehensible, that is, that it possesses a meaning and, as a consequence, can be normatively assessed. In order to develop these claims, I will return, firstly, to Davidson's critique of traditional semantic theories of metaphor. I will concentrate, secondly, on the above-mentioned contradiction between Davidson's two fundamental presuppositions.

As we saw in section 1. 1, purely semantic explanations of metaphor tend to obliterate not only the literal meaning of the words that compose it, but also the very phenomenon of metaphoricity itself. When metaphor is endowed with a supposedly metaphorical meaning that is comprehended as eclipsing the literal meaning of its words, it is reduced to a convention: metaphorical meaning is conceived either as a literal but extended meaning or as a meaning derived from a rule of application. Once this is accepted, we need to maintain that, if a metaphor is to arise, there can be no conventional link between the literal meaning of its terms and the metaphorical phenomenon itself. If the metaphorical expression is to effectively

produce the series of effects that Davidson describes, then the literal meaning of the words that constitute it has to remain active.

I want to show now, however, that Davidson's second presupposition — that of an effective dualism between literalness and metaphoricity — (and the series of epistemological oppositions that it assumes) renders impossible any explanation of why it is that metaphor is capable of provoking effects in its addressee. If we follow Davidson then we must say, on the one hand, that meaning is solely to be located in the literalness of words, that is, in a determinate cognitive content. This content is subject to learning processes: words univocally refer to meanings and the relations between them are ruled by conventions. These meanings and conventions are, on Davidson's account, the object *par excellence* of semantics. If, according to this view, convention determines meaning, it follows that the latter can always be ascertained independently of the infinite contextual particularities of communication. We must say, on the other hand, that metaphoricity is itself devoid of meaning, that is, that it lacks conventional cognitive content. As such, it is the object of pragmatics; it is the tool that allows us to use language in the world in order to cause or provoke suggestions and intimations. These suggestions and intimations are essentially indeterminate and hence radically different from accepted literal meanings. They produce effects which cannot be paraphrased, effects which are as infinite as the possible situations in which words can be used. If literalness and metaphoricity are radically disconnected, however — in such a way that metaphor, conceived as an extra-linguistic phenomenon irreducible to literal meaning, always escapes the yoke of convention — then why does Davidson consider it necessary to perceive the literalness of a metaphor's terms? Is there not a relation between literal meaning and metaphorical effects which contradicts Davidson's second presupposition?

We can begin to answer this question by concentrating on the so-called effects produced by a metaphor. Davidson's strict separation between meaning and use restricts these effects to the sphere of use and comprehends them as emotions or suggestions evoked in the addressee. These emotions are provoked by the speaker and are totally dependent on the context of the enunciation. They are purely «sensuous» and are opposed to the cognitive content of the metaphor's own terms. It follows from this that there takes place between effects and meaning only a juxtaposition, that is, an addition without connection. It is for this reason that Davidson maintains that, with regard to the literal meaning of the terms being used, the effects function merely as an accessory. If, however, according to Davidson's first presupposition, the literal meaning of the expression has to remain active in order even for metaphor to arise, then it seems impossible to avoid the inference that the understanding of literal meaning somehow causes or is related to the supposedly pure effects. The production of effects must be construed in relation to the understanding of the literal meaning. (Later on, I will examine in more detail the nature of this relation). Furthermore, it also appears self-evident that, if the arousal of effects depends upon the context in which the locution is enunciated, the literal and active meaning cannot be free from the influence of this context. It is precisely this conception of a necessary relation between meaning and effects that renders impossible Davidson's second presupposition (of a total separation between literal meaning and the contextuality of metaphorical effects). In nuce, there is a contradiction between the first and the second presuppositions of Davidson's analysis. This contradiction can only be avoided, I want to claim, if — in accordance with a proper understanding of the first presupposition — Davidson's dualistic understanding of literal meaning and metaphorical effect is replaced by a relational or dialectical one. As we will come to see, this relational or dialectical understanding will also compel us to call into question Davidson's strict separation

between (literal) meaning and (metaphorical) use: both meaning and use will be seen to play a role in the literal and the metaphorical.

Before spelling out the nature of this relation, however, I want to make a brief detour in order to discuss a question that Davidson himself raises: the question of dead metaphor. The main problem that Davidson faces with regard to this question is the following: how can we explain the emergence of a dead metaphor from what used to be a living one? Let's us take as an example a typical expression from sociology and political science, that of a social «role» (from the French «rôle»), which literally means «a behavior appropriate for some relations but not for others.»[5] In its historical origins, the living metaphor «role» emerged as something that was used to designate theatrical behavior in social life. The term was translated, that is, from a theatrical context — in which it denoted a «part in a play» — to the wider context of social life — where it was used to qualify certain public attitudes as artificial.[6] According to the «extended» literal meaning theory, the original meaning of «role» goes unnoticed when we speak of social life because the metaphor is dead and the word «role» now literally indicates a social position. The word «role» possesses, that is, at least two literal meanings. On Davidson's account, this amounts to saying that the meaningless effects that the metaphor originally triggered when it was first used as a living metaphor have transmogrified into the numerous meaningful traits of a social conduct denoted by a dead metaphor. How is it, however, that these effects, which are by definition incommensurable to meaning, can come to signify something determinate, conventional and context-independent? Because it presupposes that metaphoricity is a phenomenon wholly extrinsic to language, Davidson's account cannot explain the transformation from living to dead metaphor.

The previous considerations pose a series of problems that will lead me to formulate, in the following section (1. 3), an alternative to the theory of metaphor that Davidson advances in WMM. For the moment, I would like to advance two characteristics of the interpretation of literalness and metaphoricity that can only be comprehended by means of a relational or dialectical understanding of meaning and use. By pointing out these two characteristics, I want to once again draw attention to the above-mentioned problems with Davidson's position. Firstly, then, we have seen that Davidson totally rejects the notion of metaphorical meaning that is developed by strictly semantic theories of metaphor. The basis of this rejection — the presupposition of the persistence of the literal meaning of the terms that constitute the metaphor — is undermined, however, by the long list of oppositions that derive from the second presupposition. But if — as I have already suggested — we abandon the strict distinction between literalness and metaphoricity, we can hold on to the thesis that literal meaning remains perceptible, without needing to oppose to it a totally alien metaphorical effect. In this way, the so-called metaphorical effects can be conceived as intrinsic to language and, hence, as comprehensible, that is, as capable of being translated into a determinate meaning. This amounts to claiming that metaphoricity should be conceived as a link between meaning and use. And, conversely and in a dialectical movement, it means that the

---

[5]. Richard Sennett, *The Fall of the Public Man,* London, Boston: Faber and Faber, 1986, p. 33. In the theatrical context, this metaphor originated from another metaphor: the «rôle» or «paper» was literally the piece of paper that contained the part which each actor had to memorize.

[6]. The Baroque origin of this conception of social and political life is explained in the chapter «Theatrum Mundi: teatro, máscara y escena política,» in José M. González García, *Metáforas del Poder*, Madrid: Alianza, 1998, pp. 97-142, and in Alexander Demandt, *Metaphern für Geschichte, Sprachbilder und Gleichnisse im historisch-politischen Denken*, Munich: Beck, 1978, chapter VI.

comprehension of literalness has to be reformulated with reference to metaphoricity, that is, once again, in terms of a relation between meaning and use. Secondly, this abandonment of the strict distinction between literalness and metaphoricity implicitly refutes Davidson's claim that «there are no unsuccessful metaphors» (WMM, p. 245). If the activity of the literal meaning of words is necessary for metaphor, then — as we will see in the next sections — this meaning (and use), when understood in relation to metaphoricity, will provide us with normative criteria by metaphor can be assessed (without reducing it to literalness). The traditional priority accorded to literalness is the result of the dichotomy that undialectically opposes it to metaphoricity. Once literalness and metaphoricity enter into a relation of co-implication, the notion of this priority can be discarded.

The strict distinction between literalness and metaphoricity — as belonging to two epistemologically opposed spheres — entails an erroneous conception of both metaphor and literal meaning and, thus, of language in general. The metaphorical is produced in a world supposedly external to language. In this world, the normativity or rationality of an utterance can assume no place, since the so-called effects of language are by nature incomprehensible. Language, for its part, is reduced to a fixed number of literal meanings, determined by convention and totally context-independent. I will henceforth refer to this mistaken conception of literal meaning as «restrictive», both because it conceives literal meaning as unchangeable and as unaffected by language-use, and because it limits language to a determined and fixed amount of restrictive literal meanings. In the following sections, I want to develop a critique of the notion of restrictive literal meaning — and of the epistemological position that grounds it — by advancing a reconstruction of two other essays by Davidson. Without explicitly referring to the question of metaphor, Davidson articulates in these articles a very different conception of language to that contained in WMM. My aim is thus to develop this conception into an alternative theory of metaphor.

1. 3 «Passing Theory»: Meaning and Belief

The relation between language and world is conceived in strictly dichotomous terms when the world is conceived both as external to language and as capable of exerting an epistemological impact on knowledge. In his 1974 article «On the Very Idea of a Conceptual Scheme» (CS)[7] Davidson uses a number of deflationary arguments in order to show the unintelligibility of this conception. The fundamental characteristic of the dualist conception of language and world, Davidson claims, is that language is conceived as a system or conceptual scheme that maps or projects itself onto a worldly reality. This model recurs, Davidson maintains, in different systems and conceptions of philosophy, all of which, as we shall see, coincide in comprehending language as a set of meanings organizing a supposedly external world. It is stated, for example, that language configures empirical or sensuous data, that it provides us with a point of view or that it endows us with a number of categories by means of which we can comprehend or analyze reality. According to Davidson, these various formulations of language as a conceptual scheme entail a thesis of *relativism:* given that the world is structured by conceptual schemes, different schemes can arrange the world in divergent or incommensurable ways. In linguistic terms, different languages speak about the world in ways divergent from and untranslatable into other languages. In order to call into question — and ultimately abandon — the idea of a conceptual scheme, Davidson discusses the

---

[7].   Donald Davidson, «On the Very Idea of a Conceptual Scheme», in *Inquiries into Truth & Interpretation*, Oxford: Clarendon Press, 1984, pp. 183-198.

two main forms in which the untranslatability thesis is defended. Firstly, the form of complete untranslatability, which is said to occur when no set of categories of one language can be converted into another; and, secondly, that of partial untranslatability, in accordance with which some categories can be translated, and some cannot. Davidson's critique of the thesis of complete untranslatability dismantles a concomitant dualism between language and world, that is, it dismantles Davidson's own second presupposition in WMM. Davidson's considerations on partial untranslatability prefigure a non-restrictive conception of literal meaning. This conception of a non-restrictive literalness is more thoroughly elaborated in Davidson's «A Nice Derangement of Epitaphs» (NDE),[8] to which I will dedicate the last pages of this section.

One common manifestation of the thesis of complete untranslatability of different languages (or of the incommensurability between different conceptual schemes) is the claim that translatability is not transitive. According to this idea, given three languages, A, B and C, when A can be translated into B and vice versa, and B can be reciprocally translated into C, it does not follow that A and C can be translated into each other. Davidson argues against this claim by questioning the presupposition that — from the perspective of an A speaker, and given that A is not translatable into C (or vice versa) — an A speaker is entitled to affirm that C has been translated into B. For in order for an A speaker to verify the validity of the translation of C into B she would have to recognize that the expression is part of C. Without this act of recognition A would never be able to verify whether a B speaker who claims to have successfully translated an expression from C to B is pretending or telling the truth. An A speaker could simply *assume* that the B speaker is not lying, but she would never be able to *know* whether the expression was actually translated. For A to be able to recognize what it is that a B speaker has translated, A has to recognize the expression as part of language C. And this already entails that A and C are translatable.

A second argument in favor of complete untranslatability is postulated by those philosophers of language and science who, inspired by Kuhn,[9] maintain that when different observers possess different conceptual schemes their perspectives about the world are also radically different. Davidson suggests that this relativistic position is generally articulated as an alternative to Strawson's thesis[10] that the distinction between concept and world is linguistic — that is, that it is only *within* language that we can distinguish between a concept and its referent. Davidson shares Strawson's thesis, and in order to defend it makes a detour. The followers of Kuhn could attempt a criticism of Strawson, Davidson conjectures, by arguing the following: the analytic-synthetic distinction in language states that analytic sentences are true or false by virtue of their meaning alone, and that synthetic sentences are true or false because of their empirical content; if we accept Quine's critique of this distinction,[11] it follows that we should also give up the idea that meaning is independent of

[8]. Donald Davidson, «A Nice Derangement of Epitaphs», in *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*, ed. by Ernest Lepore, Oxford, New York: Basil Blackwell, 1986, 1989, pp. 433-446.

[9]. T. S. Kuhn, *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 1962.

[10]. P. F. Strawson, *The Bounds of Sense*, London: Methuen, 1966.

[11]. W. V. Quine, «Two Dogmas of Empiricism,» in *From a Logical Point of View*, Cambridge, Mass.: Harvard University Press, 1961.

our beliefs about what is true or false; if meaning radically depends on belief, then whenever the speakers of a language change a sufficient number of beliefs, they will not simply exchange falsehoods for truths (or vice versa); they will rather adhere to a new set of meanings, that is, to a new conceptual scheme. Davidson proceeds to question this reasoning by pointing, by means of an example, at its basic presupposition. Suppose some philosophers were to advocate that instead of using terms that apply to psychological states we use only those terms, already present in our language, that refer to the equivalent physiological events in the brain. According to the followers of Kuhn, a change of scientific belief would have here compelled the creation of a new language or conceptual scheme, one radically incommensurable with, or untranslatable into, the old language or conceptual scheme. Davidson replies to this contention by claiming that, when scientific advances like this one compel a change in the use of language, this in no way means that, after the change, we are employing a new conceptual scheme incommensurable with the old. For, if that were the case, we would not be able to determine whether the words used in the new language *actually* refer to physiological events — they could just as easily be synonyms of the old psychological terms. The criteria of the old language (the one that has apparently been superseded) must be presupposed if we are to be able to recognize that the supposedly new terms refer to physiological events. The Kuhnian argument in favor of untranslatability and, coterminously, in favor of the idea of a conceptual scheme, has to be discarded.

The third derivation of the argument for complete untranslatability discussed by Davidson also stems from the abandonment of the analytic-synthetic distinction in language: even if we give up this distinction, and drop its related conception of meaning, we can still hold on to the notion of empirical content. We can still maintain that all sentences — and not only synthetic sentences — have an empirical content. Empirical content can be interpreted on this view as the world, as reality, as the density of experience, as the totality of sensory stimuli or, more generally, as something outside language which, in order to be known at all, requires the imposition of an organized structure. Concomitantly, a conceptual scheme can be described as that which imposes the required order or form onto the empirical content. This formulation — denominated by Davidson as the «third dogma of empiricism» — merely reproduces, he claims, a dualism of world and language that is as untenable as the one informing the thesis of intransitivity and conceptual relativism. The argument proceeds in the usual manner: Davidson, firstly, presents the relationship between conceptual scheme and empirical content before, secondly, going on to prove its unintelligibility. This relationship is usually expressed, according to Davidson, in one of the following ways: either (1) conceptual schemes *organize* reality or experience or (2) they *fit* reality or experience. The dictum that conceptual schemes organize or put in order reality or experience can only possess a meaning, Davidson claims, if we take it to state that conceptual schemes organize objects within reality or sense-data within experience. For reality or experience, taken as a whole or as a single object, cannot be organized, just as it is not a closet itself that is organized, but the things inside it. Even if this criticism is accepted, however, a response remains open to those who defend the idea of a conceptual scheme. On the assumption that there exist different conceptual schemes organizing a manifold of objects within reality or experience, it could still be affirmed that, within one conceptual scheme, there might be some concepts which apply to determinate objects or sense-data and which do not have correlate concepts in another conceptual scheme. Davidson replies to this objection by claiming that we only perceive the lack of correlation between conceptual schemes if we acknowledge that there are a series of items (objects within reality or sense-data within the realm of experience) referred to by

concepts in *both* conceptual schemes. Without this background of shared understanding, there is no possibility of a comparison of concepts. The first version of the third dogma of empiricism is, then, indefensible. In the second version, (2) above the emphasis is no longer placed on concepts that individualize and hence organize reality or experience, but on sentences that fit or correspond to it. To affirm that sentences fit reality or experience, Davidson claims, is to conceive of the latter as the place where we can find evidence for the affirmation. But if to offer this kind of evidence is to find the affirmation to be true, then to say that a sentence fits reality or experience is simply to say that it is a true sentence. But, what is it, Davidson asks, that makes us recognize a sentence as true? A fact in reality or experience? The evidence that we already possessed? This is the surprising answer:

> Nothing (…) no *thing,* makes sentences and theories true: not experience, not surface irritations, not the world, can make a sentence true. *That* experience takes a certain course, that our skin is warmed or punctured, that the universe is finite, these facts, if we like to talk that way, make sentences and theories true. But this point is put better without mention of facts. The sentence 'My skin is warm' is true if and only if my skin is warm. Here there is no reference to a fact, a world, an experience, or a piece of evidence. (CS, p. 194)

To believe something to be true does not imply that there is an object in experience or reality that is related to this belief and which makes it true. Nor do we possess an epistemological device that would allow us to relate or fit together an object and a belief with the aim of making the belief plausible. On the contrary, everything we need in order to be certain about a belief resides in the meaning of the sentence itself. To know that our belief is true is to know the language in which the sentence is uttered. If, after all of this, we still wanted to maintain the idea of a conceptual scheme — as the idea of a set of sentences that fit reality or experience in ways incommensurable to other sets of sentences — we would have to reformulate it so as to affirm that sets of sentences in different languages possess different truths. This last thesis, however, Davidson claims, cannot be valid. For to maintain that truth is defined within language means to apply the predicate «is true» to all the possible sentences for which truth is in question, including those from other languages. This is Tarski's famous Convention T,[12] which claims that a linguistic statement can only be true when it can be translated into another language. The sentence «snow is white», for example, is true in English if, and only if, snow is white *(p);* and true in another language when we can find in it an equivalent expression that can be translated into «snow is white» and substituted as *p.* Sentences can be understood and asserted to be true, that is, only in relationship to other sentences — either from the same language or from other languages. Once again, the idea of a conceptual scheme as a non-translatable framework of truth has been refuted: truth cannot be conceived independently of translation.

Davidson's analysis has shown that the idea of a conceptual scheme -coterminous with the idea of conceptual relativism or complete untranslatability, and with a dualistic conception of language and world- cannot be justified by any of its most common philosophical defenses. This idea cannot be maintained, that is, either as an intransitivity between languages, as a relativism derivative from a change of belief or as a «third dogma of empiricism». What remains to be addressed, however, is the thesis of partial untranslatability. This thesis maintains that the basic conceptual presuppositions of differing views — even if they are intelligible to one another — are exclusive to their respective conceptual schemes. In reply to

---

12. A. Tarski, «The Concept of Truth in Formalized Languages,» in *Logic, Semantics, Matamathematics*, Oxford: Clarendon Press, 1956.

this thesis, Davidson inquires what it is that makes people capable of being understood and of understanding others. He tries to develop a theory of the conditions of intelligibility, that is, that gives an account of the interconnection between the attribution of belief and the interpretation of sentences. Communication and understanding is to be conceived as a two-fold process in which we can interpret others only because we attribute beliefs to them, that is, only because we attribute commitment to their sentences. By implication, this is a process in which we can trust the assertions of others only because we are able to extrapolate the meaning of what they are saying. This theory finds its most complete development in NDE. In this article, Davidson explicitly refutes the thesis of partial untranslatability and advances — as against his own restrictive account of literal meaning in WMM — a revised conception of literal meaning and, as a consequence, of metaphor. I will deal with Davidson's refutation of partial untranslatability at the end of this section. I want now to concentrate on his reworked conception of literal meaning.

The title «A Nice Derangement of Epitaphs» makes a pun on Mrs. Malaprop's infamous «a nice arrangement of epithets»; the topic of Davidson's article is precisely that kind of linguistic use that deviates from literalness. A malapropism, for instance, is a linguistic utterance in which an appropriate term is replaced by a term with a similar sound but a completely inadequate meaning. Despite their anomalousness, these utterances succeed as a form of communication. If successful communication occurs whenever the hearer understands the speaker in the way the speaker had in mind, then all it takes for communication to succeed, in the example at hand, is for the hearer to understand the malapropism as a malapropism. These preliminary remarks alone already indicate that what Davidson is envisaging here is an alternative, non-restrictive, definition of literalness. Whereas in WMM Davidson denied the possibility of meaning outside convention, in NDE he contends that the understanding of meaning can arise even in a supposedly irregular — in this case, malapropistic — circumstance. And whereas before he declared meaning to be context-independent, he now argues that «we want a deeper notion of what words, when spoken in context, mean.» (NDE, p. 434). Davidson's goal is, then, to reformulate the notion of literal meaning, without thereby making it indistinguishable from the speaker's intended meaning.[13] This distinction between literal meaning and the speaker's intended meaning cannot be given up for, if this were done, all meanings would be left to the intention of the speakers, that is, no room would remain open for linguistic understanding.

Instead of beginning with a direct refutation of the restrictive definition of literalness, Davidson prefers to interrogate the basic requirements that a new definition of literalness should satisfy once successful communication is accepted in supposedly irregular cases. A definition of literalness, Davidson claims, has to account for the fact that literal meaning «comes first in the order of interpretation.» (NDE, p. 435). I shall, for the time being — in order to avoid the customary restrictive definition of literalness — give the name «first meaning» to the literal meaning that needs to take the first place in the series of possible interpretations of an expression. First meaning is the meaning which — given the possibility of different and alternative ways of interpreting an utterance — seems to come closest to the intention of the speaker, and which, equally importantly, will be interpreted by the hearer as the speaker intends. First meaning is *simultaneously* determined, then, from both the speaker's and the listener's perspective. From the speaker's perspective, it is determined both by the

---

[13]. «Intended meaning» is the meaning that the speaker wants to give to her words, and which will be so understood by the listener.

intention to say something and by the intention that the hearer understand what is said (Davidson denominates the latter as «self-referring» or double intention). From the hearer's point of view, first meaning is determined when what the speaker says is understood, and when this understanding corresponds to what the speaker intended. In short, first meaning arises when a speaker's belief in and interpretation of an utterance corresponds to a hearer's.

Having adopted the concept of first meaning (instead of the common conception of literalness) as the basis of his theory of meaning, Davidson proceeds to analyze and modify three explanations of «irregular» cases that derive from the common conception of literalness. According to the first explanation, «first meaning is systematic», that is, it is immersed in a system of semantic relationships to other meanings. The second explanation states that «first meanings are shared», that is, in addition to belonging to a system — as the first explanation stipulates — they are also shared by speaker and listener. The third and final explanation maintains that «first meanings are governed by conventions and regularities.» (NDE, p. 436) We can begin to see the problems that these three proposals encounter when we consider the fact that malapropisms, despite their supposed irregularity, possess meaning:

> [M]alapropisms *and the like* (…) introduce *expressions not covered by prior learning* (…) Malapropisms fall into a different category, one that may include such things as our ability to perceive a well-formed sentence when the actual utterance was incomplete or grammatically garbled, our ability to *interpret words we have never heard before,* to correct slips of the tongue, or to cope with *new idiolects.* These phenomena threaten standard descriptions of linguistic competence *(including descriptions for which I am responsible).* (NDE, p. 437. My emphasis)

Malapropism, along with any idiosyncratic use of words, seems call into question those descriptions of communication which — like the three explanations summarized above — reduce meaning to a restrictive account. Even if Davidson only mentions it in passing (in his critique of the third explanation), it is important to note here that metaphor can be understood to fall within the category of unconventional uses here under discussion. Metaphor invokes — although not always — novel forms of expression, it includes words that can only be properly understood within the context of their utterance, and it sometimes interferes with or suspends grammatical rules. Davidson's avowal of responsibility in the above-quoted passage has, then, to be comprehended as a reference to his own former theses in WMM.

According to the first explanation of first meaning — that which construes it as forming part of a system — the hearer is capable of correctly interpreting a novel use of language only because she is in possession of a finite system of first meanings governed by rules. Davidson partially accepts this theory of intelligibility: the hearer does make use of a finite and recursive set of rules when she encounters novel uses or enters into different contexts. He adds, however, the following qualification: this finite and recurring set can never be known by the hearer in its totality. On the contrary, this system — which Davidson insists on calling a theory — is always geared to the occasion and created *ad hoc*. According to the second explanation of first meaning, the latter forms part of a system that is shared by both speaker and hearer. Davidson does not want to give up this thesis completely: speaker and hearer do share the same «theory» when they communicate — a theory made up not of fixed, context-independent conventions, but of context-adapted rules of use that are perpetually recreated in the act of understanding itself. This sharing of a theory does not amount, Davidson maintains, to the sharing of a language. It merely entails «the interpreter's and speaker's understanding of the speaker's words.» (NDE, p. 438)

If Davidson only modifies the conceptions of first meaning as systematic and shared, he certainly disagrees fiercely with the idea that first meaning is learned by conventions. This explanation has typically been put forward so as to preserve the distinction between literal or first meaning — conceived as furnished by conventions — and the speaker's intended meaning — conceived as resulting from the speaker's whims. As I indicated at the beginning of this section, Davidson also construes the distinction between first meaning and speaker's intended meaning as essential to communication. Nevertheless, if we comprehend this distinction as coterminous with the dichotomy conventional / unconventional we are deprived of an explanation of those unconventional uses of language, such as malapropism and metaphor, where meaning does seem to exist. In order to propose an alternative to the conventionalistic understanding of meaning, Davidson takes recourse to the following example (developed by Keith Donnellan):[14] Jones[15] is an individual who affirms that «Smith's murderer is insane,» thereby meaning that a man, who happens to have murdered Smith, is insane. According to Donnellan, this sentence is true even if the man Jones is thinking of did not in fact murder Smith, as long, that is, as Smith's murderer — whoever this is — is actually insane. A traditional, conventionalistic explanation of Jones's sentence would reject Donnellan's interpretation. It would take him to be implying that the meaning of the sentence is only a matter of the speaker's intention, without any connection with the truth of the «facts». This explanation would counter Donnellan's interpretation by claiming that literal meaning is entirely conventional and that the speaker's initiative is completely arbitrary: given that Jones identified the wrong person, the sentence is false. Donnellan responds to this accusation, however, by pointing to the danger of reducing successful communication to the initiative of the speaker. This initiative is always accompanied by the speaker's intention and expectation that her words be interpreted as they were meant (by what Davidson denominates as «self-referring» intention). And furthermore, Davidson claims, to prevent the danger of the speaker's intention taking over «[a]ll that is needed (…) is a firm sense of the difference between what *words* mean or refer to and what *speakers* mean or refer to.» (NDE, p. 439) Following Davidson's contention, we only have to distinguish in Jones' sentence between the falsity of his intended meaning and the truth of what the sentence actually means. If the murderer of Smith — whoever he is — actually is insane, then, even if Jones wrongly identified this murderer, he still said something true using a false sentence. The first meaning is true; it is only Jones's intended meaning that is false. It follows from this that if meaning were to be determined solely by convention and if Jones were to point at the wrong person, his sentence would be false. The sentence is, however, when the context is taken into account, true. As there is no danger of this truth being left entirely to the whim of the speaker, the idea of convention as a safeguard of meaning can easily be discarded.

In order to develop an alternative account of first meaning and intelligibility, Davidson combines this critique of the third explanation of first meaning with the modifications made to the first two explanations. Meaning should be understood, he contends, as a process: a speaker and interpreter can only enter into communication because each possesses their own respective *prior* theory — a set of beliefs and rules of interpretation that have been learned.

---

[14].   Quoted by Davidson from Keith Donnellan, «Putting Humpty Dumpty Together Again,» *The Philosophical Review*, 77, (1968).

[15].   Jones is the character invented by Wilfrid Sellars to exemplify «philosophical behaviorism» in *Empiricism & The Philosophy of Mind*, Cambridge, London: Harvard University Press, intro. by Richard Rorty, study guide by Robert Brandom, 1956, 1997.

Armed with their respective prior theories, speaker and listener attempt to understand one another. The speaker then utters something with the «self-referring» intention and expectation of being interpreted in a certain way. Let us suppose that this utterance is a malapropism; and that the interpreter's prior theory does not contain its meaning. What the interpreter now has to do is adjust her prior theory so as to incorporate all the contextual information available to her — information that allows her to engage in conjecture concerning the speaker's own prior theory and behavior. Conversely, the speaker needs to include in her prior theory reflections about the constitution of the hearer's, as well as about the hearer's reaction to her speech. When communication succeeds, both speaker and interpreter have substantially modified their prior theories in accordance with this adaptive process. They now use, instead of prior theories, what Davidson calls *passing* theories: «For the hearer, the prior theory expresses how he is prepared in advance to interpret an utterance of the speaker, while the passing theory is how he *does* interpret the utterance. For the speaker, the prior theory is what he *believes* the interpreter's prior theory to be, while his passing theory is the theory he *intends* the interpreter to use.» (NDE, p. 442) Whatever is understood and meant to be understood at the moment of the utterance is the first meaning: this is the only agreement of or convergence between speaker and interpreter. Even if the passing theory has been reduced to such a brief moment and such a narrow field of application, it still has to be denominated, Davidson insists, a theory: «when a word or phrase temporarily or locally takes over the role of some other word or phrase (as treated in a prior theory, perhaps), the entire burden of that role, with all its implications for logical relations to other words, phrases and sentences, must be carried along by the passing theory.» (NDE, p. 443) Because the passing theory that provides access to first meaning is totally geared to the occasion, the boundary between «knowing a language» and «knowing our way around in the world generally» (NDE, p. 446) — a boundary defended by Davidson himself in WMM — is continually eroded in a constant, and constantly unconventional, process of modification. On this new account, communication does not succeed because of any fixed set of linguistic rules, but because we are always (re)learning and modifying our beliefs and modes of interpretation.

Having reconstructed this theory of intelligibility and communication, we can conclude this section by making a brief return to Davidson's reply (in CS) to the thesis of partial untranslatability. If, in any act of successful linguistic communication, speaker and hearer share a passing theory, then any differences of belief or interpretation can only be conceived as differences in relation to this more fundamental agreement. Those who defend the thesis of partial untranslatability conclude from it that differences in belief amount to differences between conceptual schemes. If, however, as Davidson argues in NDE, there is «no general principle, or appeal to evidence, [that] can force us to decide that the difference lies in our beliefs rather than in our concepts,» (NDE, p. 197) then the thesis of partial untranslatability is as philosophically unfounded as that of complete untranslatability.

1. 4 Inversion and Normativity

Davidson's conception of a prior and a passing theory provides an explanation of intelligibility in the extreme case of malapropism. In this section, I want to formulate — on the basis of this conception, and on the basis of the conclusions drawn in CS — a view of metaphor explicitly opposed to the one developed by Davidson in WMM. The first question that needs to be asked is whether metaphor can be said to possess meaning. In order to advance an affirmative response to this question I will seek to refute Davidson's two fundamental presuppositions in WMM: firstly, that of a dichotomy between language and

world, which entails the exclusion of metaphor from language and understanding; and, secondly, that of a restrictive conception of literal meaning. Secondly, I would like to examine again the question of dead metaphor. A transition that was inconceivable in terms of the basic presuppositions of WMM — that from living to dead metaphor — will be seen to become easily comprehensible when looked at from the perspective of NDE. Thirdly and finally, I would like to ask whether a theory that maintains the necessary persistence of literal meaning (the first presupposition of WMM) can open up the possibility of a normative consideration of metaphor. The relation between literal meaning and first or intended meaning can serve as the basis, I want to argue, for such a consideration: the normativity of metaphor arises as a consequence of its being conceived as both understandable and constitutive of language.

The central feature of WMM is Davidson's exclusion of metaphor from the realm of language and understanding. As we have seen, this exclusion results in an epistemological dichotomy between literal, linguistic meaning and metaphorical, non-linguistic and worldly effects. Whereas language is formed by a set of literal meanings that can be learned, the world is the place where language is used by speakers. It follows from this dualistic account that, if language is constituted by conventions or rules that link words and meanings with things and facts in the world, then its structure must be that of a system; and that metaphor, conceived as operating outside the bounds of linguistic convention, must be said to lack any systematic meaning. The external world appears as a reality that, in order to be known at all, requires the imposition of the structure or order of a language. It also follows from this idea that — whether it is maintained that there is only one conceptual scheme or as many as there are languages — the world (or that which we know about it) is relative to the particular conceptual scheme in use. That is, whatever is real with respect to one conceptual scheme might not be real with respect to another. As we have seen, Davidson's article CS demonstrates the inconsistency of this epistemological division between language and world. This division is always articulated by means of the idea that one or many conceptual schemes organize or fit reality. Davidson successfully shows that this separation must imply a thesis of conceptual relativism or linguistic untranslatability, and he demonstrates that, in its two most common formulations, this thesis is indefensible. It can be sustained neither by the claim of linguistic intransitivity, nor by that of an untranslatability of incommensurable point of views, nor by that of an untranslatability tantamount to what Davidson terms the «third dogma of empiricism». Since Davidson's WMM relies on the same dichotomy proven inconsistent in all these cases, we can now affirm — from a perspective worked out by the latter Davidson himself — that its fundamental argument has to be rejected.

In WMM, meaning conforms to what I have termed a restrictive definition; it is conceived, that is, as something context-independent that is learned by convention. In NDE, by contrast, the definition of meaning is expanded so as to include within itself so-called non-conventional uses of language, such as malapropism and metaphor. Davidson here describes intelligibility as a communicative process that takes place between two participants: speaker and hearer. (His description must also be understood as valid, of course, for conversations in which numerous speakers and hearers participate). These roles can be understood as constantly interchangeable: speaker and hearer come to and initiate the conversation by means of a prior theory. This theory consists of a certain amount of information (beliefs about oneself, about the other, about the world, about the particular topic under discussion) and of communicative skills that enable mutual understanding. The interaction proceeds when the speaker and hearer express their beliefs and when each of them modifies their respective prior theory so as to incorporate new information. When speaker and hearer actually succeed in understanding one

another, when a conversation «goes somewhere», it is because they coincide in a passing theory — one which is ceaselessly undergoing revision and being adapted to the new information that becomes available. If the speaker uses a word or utterance that is not part of the hearer's prior theory, for example, he must be construed as having the intention and expectation that this expression be understood in the way that he intends. In the course of the conversation, therefore, or even at the same moment in which she utters the unfamiliar expression, the speaker provides the hearer with the necessary clues about how to interpret it. These clues, and the hearer's reaction to them, now form part of the speaker's own passing theory. In turn, the hearer — who is in principle unfamiliar with the utterance — also assumes that the speaker means something and that what she understands corresponds to this meaning. Because in the course of the conversation the speaker gives the hearer clues about how to interpret the utterance, the latter has the capacity to create her own passing theory. The meaning of the unfamiliar sentence is attained, therefore, when the meaning intended by the speaker converges with the meaning interpreted by the hearer. If, in WMM, meaning is understood to result from the learning of conventions, and metaphor is understood as having no meaning other than the literal meaning of its terms, then, in NDE, meaning becomes possible — with regard to both literal and metaphorical utterances — whenever speaker and hearer converge in a passing theory. Prior theories show us how to interpret utterances in advance; it is the only the process of communication, however — the constant revision of beliefs and modification of meanings — that provides us with the passing theory that enables us to understand novel utterances. Passing theories are always being discarded and recreated as communication proceeds. Nothing allows us to maintain that prior theories are necessarily shared: each participant enters into communication with different knowledge and skills. The only thing that is shared is, then, the passing theory, that is, meaning is only attained when communication succeeds, when the hearer understands what the speaker says. A passing theory is only valid on the occasion in which it is used; it cannot be fixed or detached from the context in which it originated.

In NDE, Davidson employs this theory of meaning in order to explain the linguistic use of malapropism, and in order to conclude that malapropisms possess linguistic meaning. What I want to show now, following Davidson's explanation, is that linguistic meaning can also be accorded to metaphor. With this in mind, I would like to refer back to the example from Donnellan discussed above, in order to advance a comparison of it with the case of metaphor. In this example, Jones utters that «Smith's murderer is insane»; although he is thinking of the wrong person, he nonetheless says something true. This example is both similar to and different from the case of metaphor. Jones does not realize that he is using a sentence with two meanings: his intended meaning being false, since he is thinking of the wrong person; and the first meaning of his sentence being true, since Smith's murderer is insane. In the case of metaphor we are also dealing with a sentence with two meanings, but in this case the sentence is uttered deliberately.[16] Metaphor consists of an intended first meaning — which has to be understood as such by the hearer — and a second meaning — which also has to be understood as such, that is, as non-intended and literal. Davidson's explanation of the case of Jones can therefore be extended to metaphor so as to propose the following thesis: in metaphor, the dichotomy between first meaning as conventional and literal and second meaning as unconventional and non-literal is inverted. The hearer understands the first and intended meaning as such, even if it is unconventional and non-literal. To understand and

---

[16]. Only, that is, when it is realized to be a metaphor so interpreted.

accept a first meaning does not mean, however, to erase the second, non-intended meaning. For the second meaning — commonly called literal — is also understood by the hearer: it remains active. In contradistinction to Jones's statement, however, where first meaning and intended meaning did not correspond, in the case of metaphor, the hearer realizes the correspondence between intended meaning and first meaning. This explanation of metaphor maintains the distinction, then, between first and intended meaning, but calls for an abandonment of the dichotomy between the conventional and the unconventional.

It is possible to rephrase this explanation in terms of the distinction between passing theories and prior theories, and in terms of the dialogue between speaker and hearer. I want now to examine this process, firstly from the perspective of the hearer, and secondly from that of the speaker. The hearer's prior theory provides in advance the approximate meaning of the majority of expressions emitted in conversation. When the speaker utters a metaphor — Max Weber's «politics is a strong and slow boring of hard boards»,[17] for example — with the intention that it will be interpreted to mean that politics requires tenacity and steadfastness, it is quite possible that this meaning will not be provided by the hearer's prior theory. The hearer therefore adjusts this prior theory so as to accommodate all the information available to her; what the speaker previously said about politics, what both of them said about related issues, what the speaker seemed to say about, etc. Armed with all the facts and contextual information available to her, the hearer then formulates a hypothesis as to what the metaphorical expression could mean: either that politics is boring (mistakenly taking the metaphor for a pun), that it consists in drilling wood (mistakenly taking it literally), that it is a gradual, step-by-step process or that it requires a lot of effort, etc, etc. The hearer thus manages to interpret the speaker, given all the contextual information, in the way she thinks intended by the expression. This is the hearer's passing theory. As the conversation goes on, however, the hearer's view as to what the speaker intends might be transformed — she might come to interpret the utterance as meaning, for example, that politics requires a charismatic character. This change is only possible because the hearer's passing theory is always geared to the occasion. If we now observe the same process from the speaker's perspective, we can see that, since her intention is to be understood, she tries to give — taking into account her own view of what constitutes the hearer's prior theory — all the information relevant for the understanding of the metaphor. When the speaker's and the hearer's passing theories converge, communication succeeds and the metaphor, rather than merely producing effects or suggestions, is understood.

On the basis of this analysis, we can also provide an explanation of how a living metaphor transforms into a dead one. Let us consider once again the metaphor «role». At a certain point it came to be used outside a theatrical context as a living metaphor. The hearer (or reader) of this metaphor had to realize that, in addition to the literal meaning of its words provided by her prior theory (that is, «part in a play»), the expression had to have — in order even to make sense — another meaning. This other meaning was the one intended by the speaker (or writer) who, most probably in the context of a discussion concerning social relationships, meant to suggest something like «feigned behavior», «proper function» or «expected pattern of action» etc. The hearer therefore acknowledged that this other meaning came first in the order of interpretation, and that the second (literal) meaning, although still active, was not intended by the speaker. It is precisely when this acknowledgment starts to

---

17.   Max Weber, «Politics as Vocation», in *From Max Weber: Essays in Sociology*, ed. by H. H. Gerth and C. Wright Mills, London: Routledge, 1948, 1991, pp. 77-128, p. 128.

occur immediately that the metaphor becomes effectively dead (although it can always be, of course, revitalized). Historically, for instance, the metaphor «role» has been repeatedly employed by different people both in a social and a theatrical context. Whenever somebody hears this metaphor today, her prior theory will tell her that, in a social context (in a sociology classroom, for example), «role» means «appropriate behavior»; and, in a theatrical context (when reading the list of actors in a program, for example), that it means «part in a play».

A number of other implications can be derived from this metaphorical reconstruction of Davidson's conception of prior and passing theories. We can extrapolate from Davidson's analysis, for instance, that in every process of communication the meaning of the words has to be distinguished from the speaker's intended meaning, and that *both* are equally constitutive of understanding. This article, for example, is for the most part written so that the meaning of the words roughly coincides with their intended meaning. Intended meaning and literal meaning can, however, be separated. Consider the following examples. When I use the word «first» in the expression «first meaning», its intended meaning in no way differs from its literal meaning: that which comes before others in a series or order. Even when the convergence between intended and literal meaning is obvious, this is only acknowledged in the context of its occurrence: it is intended here that «first» has a selecting function, not a temporal connotation (as in «the first female politician to gain power»). The literal meaning of the word «first», therefore, although it certainly belongs to the reader's prior theory, already presupposes an accommodation to the specific context of its use, that is, it already requires a passing theory in order to be understood. A passing theory is an unavoidable requisite whenever expressions appear the words of which do not correspond to the intended meaning given them. If a metaphor is employed, then the reader, in order to understand it, has to be able to distinguish between the literal interpretation of its words and their intended first meaning. If specific philosophical or political terms and expressions are employed, it is most probable that these words or expressions possess other meanings when used in other contexts and by other people. A field linguist, an English teacher, a multiculturalist and a Davidsonian, for example, would offer very different interpretations of Davidson's sentence «there is no such thing as a language». If the context is a quote from one of Davidson's articles (NDE, p. 445), then the interpreter will (hopefully) receive the necessary clues as to a successful interpretation with the help of my reconstruction of Davidson's article. These will allow her to modify her prior theory, to discern what Davidson intends the sentence to mean, and to decide whether she agrees with it or not. In each of these cases, the intended meaning must be distinguished from the meaning of the words that constitute the expression: Davidson does not want to claim that language does not exist, but that those traditional theories of language which claim that meanings are always furnished by conventions are wrong. Davidson's dynamic conception of meaning purports to claim that — in the case of metaphor and other linguistic uses, as much as in the case of expressions that simply mean what their words literally mean — communication can succeed. Communication succeeds when both speaker and interpreter are capable of distinguishing between the literal meaning of words and their intended meaning, of acknowledging when these two meanings coincide and of making clear (speaker) and ascertaining (hearer) the intended meaning when they do not coincide. When first meaning and intended meaning do coincide, there is no need for a complex mechanism of interpretation; all that is required is an implicit commitment both that the speaker means what the words that she uses says and that the reader acknowledges this. When first meaning and intended meaning do not coincide, speaker and interpreter have to modify their passing theories so as to render them convergent; these modified theories will then give them the new

intended meaning. The literal and the metaphorical therefore belong to a continuum of meaning and use — one in which every employment of language, even if it is transformed by its intended meaning, remains comprehensible so long as it answers to what the words literally mean.

It is possible to derive from this account a final insight into the minimal normative constraints inherent in all communication. For if a speaker, in order to communicate with others, has to commit herself to making her utterances understandable, she must employ words with a determinate literal meaning, and introduce enough contextual information so as to convey what she intends these words to say. To make herself understood, therefore, a speaker has to both abide by the usual meaning of words, and commit herself to showing how these meanings relate to the intended meaning introduced into them. A hearer, conversely, if she wants to recognize and understand what is being said, must abides by the literal meanings of the words and follow and ascertain those hints of the speaker that relate these words to their intended meaning. Understanding compels both an acknowledgment of the literal (usual) meaning of words and an endorsement of their intended meaning. (We can say the same thing when a new word, unknown to the hearer, appears in a conversation. This word's meaning will only be understood if the speaker provides enough clues with regard both to what the word literally means and is intended to mean, and if the hearer acknowledges that both of these meanings coincide. That both literal and intended meaning are constitutive of communication is revealed by the fact that, when a new word is introduced into a discourse and is intended as a metaphor, the hearer can in no way ascertain this metaphoricity unless she also understands the literal meaning of the word). Communication is subject to an inherent normativity: it necessarily depends both on the speaker's and hearer's commitment to the literal meaning of the words employed, and on their endorsement of the intended meaning as that which is actually meant. The inherent normativity of language is nothing more than the mutual presupposition of, and unavoidable commitment to, both literal and intended meaning; a presupposition and commitment to a sense of both their inextirpable difference and necessary connection. Without committing themselves to the literal, speaker and hearer simply cannot understand the usual reference of words; without committing themselves to the intended meaning in its necessary relation to the literal, speaker and hearer cannot understand one another on each new occasion of utterance.

If we accept, in contradistinction to the Davidson of WMM, that metaphor forms a part of language, then this account of the inherent normativity of language must of course be seen to compel an insight into the intrinsic normativity of metaphor. Metaphor is use of language in which the literal meaning of the words is related to, but does not coincide with, their intended meaning. Because the speaker wants to be understood, she commits herself to using words with a known literal meaning; but because she also intends to say something else with these words, she gives enough contextual information about how to relate this intended meaning to the literal meaning. And, from the hearer's perspective, metaphor is understood both because she recognizes the literal meaning of the words and because, with the help of the contextual information provided by the speaker, she acknowledges their intended meaning. It can consequently be maintained — in contradistinction to conceptions of metaphor as non-constitutive of language, or as a deviant and merely ornamental use of language that can be paraphrased — that metaphor is subject to a properly normativity. Like any other form of linguistic communication, metaphor makes use of literal meanings; and, when it is recognized as metaphor, it deliberately and explicitly formulates the speaker's intended meaning in a discursive manner. In order even to be understood, metaphor must respond both to literal

meaning and its relation to intended meaning. As against Davidson's conception of it in WMM (as the mere producer of extra-linguistic effects), metaphor can be said to possess a meaning, and to be inherently normative, because of the relationship that it establishes between the literal and intended meaning of the words that constitute it. Once we have refuted Davidson's earlier thesis according to which metaphor has no meaning other than a literal one, we can therefore term the intended meaning of metaphor «metaphorical». Metaphor has an intended metaphorical meaning that can only be comprehended in relation to the literal meaning of the words that constitute it. The linguistic normativity of metaphor resides precisely in this relation.

Sonia Arribas
Instituto de Filosofía (CSIC)
<arribas1234@yahoo.es>

THE PRIVATE LANGUAGE ARGUMENT ISN'T AS DIFFICULT, NOR AS DUBIOUS AS SOME MAKE OUT

Roger Harris

Introduction

The sections of Wittgenstein's *Philosophical Investigations* which contain the Private Language (PL) Argument are dense, cryptic and wide ranging. I argue that a specific argument against a private language can be distilled from the text that is less involved and obscure than is often supposed in the immense secondary literature. It is also far less self-contained and isolated from the mainstream of philosophy than many make out, including Brian Garrett[1] and Michael Ming Yang[2] in recent papers in this journal. It can be distinguished from arguments about rule-following, pain and the problem of other minds, forms of life, etc. and, as I have framed it, avoids Garrett's objections. Moreover, a number of would-be conclusions Garrett takes its proponents to draw from the argument, far from depending on it, are long standing[3] positions, on some of which the PL Argument itself depends.

§1. Privacy

Although, for Garrett, «It is crucial to be clear about the sense of 'private'» invoked by Wittgenstein, privacy remains virtually unexamined in his paper. These days, privacy is best discussed using some terms which have become current since Wittgenstein wrote. The realm of necessary privacy within which, Wittgenstein maintains,[4] it is not possible to follow a rule, and so not possible to have a language, is the realm of immediate first person mental life. I.e. immediate phenomenal episodes of sensory experience, memory, and the other aspects of one's mind which one owns exclusively. These are available only through introspection,

---

[1]. 'Cartesianism and the private language argument' *Sorites* **14** Pp. 57-62, 2002.

[2]. 'Privacy, Individuation, and Recognition' *Sorites* **13** Pp. 90-98, 2001.

[3]. I would say that they were *uncontroversial*, were it not for what you might call the 'Monty Python proviso' after the Python's brief sketch which went something like:
«'This is Mr. Smith, who contradicts everything anyone ever says.'
 — 'Oh no I don't!'»
It is in the nature of philosophy that every philosophical thesis has been denied by someone, somewhere at some time.

[4]. I am not going to cite textual support for every remark I make about Wittgenstein's argument, but, where I do, I will refer to his *Philosophical Investigations* by the initials 'PI' and the paragraph number preceded by '§'.

judgements about them are wholly subjective, and their properties are the *qualia* that have subsequently been presented as posing the 'hard problem' of the mind/body relation.

Wittgenstein has been taken by some to deny the reality of subjective inner mental life. I think he does not, and *must* not, if his PL Argument is to succeed, so I will devote a few paragraphs to this issue. His discussion of inner mental life turns in part on the truism that you cannot feel my pain. This exemplifies the necessarily private content of that realm of exclusivity which is the inescapable corollary of my experiences and thoughts being *mine* and not yours, and vice versa. This sphere of necessary (rather than 'logical') privacy is delineated by that proprietary possibility, namely that experiences, thought and whatever else comprises an inner mental life, can

- be *owned*,

- be known only by introspection, and

- only be reported subjectively.

This determines some key features of privacy and subjectivity. Let's look at privacy and introspection first (I discuss subjectivity in §4. below).

If *qualia* are the properties e.g. of immediate phenomenal episodes, then, just in so far as they can be owned, they must possess aspects which are not shared, and which therefore cannot enter any arena of public comparison. This does not make *qualia* ineffable, however. We can refer to them, and describe them in detail. So you cannot feel my pain, nor I yours, but, as a diagnostician, for example, I can perfectly well appreciate the finer points of your pain as pointers towards the pathology I investigate. Physicians undoubtedly enjoy a huge advantage over veterinarians in virtue of our well developed phenomenological vocabulary for pain. Wittgenstein takes the fact of public talk about private pains to show that pain descriptors cannot get their *senses* from episodes accessible only to introspection.

§2. Public Sense but Private Reference

Garrett maintains that the PL Argument:

> has been thought to undermine the Cartesian Model of how natural language sensation words get their meaning. 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.

In reality the Private Language Argument is not needed to undermine this model. It suffices merely to cite an example — 'tongle'. This expression applies to a series of my mental episodes with which I am acquainted and you are not, and necessarily cannot be. I am sure you do not understand the expression 'tongle'. It has been customary since Frege to distinguish what an expression conveys — its 'sense' — from what it picks out — its 'reference', because 'meaning' is ambiguous in this regard. There is a flat contradiction in supposing that what an expression *conveys*, when it is used by you and I, and speakers of English at large, is something which is not equally accessible to all who use it. Since I cannot feel your pain and vice versa, *what you feel* when you have a pain cannot qualify as what *I convey* by describing *my* pain, but no more can what *I feel*, since that is not accessible to you. But even

if it is only the reference of 'tongle' that it inaccessible, as I «define it with reference to my» tongles, this 'definition' on my part is not a success, for you are none the wiser, any more than Alice understood Humpty Dumpty[5] when he reassured her that

> «When *I* use a word,» Humpty Dumpty said in a rather a scornful tone, «it means just what I choose it to mean — neither more nor less.»
>
> «The question is,» said Alice, «whether you *can* make words mean different things.»
>
> «The question is,» said Humpty Dumpty, «which is to be master — that's all.»

At the risk of seeming glib, I say that these expressions, attributing pain and the like, have a *public sense* but a *private reference.*[6] *What it is (privately) to have them*, in such a way that they are your pains and no one else's, cannot provide the *public senses* of the expressions employed in the phenomenological vocabulary for pain. Humpty Dumpty said words mean what he chooses them to mean. No they don't. If they didn't *already* mean what they do mean, there would be *no meanings* for him to choose between. 'Headache' is a word in English, and I learn from others that it has a sense which allows me to refer to my private sensations.

In support of his defence of the Cartesian model Garrett cites Craig's claim that an expression's being used to refer to what can only be known to the speaker doesn't stop its being understood by others. That is no support: Garrett can *refer* to his own headache without his own headache being the *sense* of the expression 'headache' as he uses it, because he speaks English, and 'headache' *already has* a sense and a definition. If you understand the English word you are able to refer to the private pains of whomsoever you please, including yourself. Craig's claim is equally compatible with what I have claimed.

What makes this possible? The answer can only be "the public sense of the expression 'headache' which we all use." This cannot be an item known only to Garrett by introspection. An expression could not be generally intelligible if it were generalisable over nothing but a plurality of private mental episode tokens accessible solely to the introspection of a single individual. Applicability both to mental episode tokens to which you do and to those to which you don't have immediate access is a necessary condition for the intersubjective intelligibility of any mental content descriptor we might employ. So the PL Argument is not needed to dispense with private senses. These can quite independently be shown to be impossible.

§3. Private Facts

I wrote above of 'what it is (privately) to have' pains because I think their necessary privacy crucially undermines the more usual locution 'what it is like…' to have a pain, be a bat, or whatever. Suppose you grant that the realm of necessary privacy is defined by ownership and accessibility to nothing but introspection, so its contents, *qualia* are consequently excluded from any arena of public comparison. If so, then the only relations of

---

[5]. Lewis Carroll *Alice Through the Looking Glass*, http://sunsite.tus.ac.jp/coll/alice/alice_toc.html.

[6]. So, on this, I disagree with Michael Ming Yang. I think the impossibility of following a rule in a sphere known only by introspection does not result from the absence of identity criteria for e.g. pain episodes. In this context Ming Yang writes «What counts a kind, a type, an identity criterion, is hinged upon our way of life.» This perhaps confuses type identity with token identity (since it is to the latter that the slogan «no entity without identity» applies, which Ming Yang invoked). In any case, doctors can identify my pain tokens by extraneous criteria (like start and finish times), and my pain types from their diagnostic manuals.

'likeness' into which such *qualia* can possibly enter must be entirely confined to that necessarily private realm, unless these are not simple likenesses between monadic properties of *qualia* at all, but involve *relations* they bear (or are disposed to bear) to other *qualia* and to properties of items *outside* the private realm. I argue that this is just what they do involve.

If the individual qualitative character of *qualia*, disclosed by introspection, cannot belong in an arena of public comparison, then discussions of what they are *like*, and of possible undetectable 'inverted spectra' of *qualia* must surely be beside the point. Your *qualia* are not open to inspection by me, nor mine by you. The respects in which your *qualia* can resemble mine therefore cannot be those qualitative respects available only to you as the 'owner' of those *qualia*. So no question of qualitative resemblance between your individual *qualia* and mine can arise, because the individual qualitative character of *qualia* cannot enter the arena of intersubjective comparison. By the same token, then, no question of qualitative *difference* can arise either. So there can be no 'fact of the matter' about inverted spectrum claims. This is why Wittgenstein departs from a line of discussion that began with Locke[7] when he says, of his 'beetle in a box' that «It is not a *something*, but not a *nothing* either» (PI § 304). Your 'beetle' is not open to inspection by me, nor mine by you.

That is not to say that your *qualia* do not resemble mine in any respects — indeed, there must be respects in which they do. In order for us to speak of us both seeing crimson, yellow or turquoise, your *qualia* must resemble mine in respect of the *relations* they bear (or are disposed to bear) to other *qualia* and to properties of things and events in the public realm. Your perceptions and mine could not otherwise *coincide* when it comes to reacting to, acting on, or assenting to descriptions of what we perceive. So we can say that each one of my colour perception *qualia* and of yours (if you and I have accurate colour vision) belong to *relational resemblance classes*, which contain the relevant *qualia* of everyone with normal colour vision who lives in a polychromatic world. Those people whose discriminations are faulty — (can't tell red from green, say) lack *qualia* in the relevant relational resemblance classes.

"Aha!" the Cartesian will say "then those who can't tell red from green don't know what 'red' and 'green' mean, so the meanings of 'red' and 'green' must be the inner states possessed by those who *can* tell them apart." Indeed, Garrett writes:

> 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.

There is no 'so' about the identification of the senses of 'red' and 'green' with inner states. Let me take the suggestions in the previous paragraph a little further. Delineating the realm of the private turns on the distinctness of my inner life from yours, and this consists precisely in your not being privy to the private aspects of my *qualia*, and vice versa, since no properties of a private episode are open to inspection by any other. So what your experience of seeing the colour red has in common with mine cannot be a monadic property of that private phenomenal episode, (what it is for me to see something red).

---

[7]. Locke, J., 1689/1975, *Essay Concerning Human Understanding*, II, xxxii, 15, http://humanum.arts.cuhk.edu.hk/ Philosophy/Locke/echu/

Garrett has a further argument to fulfil his undertaking:

> Further, this [Cartesisn] 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.

This turns on the claim that a confusion is involved between qualitative and numerical identity when we say that you cannot feel my pain. You cannot feel my present pain token, because it is mine, but that does not show that you cannot have the same pain type as me. Indeed, it does not, but nor does it refute my earlier argument. Namely, «If *qualia* are the properties e.g. of immediate phenomenal episodes, then, just in so far as they can be owned, they must possess aspects which are not shared, and which therefore cannot enter any arena of public comparison..» If God could look into both our minds, He might see that our pains were of exactly the same type. However, since *neither of us* can do this, whether or not our pains are of the very same type must be irrelevant to the way senses are acquired by the terms used in our phenomenological descriptions of pain.

The resemblance your experience of seeing the colour red has to mine can, therefore, only consist in relations our experiences bear (or are disposed to bear) to other *qualia* and to properties of things and events in the public realm (e.g. to red and non-red things). The fact that our discriminations *coincide* is crucial (this is Wittgenstein's PI §242 «agreement in judgements»), because the *agreement* that something is red rests on this. This is a simple matter in the case of perceiving a colour. (It's much more complicated for other areas of inner mental life, which I won't discuss now.) There is a property of my immediate phenomenal episode with which, in so far as it is *mine* and not yours, you cannot be acquainted. What matters to us both are

- relations that this private property bears (or is disposed to bear) to other *qualia* and to a property of external things and

- that the property of your equally inscrutable immediate phenomenal episode enters into the same relations.

If I am colour blind and don't have the relevant *qualia* then I don't have qualia that enter into the relations to one another and external things that are relevant to the correct use of 'green' and 'red', and my judgements do not coincide with yours.

This is not behaviourism, because our immediate phenomenal episodes *have to exist* for the key relations to have their *relata*. It is because those episodes are inscrutable to all but their owners that the question cannot arise for you of what it is for me to have them, however closely our judgements coincide, and their private qualitative character can be no part of their public sense. So, without the *qualia* with the necessary relations to red and green things, our colour blind subject knows that there are colours others perceive, but does not know how to use the expressions 'red' and 'green' of red and green things. This is not because the *qualia* are the senses of these expressions, but because it is on the relations those *qualia* have to external things that the agreement in judgements rests that supports the use of (provides the senses for) 'red' and 'green'. The colour blind subject lacks the *qualia* to which 'red' and 'green' *refer*. "So he doesn't know what we mean." Well, yes, but bear in mind that 'mean' is ambiguous as between 'intend to convey' and 'denote', and in this case it is the latter which is lacking an object for the colour blind subject. This account, notwithstanding its sketchiness, has the virtue that, unlike the Cartesian model, it is not *impossible*.

Moreover, it is only because our judgements coincide, and we *agree*, that we can overcome subjectivity and distinguish *judgements* from fancies, illusions dreams and hallucinations which are not necessarily inwardly phenomenologically distinct from veridical perception. Clarification of the conditions for the possibility of this agreement is the outcome of the Private Language Argument.

This agreement, however, falls far short of epistemic warrant, since speakers notoriously agree on superstitions of every kind. Collective use of a linguistic idiom rests can objectively rest on agreement independently of the objectivity of the subject matter that idiom is used to describe. 'Language game relativism' arises from failure to mark this distinction. The possibility of agreement revealed by the PL Argument provides the justification for the general strategy involved in describing 'language games', rather than the PL Argument *requiring* this strategy for its vindication. On this I part company with Ming Yang over what depends on what in Wittgenstein's skein of argument.

Now, given all that, one might ask, if there is already an argument against private meanings, why do we need an argument against a private language? The arguments in §2 & §3 above, however, show that the senses of expressions in a *public language* cannot be mental items with which people are exclusively acquainted by introspection. The PL Argument takes aim at the idea that such mental items might nonetheless be the meanings in play in a *private language*. To show that this is impossible, it is necessary to show that rules cannot be followed in the necessarily private sphere accessible exclusively to introspection.

§4. Subjectivity and The 'is right / seems right' Distinction

There is more to this distinction than «the unobjectionable [idea] that meaning is normative» which Garrett's concedes. Indeed, it is independent of and necessarily precedes considerations of normativity. It turns on contrasting a realm only available to introspection with a sphere that embodies multiple perspectives, where the former is characteristically *subjective*. It is because it is necessarily solitary that introspection can only disclose what *seems* to be. Solitary animals and persons can check how things seem to one of their senses by bringing others into play — e.g. by touching, smelling, tasting or listening to what they see, but, once they have done that, all they can know is how things *seem* to them.

To contrast how things seem with how they really are requires subjective and objective (or at least inter-subjective) standpoints to interact. So long as there is only necessarily solitary (i.e. private) knowledge, disclosed solely by introspection there can be no distinction between how things seem and how they are. For you, if the taste of liquorice seems nice, it is nice — a paradigm of subjectivity since Hume (if not Hobbes).

It has sometimes been charged that Wittgenstein relies on simple (trivial or avoidable) memory scepticism in order to press for this conclusion. In fact he doesn't cast any doubt at all on memory *per se*. The issue, rather, is that recognition and memory are not *two separate faculties* so that one could serve as a check on the other. How do I recognise some sensation S on 2nd and subsequent occasions when I have it? Only by remembering previous occurrences of S. So, were I to go back to 'consult my memory' to check I have remembered S correctly, I would not find a second inner authority distinct from the faculty I had already employed to recognise S in the first place. Hence the remark (PI § 265) that proposing this is «As if someone were to buy several copies of the morning paper to assure himself that what it said was true.» The point is that there can be no 'is right / seems right' distinction within what is disclosed by introspection alone.

Descartes notoriously bootstrapped his way out of this problem. He sought to institute an 'is right / seems right' distinction by invoking the necessity of an alternative, infallible Divine standpoint as a source of 'clear and distinct ideas' to contrast with the possibility that his own might have been wholly deceived by the demon. Wittgenstein is not concerned with being right *tout court*, but only with the possibility of following a rule, which requires an 'is right / seems right' distinction to be applicable to putative instances of rule-following.

Nonetheless, there is nothing new about Wittgenstein's implicit assumption that, for knowledge necessarily restricted to introspection alone, there can be no 'is right / seems right' distinction. What introspection discloses is subjective. Latter-day Cartesians[8] have sometimes assimilated the impossibility of being mistaken about what is immediately disclosed by introspection to the guarantee supplied to clear and distinct ideas by their divine origin. Hence the notion that sense data supplied incorrigible knowledge. However, the impossibility of being mistaken about sensations, or whatever else is immediately disclosed by introspection, far from being a form of infallibility, is a defect consequent upon the absence from that sphere of any 'is right / seems right' distinction.

You *can't* be wrong about what is immediately before your mind, but no more can you be right, because the distinction has no application to the necessarily private sphere, any more than it has to clouds, birdsong or the stars. Where it is *impossible* to be wrong it does not follow that one *must* be right, it follows, rather, that in such a context no distinction between correct and incorrect performances can apply. So «what seems right is going to be right» and the necessary conditions are not met for following any normative rule which distinguishes correct from incorrect judgements or performances.

§5. Rule Following and Normativity

What are those necessary conditions and why are they not met? The absence of an 'is right / seems right' distinction debars rule-following because the possibility that a rule be followed requires that a *mistake* be equally possible. This is another longstanding position on which the PL rests. From Kant's argument, now known by the motto 'ought' implies 'can', a more general feature of normativity can be derived:

'ought' implies 'can', and so too does 'ought not'.

If we think of any prescriptive sentence expressing a normative rule (including those of language) which distinguishes correct from incorrect performances, we find an implicit notion of its 'normative content' (by analogy with the notion of the 'cognitive content' of declarative sentences). For there to be a rule expressed by prescriptive sentence (for it to have any normative content), both what it enjoins and what it forbids must equally be possible. A sentence can have no normative content if it states a putative rule that enjoins what no one can do, or, equally, that forbids what cannot be done in any case. It would be entirely empty and irrelevant to any possible performance: i.e. equally irrelevant both to how the performance might be produced and to how it might be judged.[9]

---

[8]. Adherents of a Cartesian theory of mind, that is, rather than of his rationalist epistemology.

[9]. What some call the 'rule-following considerations' are often lumped in with the PL Argument, because Wittgenstein mentioned them in the same tract of the *Philosophical Investigations*. The impossibility of rule-following within the necessarily private sphere is quite distinct from the question whether *public* rule following can ever successfully be captured by determinate formulations of rules. Wittgenstein casts doubt on this because

§6. The Private Language Argument Reconstructed

Now we have all the elements on which the PL Argument rests:

- **Privacy premise**: there is a sphere of necessary privacy defined by the possibility that experiences, thought and whatever else comprises an inner mental life, can be exclusively *owned*, and possess aspects known only by introspection.

- **Subjectivity premise**: it is *impossible to be wrong* regarding what is solely disclosed by introspection, but it does not follow from this that one *must* be right, it follows, rather, that, in such a context, no distinction between correct and incorrect performances can apply.

- **Normativity premise**: only in a context where following or breaking a rule are equally possible can a rule be followed in producing or judging a performance to be correct or incorrect.

None of these claims, as I see them, is a novelty introduced in the *Philosophical Investigations*. What is novel, in relation to a private language, is to combine them so that if:

1. the privacy premise applies to our inner mental life, and

2. the subjectivity premise applies to introspective knowledge, then

3. the necessary condition for rule-following (and, hence, a language) set out in the normativity premise cannot be met within the private sphere of our inner mental life.

The key challenge that Garrett poses for this argument concerns *contingently* private languages, for it is not possible successfully to argue that all languages are necessarily public. A single person might invent a language like Esperanto, or the private code in which Samuel Pepys wrote his diary. Garrett also states that

> …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.

These two sorts of contingent privacy are very different, and I doubt that a 'socially-isolated-from-birth Crusoe' could devise a language all by himself. I won't press this point, however, so let's suppose he could. Garrett's key point is that the limitations of actual privacy upon language should be the same whether or not that privacy is necessary. He takes the key steps of the PL Argument to be:

> (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

---

(a)   any explicit formulation always requires a further rule for its application, to create a regress of rules for following rules which must finish with the last rule followed «blindly» *PI* §§217 — 219.

(b)   there cannot help but be innumerably many such rules, and no end to the uncertainty as to *which* precisely is the unique rule being followed in the unfolding performances:
This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord not conflict here. PI §201

I will say no more about this, other than to reiterate that the impossibility of following a rule within a necessarily private sphere is a completely separate matter from the conditions surrounding following public rules.

(3) There cannot be a private language.

and argues that (2) must be «indifferent to whether the private language is contingently private or logically private.»

I hope that my reconstruction of the PL Argument shows that it is not privacy *per se*, but the restriction to knowledge of one's inner life by *introspection* which is responsible for the absence of the 'is right / seems right' distinction. So it is merely within the sphere of inner mental life known only by introspection that a private language is impossible. The nature of actual privacy is, therefore, not indifferent to whether it obtains necessarily or only merely contingently. It is rather like ownership: you can transfer the ownership of your house to me, but, while you live, neither your body nor your mind can similarly become mine if deeds of ownership are exchanged.

§7. Is the Characterisation of Normativity Inescapably Modal?

This, however, is not enough to dispose of the root of Garrett's objection, namely that we cannot require a modal criterion be met in order to establish an actual fact. He writes:

> …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.

For my reconstruction of the PL Argument relies no less on a modal criterion — the normativity premise above. There are several possible responses to this. I think that, despite its apparent reasonableness, we will have to abandon the presumptions in the quote above and, with them, the more general presumption that we can never require a modal criterion to be met in order to establish an actual fact.

The other possible responses are.

1. Sever the link I assumed when I wrote of 'the necessary condition for rule-following (and, hence, a language)'. Perhaps a language can be spoken without any issue ever arising regarding correct or incorrect use of its expressions. (But Garrett himself concedes the normativity of language.)

2. Devise a non-modal characterisation of normativity so as to do away with anything resembling the notion of the 'normative content' of a rule I sketched above. (This will presumably show that 'ought' need not imply 'can'.)

One or other, or both of these two possible philosophical challenges may well be met and overcome. I am not going to hold my breath. Instead I will briefly consider the implications of an inescapably modal criterion for normativity, and an inescapably normative characterisation of language.

Let's return to Garrett's claim above. If Crusoe's language can be contingently private, as Garrett claims, that can only be because it is possible to translate it into ours. I.e. if his language is only contingently private there is a possible world in which I, too, can speak 'Crusoese'. Since I speak English, I must, in that possible world, be able to translate his language into English. There would be no grounds to believe that I spoke Crusoese if I could not and, indeed, no grounds to believe Crusoese was a *bona fide* language if its translation into English was not possible. It wouldn't be a necessarily private language if it couldn't possibly be translated, it wouldn't be a language, period. These considerations are quite independent of the PL Argument.

It is false, therefore, to claim that «His speaking a language cannot rest …. partly, on such modal considerations.» Consequently «intrinsic facts about him, his world, and his intentions» cannot suffice for it to be true of Crusoe that he speaks a language. Suppose, furthermore, that Crusoe has taught the parrots on his island to parrot words in his contingently private language. The difference between Crusoe and the parrots would be that, unlike the parrots, «we might understand and correct him».

In fact, there is also the possibility that he might understand us and correct *himself*. There is something profoundly wrong with the customary[10] conventionalist 'community' account of the source of correction in rule following — i.e. that following a rule is only susceptible to external correction. Self-correction (reflexive imputation of a rule) must be no less important to rule-following than correction by others. Two people in dispute over how to follow a normative rule correctly would otherwise be compelled each to say to the other: "You can, and I cannot tell whether I am following this rule correctly, but equally, I can, but you cannot tell whether you are correctly following the rule you seek to follow in correcting me," which I hope is plainly absurd. The issue here is not the epistemic question '*How* is a correct judgement made?' (This depends on the sort of rule and practice involved). It is, rather, what the necessary conditions are for the possibility of there being a distinction between correct and incorrect judgement.

The answer must be a form of agreement which two or more people can have reciprocally with one another. (The reciprocity is all that can debar the absurd scenario of the previous paragraph.) The sort of agreement at issue here is one which may be granted, refused, broken, honoured, entered into and withdrawn from. It is a normative agreement.[11] It is not a convention. A convention assigns two or more items (of a general kind) to some arbitrary relation, which could as easily be otherwise (drive on the left, use an expression or a sign in a certain way, etc.). In order for it to be possible to institute a convention there needs to be a prior, non-arbitrary agreement as to which items are to participate in the arbitrary, conventional relation.

Most importantly, however, this sort of agreement only has a point when it obtains between two or more individuals, since it is perfectly empty to suppose that anything can turn on your «agreeing» in this sense with yourself. This is why it cannot consist solely in some sort of relation of similarity or concomitance. The reason why not is to be found in PI §215 where Wittgenstein writes, of following a rule in the *same* way on more than one occasion:

> 'But isn't *the same* at least the same?' We seem to have an infallible paradigm of identity in the identity of a thing with itself. I feel like saying: 'Here, at any rate there can't be a variety of interpretations. If you are seeing a thing you are seeing identity too.'
>
> Then are two things the same when they are what *one* thing is? And how am I to apply what *one* thing shows me to the case of two things?

In short, the relation between two instances, in virtue of which they amount to agreement in following some rule, cannot wholly *consist* in the same relation which one thing must necessarily have to itself — namely, having properties in common. Following a rule must require that properties be shared by a *plurality* of instances, but enumerating those properties

---

10. E.g. in Kripke S. *Wittgenstein on Rules and Private Language* Oxford: Blackwell 1982 amongst many others.

11. A 'contract', if you like, in the philosophical sense we derive from contractarianism.

presupposes, and does not explain, what makes them the *right* properties to have picked out in order to identify successive instances of following that rule.

Intrinsic similarity, then, is a necessary, but cannot be a sufficient condition for agreement upon a rule. Thus, if speaking a language does depend on agreement upon some rules, then the intrinsic character of the speaker and his/her world and intentions cannot comprise a set of sufficient conditions for the existence of that agreement, and, therefore, the fact that a language is being spoken.

§8. Conclusion

I hope I have shown that the argument that there can be no private language rests on three unsurprising, far from novel premises concerning privacy, subjectivity and normativity, respectively. These premises stand:

i.   so long as it is possible coherently to refer to items in the necessarily private sphere of a person's inner mental life,

ii.   so long as judgements confined to that necessarily private sphere are subjective,

iii. so long as both 'ought' and 'ought not' imply 'can', and

iv.   so long as languages embody normative rules.

I think this successfully untangles the PL Argument from the other issues Wittgenstein discusses in connection with privacy, rule-following, etc.: the considerations i.-iv. above are prior to the PL Argument; considerations regarding agreement (in judgements, language games and in forms of life) are consequent upon the PL Argument; and considerations regarding other minds scepticism and the indeterminacy of rule-following are independent of the PL Argument. So I hope I have indicated why Brian Garrett's conclusions do not follow, and, why, while I generally agree with Michael Ming Yang's substantive conclusions, I cannot agree with the route he takes to establish these.

_____

Roger Harris
Middlesex University
<r.harris@mdx.ac.uk>

FACTS, TRUTH, AND REALISM: TOWARD A MULTI-LEVEL THEORY OF KNOWLEDGE

Paul Bowyer

Introduction

It is a striking fact of modern life that the predictions of our best scientific theories in the physical and biological sciences are routinely confirmed to a high level of accuracy. This is what makes our modern technology possible. The question of why scientific theories are able to make such accurate predictions has received considerable attention in recent years (e. g. Psillos, 1999; Sankey, 2001; Stanford, 2003). The purpose of this paper is to suggest a possible answer to that question. It will defend a version of what is called «scientific realism,» which says that theories make accurate predictions because they are approximately true.

Scientific theories are expressions of human thought. It is natural to think of thought as being largely imagination, and of imagination as consisting of images that resemble things in the world. On that hypothesis, theories can predict events by a sort of *simulation*, the way we predict the behavior of a jet plane in flight from the behavior of a model jet in a wind tunnel. But there are convincing arguments that this picture, sometimes called the «copy theory of representation,» is wrong (Cummins, 1989, pp. 27-34; Fodor, 1975, pp. 174-195).[1] There is also good evidence, although in this case it is more controversial, that thoughts are more like *sentences* than images, and that the principle that causes one thought to succeed another is essentially logical inference from premises to conclusions.[2] At the same time, for a theory to make accurate predictions, it must act as a sort of map to reality. We can liken thoughts to the dots on a map, and facts to the cities represented by the dots. Then the principle by which one thought leads to another must parallel that by which one fact leads to another, much as the lines between dots on the map must parallel the roads between cities. This kind of parallel is an *isomorphism*, so that in order to make correct predictions thoughts must be at least partially isomorphic with facts. Thus when we abandon the copy theory of representation, we are forced to search for another explanation of the relationship between thoughts and facts, which can tell us how thoughts or ideas in the mind, being sentences

---

[1]. Fodor (1975, p. 174) gives an excellent summary of the arguments against the doctrine that «thoughts are mental images and they refer to their objects just insofar as (and just by virtue of the fact that) they resemble them.» He does not deny that images may play an essential role in thought.

[2]. This is the «language of thought» hypothesis (Fodor, 1975), which is defended, for example, by Marcus (1998, 2001) and Pinker (1994). For arguments against the language of thought hypothesis, see, for example, Churchland, P. S., 1986, pp. 386-399; Dennett, 1987, pp. 112-116; Millikan, 1995).

connected by logical inference, can be isomorphic with events in the world. The purpose of this paper is to supply that explanation.

I will argue in the remainder of this introduction that thoughts have a special capacity to mirror and anticipate facts because both thoughts and facts are subject to the rules of logic, that is, to the rules by which one idea «follows from» another.[3] But in order to state this relationship in our theory we must be able to talk about facts, and to do that we need to incorporate our present well-established scientific theories into an overarching logical structure, containing a *metatheory* consisting of descriptions of the sentences of those scientific theories. Using this metatheory, we can view facts as sentences related by logical inference. At the same time, the language of thought hypothesis allows us to view the physical states of the human brain as encodings of sentences, where these encodings are related by a physical process that is equivalent to logical inference. Thus from the point of view of the metatheory there is an isomorphism between thoughts and facts, which means that there is an isomorphism between theories and facts, since theories are expressions of thought. We will see that this isomorphism provides a possible solution to certain problems in epistemology, such as the problem of representation (Cummins, 1989), and the semantics of the propositional attitudes (Quine, 1986). Our metatheory also suggests how statements about such concepts as «reference» and «truth» can be incorporated into a parsimonious theory that makes verifiable predictions about observable facts — in this case the fact of the success of science. In the following sections, I will introduce a thought experiment by means of which we can understand in detail the nature of the isomorphism between mental sentences and facts. The last two sections include a brief defense of the theory and the conclusion.[4]

I have given a central role to the relationship between sentences that we call «logical inference.» When we try to get a clear idea of the nature of this relationship, we find that it consists of the process of *formal inference*. Let us remind ourselves how formal inference works by means of a simple example. Consider how we would derive the conclusion «Socrates is mortal» from the premise «all men are mortal and Socrates is a man.» This premise is a combination of «all men are mortal,» which is called the «major premise,» and «Socrates is a man,» which is called the «minor premise.» In formal logic, we would paraphrase «all men are mortal» as «for all values of the variable x, if x is a man, then x is mortal.» We then match the phrase «x is a man» with «Socrates is a man,» assigning the word «Socrates» as the value of the variable «x», and then substituting this value in «x is mortal» to get the conclusion, «Socrates is mortal.» We see that formal logic acts something like an assembly line for sequences of symbols, processing old sentences into new ones according to mechanical rules. We can call such a series of sentences a «chain of inference.» Notice that our explanation of formal logic required us to *describe* the sentence, «Socrates is a man» as consisting of the words «Socrates,» «is,» «a,» and «man,» in that order, since otherwise the phrase «corresponding positions» would have had no meaning. We also needed to know that the symbol «x» functions as a variable. Let us define the «description» of a sentence as the collection of all these statements taken together, that is, the collection of all the various pieces of information that are important to know about a sentence when we apply

---

[3]. For a related argument that sentential processing evolved because it has an adaptive advantage resulting from its correspondence with the structure of nature, see McGinn (1995).

[4]. There is not sufficient space in this paper to both explain the theory and give a proper defense of it. I will therefore concentrate on the explanation and provide arguments wherever possible.

formal inference to it. Let us also define the phrase «higher logical level» as follows: one sentence is at a higher logical level than another if the first sentence is part of a logical system that contains a description of the second. Thus the sentence, «'Socrates is mortal' consists of the words 'Socrates,' 'is,' and 'mortal' in that order,» is at a higher logical level than the sentence, «Socrates is mortal.»

If we want to develop a theoretical approach that can deal with the isomorphism between theories and facts, we must be able to refer to such things as «theories» and «facts,» for the same reason that we must be able to refer to atoms and electrons if we wish to discuss chemistry. We can think of facts as whatever it is that causes our experience. In science, they are whatever makes our experiments and other observations turn out the way they do. It is a difficult question how to refer to facts in a theory (we will return to this question later), but let us provisionally identify facts with the sentences of an idealized theory of nature — the ultimate theory that accounts for all observations (see Peirce, 1878/2001). When we need, for expository purposes, to identify particular facts, we will equate them with sentences of our well-established physical and biological theories, since these are the sort of thing that, in everyday life, we provisionally regard as facts. Because our metatheory consists of descriptions of those sentences that we have chosen to call «facts,» we are thus able to refer to facts in our metatheory. Notice that scientists, when they criticize and test their theories, must form descriptions of the sentences of those theories in order to determine whether those sentences are consistent with each other and what their logical implications are. They must do this whether or not they are clearly aware that that is what they are doing. Thus our metatheory resembles a formal statement of what scientists say about their theories when they criticize them. For example, one of our physical facts will be «electrons have a negative charge,» which we can translate into a somewhat more formal language as «charge electron negative,» putting «charge» as the predicate and «electron» and «negative» as the arguments. Now we can say that the description of this sentence in our metatheory will include the sentences, «The sentence 'charge electron negative' consists of the symbols 'charge,' 'electron,' and 'negative,' in that order,» «the predicate of the sentence is 'charge,'» «the first and second arguments of the predicate are 'electron' and 'negative,'» and so on.

We have seen how we can talk about *facts* in our metatheory, but we also need to be able to talk about *theories*. In this connection the word «theory» refers, not to those idealized statements that we regard as facts, but to hypotheses that we might want to test. Let us refer to speculative hypotheses of this kind as «thoughts,» since they are the thoughts of scientists trying to understand the causes of some set of phenomena. What role should thoughts play in our overall theoretical structure? I will adopt here the «language of thought» (or LOT) hypothesis (Fodor, 1975), because I think it is by far the simplest explanation we have of human behavior. According to that hypothesis, as I will use it here, the cognitive states of the human brain can be thought of as sentences in a mental language, or «mental sentences.» But if this view is correct, then there are physiological states of the brain that can be described at a certain level of abstraction as sequences of symbols that form sentences. (In much the same sense we refer to certain states of the memory of a computer as encodings of sentences.) Since descriptions of the state of the brain belong to neurophysiology, which is part of biology, a statement like, «Einstein believed that electrons have a negative charge» is really a statement about the structure of Einstein's brain, and so is part of the logical system of statements that constitute the science of biology. Here the sentence, «Electrons have a negative charge,» is being used to identify a biological state of Einstein's brain.

Because we know so little about how information is encoded in the human brain, we need to turn to computers to see how it is possible for a sentence to identify a state of a physical system. We can encode the symbols in a computer's memory as sequences of «on» and «off» settings of electronic switches. For example the letter «A» might be encoded as the pattern «11000001,» where «1» means «on» and «0» means «off.» By arranging these encodings of letters into longer sequences we can encode words and other symbols, and whole sentences. If an engineer wants to describe a certain state of the computer's memory, she can express it as a certain pattern of ones and zeroes, but she can also express it, for example, by saying that it contains encodings of the words «charge,» «electron,» and «negative,» in that order, and that «charge» is the first symbol in the sequence, and so on. In other words, the engineer describing the state of the computer's memory can use very much the same types of expressions as our metatheory uses to describe the facts. Furthermore, if our computer is programmed to carry out the procedures of formal inference as described in the «Socrates is mortal» example, then one of the principles by which its internal states succeed one another will be formal inference. And according to the view presented in this paper, the internal states of the brain that we refer to as *thoughts* are analogous to the internal states of our computer, and so they also sometimes succeed one another according to the rules of formal inference. But since we view facts as sentences of idealized theories, they are also connected by formal inference. Thus we have reestablished the isomorphism between thoughts and facts that was lost when the copy theory of representation was abandoned.

Our complete theory thus has a structure made up of three logical levels, the metatheory, the fact level, and the thought level, each of which can be viewed as a set of sentences. The sentences of the metatheory constitute descriptions of those of the fact level, and some of the sentences of the fact level are descriptions of the sentences of the thought level. However, although for clarity in introducing the theory I have spoken of the fact level as composed of sentences, it should properly be thought of as *sets of possible worlds*, as those are described in intensional semantics (Carnap, 1956, Montague, 1974). That is, everything we say in the metatheory should properly be translated into the language of intensional semantics, with sentences being translated as sets of possible worlds, predicates as functions from possible worlds to sets of n-tuples, and so on. Of course the word «fact» should only be applied to those sets of possible worlds that correspond to true sentences, that is, those sets that contain the actual world (the word «proposition» can be used to refer to any set of possible worlds, whether the sentence it corresponds to is true or false). These are what we should regard as the causes of our experience. We will find that none of the predictions of our three-level system depend on identifying any specific fact. They require only that we posit that sets of possible worlds exist and have certain properties. Thus we do not need to assume an idealized ultimate theory, but only that the sentences of our theories can become approximately isomorphic to sets of possible worlds that contain the actual world. This means that we should regard the objects that are spoken of in intensional semantics as real things. The sentences of our physical and biological theories are true to the extent that they correspond to sets of possible worlds that contain the actual world, and in this way our system inherits all their predictions, while making the additional prediction that those theories will be successful under the proper conditions. But because intensional objects are isomorphic with sentences and their parts, I will continue to speak here as though facts were sentences.[5]

---

[5].   What I call the «fact level» might better be called the «proposition level,» since it contains all sets of possible worlds whether or not they include the actual world. But I will continue to use the more evocative term «fact level.»

I want to make a very important point here — that one sentence can imply another only if they are at the same logical level. For example, in a physiological theory of how the brain perceives objects, there will be a logical chain of inference leading from a description of the perceived object to a description of the resulting brain state, which, I contend, can be translated into a description of a set of mental sentences. The point to note is that the description of the object is logically connected, not to the mental sentences themselves, but to a *description* of them. Thus any given axiomatic system is confined to a single logical level. There is a sense in which a higher level can contain a sentence at a lower level, but only in the sense that the higher level contains a description of the lower-level sentence. This is, I believe, the nature of the embedding of one sentence in another that we see in the propositional attitudes. For example, if we state it as a fact that Einstein believed that electrons have a negative charge, we mean that the fact level contains a description of the sentence «electrons have a negative charge» as part of its description of the state of Einstein's brain.

Our three-level system can make the *metaprediction*, or prediction about other predictions, that theories that are strongly supported by evidence will tend to make accurate predictions, because of the isomorphism between theory and fact (compare Cummins' [1989, Chapter 8] concept of an isomorphism between mental states and their «interpretations»). With regard to the truth of scientific theories, which is viewed as a relation of isomorphism between the thought and fact levels, our system is a form of correspondence theory (see Kirkham, 1992). But with regard to the system as a whole I take an «instrumentalist» approach (see, e. g., van Fraassen, 1989), since our reason for accepting it is that it provides the simplest account of a set of observable facts — the success of scientific theories. It is therefore not subject to Putnam's (1981) criticism that correspondence theories require us to take a «God's eye view.» Our isomorphism between thoughts and facts is analogous to the isomorphism between a formal system and its Goedel numbers, used by Kurt Goedel in his proof of the undecidability of arithmetic (Nagel & Newman, 1986), in that our metatheory corresponds to the metalanguage in which Goedel's isomorphism is described, our fact level to the formal system that is the subject of Goedel's proof, and our thought level to the Goedel numbers themselves. It is also related to Tarski's (1983) theory of truth, in that our thought level corresponds to his «object language,» our fact level to his «metalanguage,» and our metatheory to his «metametalanguage.» I think that a number of difficulties with Tarski's theory, that are pointed out by, for example, Soames (1999) and Field (2001), are due to the attempt to define truth in the metalanguage, and are resolved when we define it in the metametalanguage, which corresponds to our metatheory.

Before continuing, let us get a clearer understanding of the contents of the thought level, that is, what it means to have a mental sentence «encoded» in the brain. How can we justify talking about the human brain as if it were a sheet of paper with sentences written on it? To answer this question we must first recognize that a sentence is not an object but a *state* of an object, in the same way that the «on» or «off» position of an ordinary light switch is a state of the switch. The state of a computer's memory, for example, consists of bit patterns that are made up of the «on» or «off» settings of electronic switches that serve as memory elements. Each «on» or «off» setting of a memory element, represented by the digits «1» and «0» respectively, is an arrangement of its electrons that is measured by a quantity called the voltage. These settings, like those of a light switch, are *discrete states* in that the exact voltage does not matter to the functioning of the computer, but only whether the voltage falls in the

«off» or the «on» range.[6] We saw earlier that these bit patterns can be used to encode sentences. But the concept of a sentence is more abstract than that of a bit pattern, in that a sentence can be encoded in a number of ways — as a stream of sounds in a spoken sentence, a stream of printed letters on paper, magnetized areas on a magnetic tape, and so on. I suggest that the brain is a discrete system, and thoughts are its mental sentences, making the brain what can be called a «sentential processor,» that is, a system to which the language of thought (LOT) hypothesis applies (see Aydede, 1997, for a discussion of the issues involved in defining the range of application of the LOT hypothesis). That it is possible to characterize the brain in this way is suggested, for example, by the work of Shastri and his associates (Shastri, 1999), and of Marcus (1998, 2001).

The thought experiment

We will now examine closely the internal structure of our three-level theory, to see how our two formal chains of inference — one at the thought level and one at the fact level — can be constructed, and under what circumstances they will be isomorphic. We will look at the «internal clockwork» of our system to see how it can make metapredictions about the success of scientific theories. This can be done via a «thought experiment,» in which we apply the principles of our theory to a hypothetical situation which is designed to reveal certain consequences of the theory. We will choose our hypothetical situation so that it contains a sentential processor whose internal structure is known to us. For this purpose we will posit a robot named «Rob,» controlled by a computer, which is programmed to carry out formal logical inference on sentences encoded in its memory. In our posit, which belongs to the fact level of our three-level system, we describe both Rob's internal structure and his environment, which is a «blocks world» consisting of solid blocks in various arrangements. We can think of the posit as a sort of «user manual» for Rob the robot. In it we describe Rob as having a visual system consisting of a mechanical «eye» that can focus an image on an electronic «retina,» which sends to his computer a matrix of numbers representing the light intensity at each point on the image. This matrix is processed in Rob's computer by a program similar to Winograd's SHRDLU (see Dennett, 1991, pp. 92-93), containing a component developed by Winston that translates a matrix of intensities into a simple description of the scene that Rob is looking at, thus giving him a simple perceptual system. The output of this perceptual system is a set of bit patterns, each of which is a series of «on» or «off» settings of electronic switches that fits our definition of a «mental sentence.» These bit patterns are composed of shorter bit patterns called «symbols,» which are the basic units that are strung together to form Rob's mental sentences. Like SHRDLU, we will assume that his computer contains a «translation table» that pairs Rob's symbols (i.e., the symbols he uses as predicates and individual constants) with English words and so allows his mental sentences to be translated into English sentences. So, for example, if Rob's environment contains two blocks, with one shaped like a pyramid and the other like a cube, and the pyramid is on top of the cube, and if Rob is in a position to see this arrangement of blocks, then his visual system will produce a bit pattern that can be translated into English as «the pyramid is on the cube.» Since these bit patterns represent formal sentences subject to formal inference, a better translation (and

---

6.   I have defined a «discrete state» by the all-or-nothing character of the response of the system to which the state belongs. But it should properly be defined in terms of the system's *state space*, which in the case of discrete states can be divided into compartments, so that the system's processes move its state from one compartment to the next in such a way that we can predict the next compartment it will enter without knowing the exact location of the state within the compartment it currently occupies.

more convenient for our purposes) would be the formal English sentence «(pyramid on cube).»

Let us use our thought experiment to give a more definite form to the ideas discussed in the previous sections. I said there that the metatheory contains descriptions of the sentences of the fact level, and that some of the sentences of the fact level were descriptions of sentences of the thought level. We want to examine how these descriptions are formulated and how they relate to each other. These relationships are quite complex, involving as they do not only descriptions of sentences, but also descriptions of those descriptions. But I believe that such an examination is the key to understanding the relationship between mind and reality, and the semantics of the propositional attitudes (Quine, 1986). We can formulate these descriptions of sentences as what I will call «positional descriptions,» which are a form of what Tarski (1983, pp. 156-157) calls a «structural descriptive name.» For example, the positional description of the formal sentence «(pyramid on cube)» says essentially «the sentence consists of the words 'pyramid,' 'on,' and 'cube,' in that order.» We can express this as follows:

The posit contains a sentence we will refer to as «S1.»

The symbol in position 1 of S1 is «pyramid.»

The symbol in position 2 of S1 is «on.»

The symbol in position 3 of S1 is «cube.»

We can abbreviate this by adopting the convention of using «(pos S1 <pyramid> 1)» to mean «the symbol in position 1 of S1 is 'pyramid.'» Now we have:

1)   The posit contains S1

and (pos S1 <pyramid> 1) and (pos S1 <on> 2) and (pos S1 <cube> 3)

Here the angle brackets surrounding the words «pyramid,» «on,» and «cube» serve a function similar to quotation marks. The advantage of formulating the description in this way is that it conforms to the syntax of formal logic and lends itself easily to manipulation by the rules of logic. We should view sentence (1) as being contained in the metatheory, part of which consists of sentence-by-sentence descriptions of our posit, which is our user manual. Let us call this description of the sentences of the posit, contained in the metatheory, the «metaposit.»

But it is also the case that the posit contains descriptions of Rob's mental sentences, which are bit patterns in Rob's computer. Let us see how these descriptions of Rob's mental sentences can be formulated as positional descriptions. An example of one of Rob's bit patterns might be «10000000/01100100/11001000.» Suppose that each series of eight bits, as I have indicated by the slashes, forms a piece of information that tends to be processed as a unit. Let us then develop a vocabulary for referring to these units, which we can call «symbols.» We can use the fact that each sequence of ones and zeroes can be interpreted as a number, so that «10000000» is equivalent to the number 128, «01100100» to 100, and «11001000» to 200. Let us form terms for Rob's symbols by placing a «B» in front of the corresponding number, so that we have «B128,» «B100,» and «B200» as terms for our three symbols. Then we can write the longer bit pattern composed of these symbols as «B128, B100, and B200, in that order,» or in our positional description format, as follows:

2)   Rob contains BP1

   and (pos BP1 B128 1) and (pos BP1 B100 2) and (pos BP1 B200 3)

Here we use «BP1» to refer to the entire bit pattern. (We have used the same predicate «pos» as in the metaposit, but we could have used a different term without affecting any of our conclusions.) Let us now take a further step to simplify our description of Rob's bit patterns, using our translation table. Suppose the translation table associates the symbol «B128» with the word «pyramid,» «B100» with «on,» and B200 with «cube.» Then we can develop an alternative term for each of our symbols by enclosing the corresponding English word in angle brackets. For example, «<pyramid>» will refer to the same bit pattern as «B128.» Our positional description can then be expressed as follows:

3)   Rob contains BP1

   and (pos BP1 <pyramid> 1) and (pos BP1 <on> 2) and (pos BP1 <cube> 3)

Comparing sentence (3) with sentence (1), we see that our positional descriptions of Rob's bit patterns in the posit resemble our descriptions of sentences of the posit in the metaposit. This depended critically on our use of the translation table between Rob's mental symbols and English words. By using the translation table in this way we have temporarily bypassed an important problem, called by Quine (1960) the «indeterminacy of translation.» I will return to this problem later.

But Rob's mental sentences are bit patterns, and the sequence of the symbols they contain must depend on the set of electrical connections between the switches in Rob's computer that tells the system which symbol comes next in the sequence. These connections define the meaning of the word «position» and of the predicate «pos,» as these terms apply to the order of the symbols in Rob's mental sentences. On the other hand, the same predicate «pos,» when applied in the metaposit to sentences of the posit, refers to the left-to-right order of words on the printed pages of our user manual. But in a sense the predicate «pos» in our positional descriptions has the same meaning in both cases, because both the sentences of the posit and Rob's mental sentences are related to each other by formal inference. We can take advantage of this fact by a process which I will call «schematization,» which translates a positional description into a sort of «picture» of the sentence described. Let us see how our positional description in sentence (3) would be schematized. The rule we will follow has the following form:

   If x, s, y, z, and w are variables, and the positional description has the form «(x contains s) and (pos s <y> 1) and (pos s <z> 2) and (pos s <w> 3),» then schematize it as «(x contains <y z w>)»

We simply take our terms for the three words in the described sentence, removed the angle brackets surrounding them, arranged them in the sequence prescribed by the positional description, and enclosed the entire sequence in angle brackets. Thus the description in (3) would be schematized in the posit as follows:

   (Rob contains <pyramid on cube>)

Notice that in the schema inside the angle brackets, we are using the order of words in the user manual as a sort of *picture* of the order of symbols in Rob's computer.

The isomorphism

Now let us examine in some detail how the isomorphism comes about between the sentences of the posit and Rob's mental sentences. We saw that the isomorphism is due to the fact that the sentences of both fact and thought levels are related by logical inference. In our thought experiment, this means that the rules of inference apply both to the sentences of the posit and to Rob's mental sentences. We can give explicit form to these rules of inference as follows:

Match rule:

For-all σ ρ

IF for-all n, α, β

      (IF (n <= max)

      THEN (IF (pos σ α n) and (pos ρ β n)

      THEN (α = β) or ((variable α) and (name β))))

THEN (match σ ρ)

Assignment rule:

For-all σ ρ α β

IF there-exists n

((match σ ρ) and (pos σ α n) and (variable α) and (pos ρ β n) and (name β))

THEN (assign σ ρ α β)

Substitution rule:

For-all φ ι α β

(subst φ ι α β)

IF AND ONLY IF

      For-all n

      IF (pos φ α n) and (pos ι β n)

      THEN (α = β) or (assign φ ι α β)

Complete rule of inference:

For-all ζ ι π ρ σ α β

IF (major-premise ζ π) and (minor-premise ζ ρ) and (antecedent π σ)

and (consequent π φ) and (assign σ ρ α β) and (subst φ ι α β)

THEN (logically-implies ζ ι)

Example:

If we apply the support rule: For-all x if (x on cube) then (cube supports x)

To the minor premise: (pyramid on cube).

To get the conclusion: (cube supports pyramid)

Then we will have the following variable assignments:

ζ = «(For-all x if (x on cube) then (cube supports x)) and (pyramid on cube)»

π = major premise = «(For-all x if (x on cube) then (cube supports x))»

ρ = minor premise = «(pyramid on cube)»

σ = antecedent of major premise = «(x on cube)»

ϕ = consequent of major premise = «(cube supports x)»

ι = conclusion = «(cube supports pyramid)»

These rules state in a formal way the procedure we applied in the «Socrates is mortal» example in the introduction. This particular statement of the rules of inference may not be free of errors, and is certainly not complete, for example, it does not deal with compound antecedents. Its purpose is only to provide an explicit statement of these rules, as a basis for discussion. Notice that these rules of inference have the special property of being rules for applying other rules. They operate on positional descriptions of the major and minor premises of a logical inference, and produce a positional description of the conclusion. To see how this works, consider a simplified example, using a principle which I will call the «support rule,» that might apply to Rob's world. The support rule states the trivial fact that if anything is on top of the cube then the cube supports it. It will serve as our major premise, and can be formulated as follows:

4)   (For-all x if (x on cube) then (cube supports x))

We will apply our rules of inference to the support rule (remember that the rules of inference are rules for applying rules), in combination with the minor premise «(pyramid on cube)» to obtain the conclusion «(cube supports pyramid).»

The first of our rules of inference, called the «match rule,» concludes that the antecedent of the support rule «(x on cube)» matches the minor premise «(pyramid on cube)» because at every position, either they contain the same symbol or the first has a variable where the second has a name. The other rules of inference carry out the remainder of the inference process, assigning a value to each variable and substituting those values in the consequent of the major premise to get the conclusion. Notice particularly that the following part of the match rule:

(pos σ α n) and (pos ρ β n)

is the antecedent of a higher-order major premise that applies, in our example, to the higher-order minor premise «(pos S1 <x> 1) and (pos S2 <pyramid> 1),» where «(pos S1 <x> 1)» is part of the positional description of the antecedent of the support rule «(x on cube),» and »(pos S2 <pyramid> 1)» is part of the positional description of the minor premise «(pyramid on cube).» In the course of applying this higher-order major premise to the higher-order minor premise, we make a set of variable assignments, which include the following:

α = <x>, β = <pyramid>

Then we substitute these values into the consequent «(α = β) or ((variable α) and (name β)),» which becomes, «(<x> = <pyramid>) or ((variable <x>) and (name <pyramid>)).» This statement is true, since <x> is a variable and <pyramid> is a name. The important fact to notice is that in applying our rules of inference we are carrying out at a higher logical level the same procedure as in the «Socrates is mortal» example, and that is the very procedure that is *described* by these rules. It is clear that if we want to describe in formal language the procedure we use in applying our rules of inference, that description must again take the form of these same rules of inference.

Now we can see more clearly the reason for the isomorphism between the fact and thought levels, represented in our thought experiment by the isomorphism between sentences of the posit and Rob's mental sentences. This isomorphism depends on the fact that we find a copy of the rules of inference in the metaposit, which we can call the «higher-order rules of inference,» and another copy in the posit, which we can call the «lower-order rules of inference.» This gives concrete form to the idea that both facts and thoughts are related by logical inference. The higher-order rules of inference that we find in the metaposit describe the fact that the sentences of the posit must be logically consistent, and that any sentence that follows logically from sentences of the posit can be treated just as though it were itself written in our posit/user manual. The lower-order rules of inference that we find in the posit describe Rob's capacity for applying formal inference to his mental sentences. When bit patterns encoding the major and minor premise of a syllogism are presented to Rob's inference mechanism, his internal electronic connections are such that the bit pattern that encodes the conclusion of the syllogism is produced. Thus the lower-order rules of inference summarize certain facts about Rob's electronic connections and the laws of physics that apply to them.

To see the isomorphism more clearly, imagine that the support rule is a fact stated in the user manual and is also known to Rob, so that it is encoded as a mental sentence in his brain. We can list the sentences of the metaposit that would constitute a description of this situation as follows:

One-level inference:

Major Premise:

(The posit contains R1) and (pos R1 <**for-all**> 1) and (pos R1 <**x**> 2) and (pos R1 <**if**> 3)

and (pos R1 <**C1**> 4) and (pos R1 <**then**> 5) and (pos R1 <**C2**> 6)

and (pos C1 <**x**> 1) and (pos C1 <**on**> 2) and (pos C1 <**cube**> 3)

and (pos C2 <**cube**> 1) and (pos C2 <**supports**> 2) and (pos C2 <**x**> 3)

Minor Premise:

(The posit contains S1) and (pos S1 <**pyramid**> 1) and (pos S1 <**on**> 2) and (pos S1 <**cube**> 3)

Conclusion:

(The posit contains S2) and (pos S2 <**cube**> 1) and (pos S2 <**supports**> 2) and (pos S2 <**pyramid**> 3)

Two-level inference (abbreviating «(the posit contains S3)» as «(PC S3)»):

Major Premise:

(PC S23) and (pos S23 <Rob> 1) and (pos S23 <contains> 2) and (pos S23 **<RP1>** 3)

and (PC S24) and (pos S24 <pos> 1) and (pos S24 <RP1> 2) and (pos S24 **<<for-all>>** 3) and (pos S24 <1> 4)

and (PC S4) and (pos S4 <pos> 1) and (pos S4 <RP1> 2) and (pos S4 **<<x>>** 3) and (pos S4 <2> 4)

and (PC S5) and (pos S5 <pos> 1) and (pos S5 <RP1> 2) and (pos S5 **<<if>>** 3) and (pos S5 <3> 4)

and (PC S6) and (pos S6 <pos> 1) and (pos S6 <RP1> 2) and (pos S6 **<<CP1>>** 3) and (pos S6 <4> 4)

and (PC S7) and (pos S7 <pos> 1) and (pos S7 <RP1> 2) and (pos S7 **<<then>>** 3) and (pos S7 <5> 4)

and (PC S8) and (pos S8 <pos> 1) and (pos S8 <RP1> 2) and (pos S8 **<<CP2 >>** 3) and (pos S8 <6> 4)

and (PC S9) and (pos S9 <pos> 1) and (pos S9 <CP1> 2) and (pos S9 **<<x>>** 3) and (pos S9 <1> 4)

and (PC S10) and (pos S10 <pos> 1) and (pos S10 <CP1> 2) and (pos S10 **<<on>>** 3) and (pos S10 <2> 4)

and (PC S11) and (pos S11 <pos> 1) and (pos S11 <CP1> 2) and (pos S11 **<<cube>>** 3) and (pos S11 <3> 4)

and (PC S12) and (pos S12 <pos> 1) and (pos S12 <CP2> 2) and (pos S12 **<<cube>>** 3) and (pos S12 <1> 4)

and (PC S13) and (pos S13 <pos> 1) and (pos S13 <CP2> 2) and (pos S13 **<<supports>>** 3) and (pos S13 <2>4)

and (PC S14) and (pos S14 <pos> 1) and (pos S14 <CP2> 2) and (pos S14 **<<x>>** 3) and (pos S14 <3> 4)

Minor Premise:

(PC S15) and (pos S15 <Rob> 1) and (pos S15 <contains> 2) and (pos S15 **<BP1>** 3)

and (PC S16) and (pos S16 <pos> 1) and (pos S16 <BP1> 2) and (pos S16 **<<pyramid>>** 3) and (pos S16 <1>4)

and (PC S17) and (pos S17 <pos> 1) and (pos S17 <BP1> 2) and (pos S17 **<<on>>** 3) and (pos S17 <2> 4)

and (PC S18) and (pos S18 <pos> 1) and (pos S18 <BP1> 2) and (pos S18 **<<cube>>** 3) and (pos S18 <3> 4)

Conclusion:

(PC S19) and (pos S19 <Rob> 1) and (pos S19 <contains> 2) and (pos S19 **<BP2>** 3)

and (PC S20) and (pos S20 <pos> 1) and (pos S20 <BP2> 2) and (pos S20 **<<cube>>** 3) and (pos S20 <1> 4)

and (PC S21) and (pos S21 <pos> 1) and (pos S21 <BP2> 2) and (pos S21 <**<supports>>** 3) and (pos S21 <2>4)

and (PC S22) and (pos S22 <pos> 1) and (pos S22 <BP2> 2) and (pos S22 **<<pyramid>>** 3) and (pos S22 <3>4)

Under the heading «one-level inference» we see the major and minor premise and conclusion of the support rule in the posit, as they would appear in the form of positional descriptions in the metaposit. Under the heading «two-level inference,» we see the description in the metaposit of the description in the posit of the mental sentences in Rob's computer that encode the two premises and the conclusion. (Here we have an example of a description of a description.) For simplicity, let us schematize these sentences. (The words in bold letters are those that will appear in the schematizations.) In the case of the sentences labeled «two-level inference,» we will use «double schematization.» We can illustrate how double schematization works as follows:

(pos S15 <Rob> 1) and (pos S15 <contains> 2) and (pos S15 <BP1> 3)

      schematized as: S15 = <Rob contains BP1>

(pos S16 <pos> 1) and (pos S16 <BP1> 2) and (pos S16 <<pyramid>> 3) and (pos S16 <1> 4)

      schematized as: S16 = <pos BP1 <pyramid> 1>

(pos S17 <pos> 1) and (pos S17 <BP1> 2) and (pos S17 <<on>> 3) and (pos S17 <2> 4)

>   schematized as: S17 = <pos BP1 <on> 2>

(pos S18 <pos> 1) and (pos S18 <BP1> 2) and (pos S18 <<cube>> 3) and (pos S18 <3> 4)

>   schematized as: S18 = <pos BP1 <cube> 3>

>   Those three schemas can in turn be schematized as: BP1 = <<pyramid on cube>>

As a result we obtain two parallel chains of inference, the first of which is derived from the sentences labeled «one-level inference» as follows:

5)   The posit contains <For-all x if (x on cube) then (cube supports x)>

6)   The posit contains <pyramid on cube>

7)   The posit contains <cube supports pyramid>

and its parallel, from the sentences labeled «two-level inference» as follows:

8)   The posit contains <Rob contains <For-all x if (x on cube) then (cube supports x)>>

9)   The posit contains <Rob contains <pyramid on cube>>

10)   The posit contains <Rob contains <cube supports pyramid>>

We see that there is an exact parallel between our chain of inference in the posit (as described in the metaposit) and our chain of inference linking Rob's mental sentences as they are described in the posit, those descriptions being in turn described in the metaposit.

Notice that schemas can provide us with an enormous simplification in the process of deriving logical inferences from positional descriptions. This is particularly striking in the case of sentences (8), (9), and (10), as compared to the sentences above labeled «two-level inference.» We see that one way to get from the premises to the conclusion is by means of an extremely complicated process of applying the higher-order rules of inference to a positional description of the lower-order ones, and by that means applying the lower-order rules to positional descriptions of mental sentences. But we can get the same result by applying our own capacity for logical inference directly to the schematized sentences inside the angle brackets. Thus schemas can serve as «contexts» as they are used in artificial intelligence (see Sowa, 2003; compare the concept of «situation» as used, for example, by Reconati, 2000). I contend that schemas are the normal way in which our brains represent positional descriptions, and that the propositional attitudes are expressions of these schemas, so that, for example, the sentence «(Rob contains <pyramid on cube>)» has essentially the same meaning as the sentence «Rob thinks that the pyramid is on the cube.» Notice that sentences (8), (9), and (10) are schematizations of positional descriptions, which contain symbols in double angle brackets, such as «<<pyramid>>.» The symbol «<<pyramid>>» is a name in the metaposit of a name in the posit for one of the symbols in Rob's computer. This symbol is in turn Rob's name for an object, the pyramid. Here we see the relationship between terms that occur inside the content of propositional attitudes and those occurring outside it (Quine, 1986, pp. 32-34).

Our higher-order inference rules provide the means of giving a definition in the metaposit of the relation of logical implication between sentences of the posit, which we can

call «logically-implies$_1$,» and the combination of the higher- and lower-order inference rules provides the means of defining the corresponding relation between bit patterns in Rob's computer, which we can call «logically-implies$_2$,» using the subscript to distinguish these two uses of the word. We can then state the isomorphism in the metaposit by saying that whenever one of Rob's bit patterns logically-implies$_2$ another, the sentence of the posit corresponding to the first logically-implies$_1$ that corresponding to the second.

We have not yet provided in our model for the fact that we are able to think in terms of these three levels. To represent that ability, we need to break the thought level into three sublevels, corresponding to the three levels of the model. Suppose, for example, that the metaposit contains the sentence, «It is a fact that Rob thinks that it is a fact that Sally thinks that the book is on the table,» which in our more formal terminology is, «The posit contains <Rob contains <the facts contain <Sally contains <book on table>>>>.» Then the sentence, «<<<<book on table>>>>,» will be placed in the lowest of these three sublevels of the thought level. Of course, the metaposit can also contain the sentence, «It is a fact that Rob thinks that it is a fact that *Rob* thinks that the book is on the table,» allowing him to attribute beliefs to himself, and thus to reason about the possibility of his being wrong, and about how he might go about correcting errors in his thinking. This makes the system capable of «metareasoning» (Russell and Wefald, 1991; Costantini, 2002). These multi-level mental sentences never need to take the form of positional descriptions, because they will always be in schematized form. The information provided by positional descriptions can always be obtained from schemas as needed. In the terms used by Cosmides and Tooby (2000), the sublevels of the thought level represent different levels of «decoupling.» Each of these five levels, the metatheory, fact level, and three sublevels of the thought level, can be divided into one or more sections, one section containing sentences about things, another descriptions of such sentences, another descriptions of descriptions, and so on. The relation of *reference* can then be represented as a relation between symbols at the top sublevel of the thought level and symbols in the corresponding section of the fact level.

Some objections to the theory

I have concentrated in this paper on explaining the three-level theory, rather than on defending it. I want here to respond briefly to some criticisms of the sort of theory presented in this paper. The first criticism is based on the concept of the indeterminacy of radical translation (Quine, 1960; Dennett, 1987, pp. 40-42), which says essentially that it is meaningless to talk of a correspondence between mental sentences and facts, because any brain state could always be translated into sentences in more than one way. But suppose that all meaningful logical systems share a subset of their axioms, and this subset contains all the undefined terms of the system. Suppose further that this subset of axioms has a high density, in Hayes' (1985a) sense that the ratio of statements to concepts is high. Then this subset could act as a sort of «Rosetta stone» for establishing the translation table between the languages in which different systems were expressed.[7] Hayes (1985a) calls such a subset of axioms a «kernel theory.» Let us call the concepts contained in it the «conceptual base.» It seems to me that recent research in artificial intelligence on formal theories of common-sense reasoning shows that such a conceptual base may exist, consisting of basic geometrical and physical

---

[7]. In our three-level theory, Cummins' (1989) «specification problem» reduces to the problem of finding a translation table between the mental sentences of any given sentential processor on the one hand, and the sentences of the fact level on the other.

axioms (e. g., Hayes, 1985b; Varzi, 1997; Davis, 1998; see also Smith and Casati, 1994, for a different perspective on this research).[8] The indeterminacy argument assumes without good reason that such a conceptual base does *not* exist. In response, we could reverse the argument and say that the fact that widely different cultures can communicate, and the demonstrated success of modern science, indicate respectively that mental sentences in different brains, and mental sentences and facts, do in fact share a conceptual base. As an example of how such principles might operate in everyday problem-solving, I was once faced with the problem of having to cut a bagel for my son's lunch using only a plastic fork. I solved the problem by repeatedly pushing the tines of the fork into the bagel at different spots and moving it back and forth to cut away the substance, although as far as I remember I had never before used a fork to cut something in that way. Using as a model the sort of commonsense reasoning discussed in Varzi (1997) and Davis (1998), we can explain my behavior by saying that I used my intuitive knowledge of geometry and physics to understand that I needed to cut away the soft material between the two halves of the bagel in order to separate them, and I knew that moving the fork in that way would necessarily accomplish that.

The language of thought hypothesis has also been criticized on the grounds that it would be too slow to account for the rapidity with which we are able to respond intelligently to circumstances (e. g., Dennett, 1998, pp. 86-87; Churchland, 1986, pp. 459-460). This criticism is contradicted by the work of Shastri and his associates (Shastri, 1999) who have developed a model that demonstrates how a sentential processor can be implemented in a brain-like system that can account for the speed of human problem solving. Another objection (e. g., Dennett, 1998, pp. 279-281), that sentential processing cannot explain human emotion, is contradicted, for example, by the work of Ortony, Clore, and Collins (1988), and Anderson and Lebiere (1998), who suggest how such an explanation could be developed. In connection with the treatment of emotion in our system, I am of course not saying that logical inference is the only kind of causal relationship that can exist between thoughts, but only that it is the logical connection between thoughts that makes it possible for them to be isomorphic with facts. Since thoughts are states of the brain, they can have properties, such as activity levels, degrees of belief, and so on, that do not affect their logical implications. Because we often do not know what premises are correct, but instead have to make the best of a mass of contradictory information, the control of the thought process that is exercised by emotions is necessary for our survival.

The apparent problem of *externalism* raised by Putnam's (1981) «Twin Earth» thought experiment can be answered if we recognize that the entry in our translation table for the word «I» is *conditional*, in that a token of this word refers to whatever system contains the sentence that contains that token, so that identical mental sentences in different brains can have different referents, just as identical thoughts in different minds can have different meanings. Fodor's (2001) objection to his own CTM theory that it cannot be a general theory of

---

[8].   I do not mean to say that these authors draw the same conclusions I do. In fact Hayes (1985a, p. 17) expresses doubt that there is such a kernel theory for common-sense. But I think that Hayes' doubts are based on the fact that most common-sense concepts cannot be *defined in terms of* those of the conceptual base. My suggestion is, rather, that our qualitative concepts are efficient approximations to the correct quantitative theory that includes the kernel theory; that we use metareasoning to create and modify these qualitative approximations as needed; and that we can fall back on the underlying quantitative theory when necessary, for example when a husband and wife are confronted in a store with an unlabeled piece of furniture intermediate between a chair and a loveseat, and one says to the other, «Let's sit down on it and if we both fit, it's a loveseat!» Thus qualitative and quantitative concepts are not related as a building to its foundation, but as a ship at anchor to the sea-floor where the anchor rests.

cognition fails to take into account the fact that, while each individual computation is local, each successive computation applies to a different overlapping domain, so that any global interaction can be computed in a finite number of steps.

Conclusion

I have argued in this paper that, in order to provide an adequate explanation of the success of scientific theories, we need to add a level of theory, which I call a metatheory, above that of our ordinary physical and biological theories. The metatheory makes it possible for us to describe the isomorphism between thoughts and facts, as well as the isomorphism between theories and facts, since theories are expressions of thought. This explains how theories can serve as maps to the facts, allowing us to make correct predictions. Consider what it would mean in our thought experiment for one of Rob's predictions to come true. The truth in Rob's world is whatever the posit says it is, much as the truth in a novel is whatever the author says it is. The purpose of the posit is to tell us what would happen if there really were a robot like Rob who lived in a blocks world environment like the one described. Under what circumstances would such a robot be able to make correct predictions about its environment? Suppose that one sentence of the posit says, «At 9:50 A. M. Rob contained the sentence <the cube will fall off the block at 10:00>,» and another says, «the cube fell off the block at 10:00.» Then the posit is telling us that Rob's prediction was confirmed, in other words, what he thought would happen *did* happen. So in general, we predict that a robot like Rob will make correct predictions if its mental sentences correspond with the relevant sentences of the fact level (which includes our posit). This amounts to a *metaprediction* (a prediction about other predictions) about the conditions under which predictions will be correct.[9] But this metaprediction could not have been stated or proved in the posit, because it requires us to be able to refer to the sentences of the posit, and this we can only do in the metaposit. Since our posit consists of the physical theory of Rob's environment and his internal electronics, our metaprediction of the success of Rob's predictions is a generalization that is true of our physical theory of Rob and his environment, but cannot be stated or proved within that physical theory. This suggests that our ordinary theories of nature have the property that mathematicians call «omega incompleteness,» which means that there are generalizations, each instance of which is a theorem, although the generalization itself is not a theorem. We see that including the additional level of the metaposit allows us to account for the success of predictions in a way that would otherwise not be possible. Notice that this isomorphism is implicit in our accepted scientific theories, since those theories predict that computers and other similar mechanisms will be capable of formal logical inference. (Consider, for example, a physical theory that includes a description of an exploding star as well as a description of a computer programmed to simulate the exploding star. We would see a parallel between the theory's predictions about the internal states of the computer and its predictions about the star, but the theory itself would not contain a general statement of this parallel.)

So far, I have shown only how it is possible to make accurate predictions if we know the correct premises, but not how we come to know what those premises are. How do we know, for example, that all men probably *are* mortal? I want to finish by suggesting very

---

[9]. In order for an organism to make use of such predictions, it must have the capacity for *perception* and *action*. Perception, when successful, creates mental sentences that correspond to observable facts, and action, when successful, does the opposite, starting from mental sentences of the kind we call «intentions,» and causing motions of the body that in turn bring about states of affairs that correspond to those intentions.

briefly what direction such an explanation might take. In science as well as in everyday life, we assign probabilities to hypotheses based on evidence. For these hypotheses to make reliable predictions, a combination of *randomness* and *regularity* is required: randomness to prevent the evidence from routinely containing spurious patterns, and regularity so that there *are* real patterns to discover. Under what circumstances can we expect our world to show the necessary combination of regularity and randomness? To see how this is possible, let us add a quantitative aspect to our metatheory, by assigning *probabilities* to possible worlds, forming what is called an «initial distribution.» This makes our metatheory into a version of the Bayesian approach to the theory of probability (see, for example, Earman, 1992; Howson, 2000; Franklin, 2001).

To see how this extension of our metatheory can help us, consider a simplified example, in which each of our possible worlds consists of a series of one thousand tosses of a fair coin, and we choose our initial distribution so that half of these worlds have heads on the first toss and half tails; half of the resulting two groups have heads and half tails on the second toss; half of the resulting four groups have heads and half tails on the third toss; and so on. Then the vast majority of these worlds will show the familiar statistical pattern of a series of coin tosses, with an apparently random sequence of heads and tails, about one half being heads. Notice that there is also regularity here, in that the coin always falls either heads or tails. The results depend on the *symmetries* of the coin and of the law of gravity — that the coin is physically the same on both sides; that it is shaped so that it always falls on one side or the other; and that the same coin is used on every toss. So an initial distribution constructed according to simple rules has conferred on the vast majority of our possible worlds the properties of regularity and randomness that we were looking for.

I want to suggest very briefly, as a program for future investigation, that we look for a set of rules that produce in a similar manner an initial distribution that can account for the pattern of events in the actual world. These rules would, perhaps, resemble what is called «symmetry breaking» in modern physics (Pagels, 1985), where the randomness is introduced by broken symmetries and the regularity by those that remain unbroken. Such symmetries are expressed in terms of geometric distributions of matter or fields, so that their use in generating the initial distribution requires that all the possible worlds in our distribution share the principles of geometry and physics that make up our conceptual base. An initial distribution of this kind would confer on the vast majority of worlds the property that evidence obtained by the inhabitants of any given world would tend to be representative of the actual patterns that hold in that world. If we make the assumption that what holds in the vast majority of possible worlds holds also in the actual world, this gives us an explanation for the past success of science in our own world.

But can we predict that science will be successful in the future? Let us take our initial distribution as the basis for a definition of the concept of *probability*. To see the significance of having a definition of probability, consider that we call a factual statement «knowledge» when its probability is high enough that it can provisionally be treated as though it were certainly true (see Carnap, 1962). Thus if we know how to define probability we know how to justify our factual statements. In this way we introduce into our three-level theory a medicinal dose of *metaphysics*, the need for which is, I think, the main lesson to be drawn from Goodman's (1983) «new riddle of induction.» Because of their special role in the rules for constructing our initial distribution, we have singled out *a priori* the predicates contained in the conceptual base as, in Goodman's terms, «projectible.» (Projectibility is, after all, a

symmetry between examined and unexamined cases.) We have also phrased our definition of probability so that it expresses *a priori* conditions under which the inhabitants of a world can come to an approximate knowledge of that world through the systematic collection and careful analysis of experience. We may say that our theory of *truth* is non-epistemic, whereas our theory of *probability* is epistemic (Kirkham, 1992). The isomorphism discussed in the previous sections of this paper accounted for our ability to apply knowledge to the making of accurate predictions. The addition of an initial distribution to our metatheory accounts for our ability to obtain that knowledge, in the form of statements of probability, by means of the scientific method.

References

Anderson, J. R. and C. Lebiere (1998). *The atomic components of thought*. Mahwah, N J: Lawrence Erlbaum Assoc.

Aydede, M. (1997). Language of thought: The connectionist contribution. *Minds and Machines*, 7:57-101.

Costantini, S. (2002). Meta-reasoning: a survey. In *Computational logic: From logic programming into the future — special volume in honour of Bob Kowalski.* Springer-Verlag. Invited paper.

Carnap, R. (1956). *Meaning and necessity: A study in semantics and modal logic*. Chicago: The University of Chicago Press.

Carnap, R. (1962). *The logical foundations of probability 2nd Edition*: Chicago: The University of Chicago Press.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind/brain*. Cambridge, MA: MIT Press/A Bradford Book.

Cosmides, L. and Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentations. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective*. Oxford: Oxford University Press.

Cummins, R. (1989). *Meaning and Mental Representation*. Cambridge, MA: The MIT Press/ A Bradford Book.

Davis, E. (1998). The naïve physics perplex. *Artificial Intelligence*. Vol. 19, pp. 51-79.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: The MIT Press/ A Bradford Book.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown & Co.

Dennett, D. C. (1995). *Darwin's dangerous idea*: *Evolution and the meanings of life*. New York: Simon & Schuster / A Touchstone Book.

Dennett, D. C. (1998). *Brainchildren: Essays on designing minds*. Cambridge, MA: The MIT Press / A Bradford Book.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: The MIT Press / A Bradford Book.

Field, H. (2001). Tarski's theory of truth. In *The nature of truth: Classic and contemporary perspectives.* M. P. Lynch (Ed.). Cambridge, MA: The MIT Press / A Bradford Book.

Fodor, J. A. (1975). *The language of thought*. New York: Thomas Y. Crowell.

Fraassen, B. C. van (1989). *Laws and symmetry*. Oxford: Clarendon Press.

Franklin, J. (2001) Resurrecting logical probability. *Erkenntnis,* 55. pp. 273-305.

Goodman, N. (1983). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Hayes, P. J. (1985a) The second naive physics manifesto. In Hobbs, J. R. & Moore, R. C. (Eds.). *Formal theories of the commonsense world*. Ablex.

Hayes, P. J. (1985b) Naïve physics I: Ontology for liquids. In Hobbs, J. R. & Moore, R. C. (Eds.). *Formal theories of the commonsense world*. Ablex.

Howson, C. (2000). *Hume's Problem: Induction and the justification of belief*. Oxford: Clarendon Press.

Kirkham, R. L. (1992). *Theories of truth: A critical introduction*. Cambridge, MA: MIT Press / A Bradford Book.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*. Vol. 37. No. 3, pp 243-282.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: The MIT Press / A Bradford Book.

McGinn, C. (1995). Logic, mind, and mathematics. In B. Dahlbom (Ed.) *Dennett and his critics: Demystifying mind.* Cambridge, MA: Blackwell Publishers, Ltd.

Milliken, R. G. (1995). On mentalese orthography. In B. Dahlbom (Ed.). *Dennett and his critics: Demystifying mind*. Cambridge: Blackwell Publishers, Ltd.

Montague, R. (1974). Pragmatics and intentional logic. In R. H. Thomason (Ed.) *Formal philosophy: Selected papers of Richard Montague.* New Haven and London: Yale University Press.

Nagel, E. & Newman, J. K. (1986). *Goedel's Proof.* New York: New York University Press.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge: Cambridge University Press.

Peirce, C. S. (1878/2001). How to make our ideas clear. In *The nature of truth: Classic and contemporary perspectives.* M. P. Lynch (Ed.). Cambridge, MA: The MIT Press / A Bradford Book.

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York: HarperCollins.

Pagels, H. (1994). *Perfect Symmetry.* New York: HarperCollins.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. New York: Routledge.

Putnam, H. (1981). *Reason, truth, and history*. Cambridge, England: Cambridge University Press.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Quine, W. V. O. (1986). *Philosophy of Logic*. Cambridge, MA: Harvard University Press.

Rapaport, W. J. (1995). Understanding understanding. In J. F. Tomberlin (Ed). *A. I., connectionism, and philosophical psychology. Philosophical perspectives*, Vol. 9, Atascadero, CA: Ridgeview. Pp 49-88.

Recanati, F. (2000). The iconicity of metarepresentations. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective*. Oxford: Oxford University Press.

Russell, S. & Wefald, E. (1991). Principles of metareasoning. *Artificial Intelligence*. Vol. 49, pp. 361-395.

Sankey, H. (2001). Scientific realism: An elaboration and a defense. *Theoria*, Vol. 98, pp. 35-54.

Shastri, L. (1999). Advances in SHRUTI: A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. *Applied Intelligence*. Vol. 11, No. 1, pp. 79-108.

Smith, B and Casati, R. (1994). Naïve physics: An essay in ontology. *Philosophical psychology*. Vol. 7, No. 2, pp 225-294.

Soames, S. (1999). *Understanding truth*. Oxford: Oxford Press.

Sowa, J. F. (2003). Laws, facts, and contexts: Foundations for multimodal reasoning. In V. F. Hendricks, K. F. Jorgensen, & S. A. Pedersen (Eds.) *Knowledge contributors*. Synthese Library, Vol. 322. Dordrecht: Kluwer Academic Publishers.

Stanford, P. K. (2003). Pyrrhic victories for scientific realism. *The journal of philosophy*, Vol. C, No. 11.

Tarski, A. (1983). *Logic, semantics, metamathematics: Papers from 1923 to 1938* (2nd Ed.). U. S. A.: Hackett Publishing Co.

Varzi, A. C. (1997). Boundaries, continuity, and contact. *Nous*, 31, pp. 26-58.

_____

Paul Bowyer, PhD.
<paul.bowyer@truevalue.com>

A DILEMMA FOR THE WEAK DEFLATIONIST ABOUT TRUTH

Glen Hoffmann

——————————————————————

1 Introduction

The deflationist about truth is committed to a triviality or transparency thesis: the content of the truth predicate is exhausted by its involvement in some version of the truth-schema [P] is true iff P (where '[P]' stands for any declarative propositional object and 'P' stands for '[P]'s object-level equivalent). Within this classification, deflationists can usefully be divided into two camps according to the extent of their ontological commitment: a strong camp and a weak camp. Strong deflationism holds that the truth predicate doesn't designate a property or that there is no property of truth (e.g., Ramsey (1927), Ayer (1946), and Grover, Camp and Belnap (1975)). Weak deflationism holds that the truth predicate *does* designate a property (e.g., Horwich (1998a, 1999), Sosa (1993) and Soames (1997, 1999)). It is just that the property of truth, on weak deflationism, is deflated in some sense, e.g., it is not *substantial*, *theoretically important*, *interesting*, *explanatory*, *a natural kind*, or anything in this general vicinity. For example, Horwich, in making the latter kind of claim, denies that truth is «an ordinary sort of property — a characteristic whose underlying nature will account for its relations to other ingredients of reality» (1998a, p. 2).

As things currently stand, the consensus is that weak deflationism is the superior deflationary alternative. The problem with strong deflationism is that its ontological thesis — that truth is not a property — seems to be underwritten by a semantic claim that is widely regarded as indefensible: 'is true' does not function semantically as a predicate (i.e., it does not denote a property). Among others, Horwich (1998a, p.3) and Gupta (1993, pp. 366-7) have persuasively argued the speciousness of this thesis on the grounds that '[P] is true' and 'P' are not synonymous,[1] implying that 'is true' is ineliminable, and correspondingly, that it must designate a property of some sort.[2]

—————————————

[1]. Gupta appeals primarily to the indispensable generalizing function of the truth predicate to undermine the synonymy claim. Horwich appeals to the denominalizing function of the truth predicate in general to undermine the synonymy claim.

[2]. See Horwich (1998a) for the *general* line of argument thought to undermine strong deflationism's ontological thesis: «… 'is true' is a perfectly good English predicate — and (leaving aside nominalistic concerns about the very notion of 'property') one might well take this to be a conclusive criterion of standing for a property of *some* sort» (p. 37); «No doubt truth is very different from most properties insofar as it has no underlying nature; but in light of the inferential role of 'true' as a logical predicate, it is nonetheless a 'property', at least in some sense of the term» (p. 125).

In opposition to the prevailing view, I argue here that weak deflationism's initial promise is illusory, that it falls short of being a viable alternative to strong deflationism and to inflationary theories of truth (theories according to which truth is a substantial property of truth bearers). Weak deflationism, it turns out, is on shaky ground since it is vulnerable to an inexorable instability objection the general form of which has been highlighted by Boghossian (1990) and Wright (1992, 1996). Contra Boghossian and Wright, though, it turns out that the strongest variation of the instability objection gives rise to a dilemma concerning *truth-property ascriptions* rather than the *concept of* truth.

2 The Dilemma

Weak deflationism (hereafter 'WD'), unlike strong deflationism, is a theory of truth that takes the ontological middle road. It is committed to a two-pronged ontological thesis: truth is both i) a property and ii) deflated. It has been speculated that WD's ontological thesis generates an incoherence or instability in its view of truth of the following basic form: either truth is a property in which case 'is true' is susceptible to explicit analysis or truth is not a property in which case 'is true' is not susceptible to explicit analysis.[3] For example, Boghossian (1990) advances an instability objection to WD along such lines on the basis of the conceptual import of the T-schema, while Wright (1992, 1996) advances an instability objection to WD along such lines on the basis of the normative import of the concept of truth (i.e., on the basis of the fact that truth is a distinct norm from justification).

Perhaps the most potent variation of the instability objection, though, gives rise to a truth-property ascription dilemma that in effect calls into question the very coherence of the notion 'deflated property'. Of course, instability/incoherence dilemmas are not entirely new for the notion 'deflated property'. It has been relatively widely speculated that for any property P there must, metaphysical speaking, be some basis for ascribing P to any entity E that bears it, but that the deflationist about any property P does not have the explanatory resources necessary to supply us with the basis for such ascriptions. In other words, it has been surmised, contra the deflationist about properties, that to supply the metaphysical basis for ascribing a property P to any entity E, P must be inflated.

The instability dilemma for WD fits this *general* mold. While its details need some spelling out (see §3), the instability dilemma for WD can be framed in simple terms:

> Assuming truth is a property, like other properties, it is attributable to entities paradigmatically thought to bear it (propositions, sentences, beliefs, etc.). For example, if truth is a property it is attributable to the following statements:

---

[3]. In the postscript to his (1998a), Horwich attempts to short-circuit the instability objection by qualifying WD's (or 'minimalism's') ontological thesis, claiming that WD does not in itself answer the question of whether or not truth is a property, but does so only in conjunction with particular conceptions of *property* (p. 141):

> Minimalism does not involve, in itself, any particular answer to this question. For it may be combined with a variety of different conceptions of property, some of which will yield the conclusion that truth predicate does stand for a property, and some that it doesn't.

On the surface, though, such a maneuver does not appear to confront the heart of the instability objection. Horwich's maneuver, it would seem, can be glossed as follows: depending on what conception of *property* is endorsed, WD implies either i) that truth is a deflated or 'logical' property or ii) that truth is not a property. The problem is that on i), WD is committed to truth being a deflated property of some kind in which case it still gives rise to the instability objection, and on ii), WD is committed to there being no property of truth in which case WD looks indistinguishable from strong deflationism or the so-called 'redundancy theory of truth', a theory widely regarded as patently untenable (see Horwich, note 2).

(S$_1$)   snow is white,

(S$_2$)   grass is green,

(S$_3$)   the earth is round,

and the like. WD, in view of its ontological thesis, is thus subject to the following desideratum: its proponent must furnish us with the basis, metaphysically speaking, for truth-property ascriptions of this kind. In particular, she must answer the following question: what is the basis for ascribing the property of *truth* to all the statements in question? Or to put the point another way, what prevents us from ascribing some other property to the statements in question, or from ascribing a different property to each statement (since each statement has a different truth condition)?[4]

The quagmire is that there appears to be only two basic maneuvers available to the proponent of WD in this case, neither of which seems to enable her to retain her unique, two-pronged ontological thesis about truth — i) that it is a property and ii) that it is deflated. Either 1) she must argue that there is no metaphysical basis for the ascription of a property of truth to truth bearers (there is nothing that makes them all true) in which case she relinquishes thesis i), or 2) she must argue that there is some basis for truth-property ascriptions in which case she relinquishes thesis ii).

3 First Horn of the Dilemma

To begin, consider the first (minor) horn of the dilemma. To elude this horn, the proponent of WD must argue that there is no metaphysical basis for ascribing truth to truth bearers. In other words, what must be argued in this case is that for each and every truth bearer, e.g., [snow is white], [grass is green], etc., there is nothing that makes them true, nothing by virtue of which we can ascribe a property of truth to them. Such a maneuver is clearly a non-starter, though, one that surely will not tempt the proponent of WD. Indisputably, if there is nothing that makes alleged truth bearers true by virtue of which we can ascribe a property of truth to them, entitlement is relinquished for ascribing the property of truth to them.

A cursory examination of property ascriptions bears this point out. The following seems to be a standard principle constraining property ascriptions:

(POP) To be justified in ascribing some property P to an entity or set of entities E relevant grounds G must be available to license the ascription by explaining why or how E possesses P.

It seems uncontroversial, *viz.* (POP), that property ascriptions require relevancy grounds. For example, to be justified in ascribing the property of *tallness* to Evan and Sally one must supply some reason explaining why Evan and Sally are tall, to be justified in ascribing the property of *heaviness* to Rasheed and Joseph one must supply some reason explaining why Rasheed and Joseph are heavy, etc.[5]

---

[4]. In other words, ascribe [snow is white] property A, [grass is green] property B, [the earth is round] property C, etc. (where none of A, B, C and the like are the property of truth).

[5]. This seems to follow regardless of which account of the notion *property* one embraces — whether properties are particulars, universals, tropes, the meanings of the predicates that designate them, or something else. For more on the different theories of properties see Mellor and Oliver's anthology (1997).

But assuming the legitimacy of (POP), it plainly is not a maneuver available to the proponent of WD to contend that there is no basis for truth-property ascriptions — that there is nothing that makes truth bearers true (by virtue of which we can ascribe the property of truth to them). The problem is that claiming there is no basis for truth-property ascriptions violates (POP), since by failing to furnish the basis for truth-property ascriptions, the proponent of WD fails to provide *any* grounds for ascribing the property of truth to truth bearers, let alone *relevant* grounds.

In short, WD cannot elude the first horn of the truth-property ascription dilemma since by claiming there is no metaphysical basis for truth-property ascriptions — nothing that makes alleged truth bearers true — the proponent of WD, assuming the judicious (POP), relinquishes entitlement for ascribing the property of truth to any statement or for claiming that any statement is true. The implication in this case is that WD's view of truth collapses into that of strong deflationism: i.e., no property is ascribed to statements when we describe them as 'being true' (that is, there is no property of truth).

4 Second Horn of the Dilemma

Next, consider the second (major) horn of the dilemma. To elude this horn, the proponent of WD must argue that there is a basis for ascribing a property of truth to truth bearers — there is something that makes truth bearers true that licenses the ascription of a property of truth to them.

Crucially, what requires underscoring at this stage is that there is a widely held, intuitively appealing perspective from which it would appear WD will not be able to invoke this maneuver either. This perspective can best be appreciated by looking more closely at (POP): to be justified in ascribing some property P to an entity or set of entities E relevant grounds G must be available to license the ascription by explaining how or why E bears P. A natural extension of (POP), one that the proponent of WD will reject, is (POP*): to be justified in ascribing some property P to an entity or set of entities E relevant grounds G must be available to license the ascription by explaining why E bears P, *where in order for G to explain why E bears P, G must refer to one or more sortals (i.e., defining characteristics) of E by virtue of which E bears P*. (POP*) stipulates, for example, that to be justified in ascribing the property of *tallness* to Evan and Sally one must refer to one or more sortals characterizing Evan and Sally by virtue of which the property of *tallness* can be ascribed to them (e.g., they both measure at over six feet in height), to be justified in ascribing the property of *heaviness* to Rasheed and Joseph one must refer to one or more sortals characterizing Rasheed and Joseph by virtue of which the property of *heaviness* can be ascribed to them (e.g., they both weigh over two hundred pounds).[6]

---

[6]. Of course, (POP*) does not imply that supplying relevant grounds for ascribing a property to an entity or set of entities must involve appealing to a separable *property* P^{ii} of the entity in question in virtue of which the original property P^i is instantiated. Instead, (POP*) implies that the grounds licensing the ascription of a property to an entity must refer to sortals individuating the property as such, whether such sortals are themselves property constituting or not. The disclaimer renders (POP*) at least *prima facie* plausible by rescuing it from a prompt rebuttal to the effect that if property ascriptions must be grounded in the isolation of a separable property of the target of the ascription a vicious, infinite regress of ascriptive justification will be generated. (POP*), in conformity with this concern, suggests that the regress of ascriptive justification terminates eventually in the isolation of sortals by virtue of which the property of truth is instantiated but that are not themselves property constituting.

The intuitive plausibility of (POP*) stems from its pre-reflectively appealing view of what relevant grounds for property ascriptions must look like. In particular, (POP*) endorses a supposition about property-ascription justifications that is difficult to resist *ceteris paribus*: i.e., relevant grounds G for ascribing a property P to an entity E must *explain* why or how E bears P. Relevant grounds for property ascriptions would seem to need to be *explanatorily* relevant in some sense. It is surmised, though, that the only sorts of grounds G that seem capable of explaining why (or how) an entity E bears P must make reference to one or more sortal characteristics of E. The crucial point is that it is difficult to fathom what might distinguish E as bearing P without recourse to some kind of sortal categorization of E, in which E comes out as a subset of P things. At the very least, this view of things seems prescribed by ordinary canons of explanation and analysis which dictate that only explanations (or analyses) drawing out *explicit* sortal connections between *explanans* and *explanandum* can be sufficiently illuminating.

From the intuitive, (POP*)-based perspective, though, WD's justifications for truth-property ascriptions plainly do not satisfy the explanatory relevance criterion for property ascriptions. In particular, WD cannot justify the ascription of the property of truth P to a statement or set of statements S by appealing to sortal characteristics of S that explain why S bears P. The problem for WD is that it holds that truth is an indefinable and unanalyzable property — a property whose explanatory role is exhausted by its involvement in the schema: [P] is true iff P. This means the proponent of WD is forbidden from providing justifications for truth-property ascriptions with the explicit form deemed necessary for explanatory illumination: justifications that draw out explicit sortal connections between *explanans* and *explanandum*.

5 Second Horn of the Dilemma Revisited

Unquestionably, though, the challenge remains of examining whether the intuitive, (POP*)-based perspective obstructing our opponent's evasion of the second horn of our dilemma is in fact a viable one. In particular, to assess the legitimacy of this intuitive perspective, it is important to examine WD's specific proposal(s) for grounding truth-property ascriptions to determine whether it satisfies uncontroversial requirements for explanatory relevance (e.g., those stipulated by (POP)).

In my view, only one basic option is available to the advocate of WD in this case. This option is pursued in a proposal made by Horwich (1998a, chapter 7).[7] Horwich proposes that the basis for truth-property ascriptions is in effect that truth bearers in a specific sense 'legitimately instantiate the T-schema' (not the standard reading of this phrase): [P] is true iff P. In particular, Horwich claims that the license for ascribing the property of truth to a proposition, sentence or belief is that it *legitimately instantiate the T-schema* in the following sense: i) it can be plugged into the schema, and ii) the right-hand side of the schema in some sense *comes to pass*. In other words, the license for ascribing the property of truth to [snow is white] derives from the legitimacy of the corresponding T-sentence — [snow is white] is true iff snow is white — and that snow *is* white, the license for ascribing the property of truth to [grass is green] derives from the corresponding T-sentence — [grass is green] is true iff grass is green] — and that grass *is* green, and so forth.

---

7. Horwich's explicit formulation of the proposal differs somewhat from that given in the text. But Horwich's basic line of reasoning is captured by our formulation of the proposal.

To be sure, this maneuver is *a fortiori* plausible. It supplies us with the only justification for truth-property ascriptions that seems permissible on the terms of WD: a statement must *legitimately instantiate the T-schema*. Upon closer examination of the functioning of the T-schema, though, — how it functions, in conjunction with actualities (e.g., that snow *is* white) to pick out certain kinds of statements and not others as true — it becomes less manifest that the grounds supplied by Horwich for truth-property ascriptions satisfy uncontroversial criteria for explanatory relevance (stipulated by (POP)).

A potential concern with Horwich's maneuver in this case is that any explanation of the T-schema, how it functions to only pick out true propositions, must amplify the material equivalence of the '[P]' and 'P' upon which our understanding of the T-schema is founded, if it is to be minimally comprehensible and thus satisfy an indisputable relevancy requirement (stipulated by (POP)). For example, consider one such explanation of the T-schema:

(TS) A statement is true (i.e., legitimately instantiates the T-schema) if and only if what it says to be the case is in fact the case (compare with Boghossian, 1990).

To be sure, there are other ways of explaining the T-schema, but if I am correct an explanation of this general *form* will be needed: an explanation deploying semantic notions or phrases such as 'saying what is the case', 'expressing what is the case', 'facts', and so forth. The problem with any such explanation, of course, is that it incurs ontological commitments forbidden by WD, since it says things about truth its advocate is not permitted to say (given her ontological thesis),[8] things that would seem to commit her to truth being a property with an underlying nature (i.e., some sort of correspondence between statements and facts). Patently, to analyze truth (or derivatively truth-bearing) via the deployment of semantic *explanans* is to commit oneself to an inflationary view of truth.

Not surprisingly, Horwich has a reply to this kind of objection to his proposal for justifying truth-property ascriptions (1998a, pp. 34-35 and 50-51): the notion of *legitimately instantiating the T-schema* does not require amplification since the T-schema is explanatorily basic or primitive. In addition to Horwich, Soames (1999, p. 231) also endorses some such form of the so-called 'primitivist thesis' regarding the T-schema.

The question is: what might be meant by the claim that the T-schema that truth bearers must legitimately instantiate is explanatorily basic? The primitivist thesis is difficult to pin down but is typically fleshed out in the following way (Horwich (1998a, pp. 50-1, 121) and (1998b)): the T-schema is explanatorily basic in the sense that the need to explain its functioning does not arise. The interpretation of the T-schema, according to this line of argument, is not a thing it is *possible* to explain inasmuch as it is already *implicitly* fixed by our dispositions to assent to its instantiations.[9]

---

8. Refer back to section 1.

9. And in effect, according to this line of argument, such dispositions fix the meaning of the word 'true'. This is the sense in which, for Horwich, the dispositions to assent to the instances of the T-schema (in his case, the equivalence schema) are supposed to be truth constituting (1998a, pp. 34-35). Many of the background assumptions underlying this line of argument can be found in Horwich's (1998b). Soames does not make this sort of claim in defense of the explanatory primitiveness of the T-schema. But since neither Soames nor any other proponent of WD provides an alternative defense of the primitivist thesis, we can restrict our examination to Horwich's dispositionalist defense.

If the interpretation of the T-schema is implicitly fixed by our dispositions to assent to its instantiations, the obligation to explicitly explain it, i.e., by importing inflationary semantic notions or phrases such as 'saying something to be the case', 'expressing the fact that', etc. will be discharged. The snafu is that the Horwichian dispositionalist defense of the primitivist thesis about the T-schema flies wide of the mark: dispositions to assent to the instances of the T-schema seem not to implicitly fix its meaning. Apparently, there are a wide variety of questions regarding how any version of the T-schema (e.g., the disquotational schema, the equivalence schema, etc.) is to be read that are not resolvable by appeal to dispositional assent.

Consider Horwich's version of the T-schema, the equivalence schema:

(ES)   <P> is true iff P (where '<P>' stands for 'the proposition *that* P' and 'P' stands for <P>'s object-level equivalent).

Instances of this schema include

(ESI)   the proposition *that* snow is white is true iff snow is white,

(ESI)   the proposition *that* grass is green is true iff grass is green,

and the like. As it has recently been argued (e.g., Collins (2002, pp. 668-674), Davidson (1990) and Wiggins (1980)), one crucial aspect of how to read the equivalence schema (and its various instances) that is not fixed by our disposition to assent to propositions such as these is how to read the 'iff' operator in it. There are a variety of ways in which the 'iff' operator might be read in the equivalence schema (or in any version of the T-schema). Does it represent extensional equivalence, intensional equivalence, cognitive equivalence, etc.?[10] Dispositions to assent to the instances of the equivalence schema are of no help at all here in isolating which of these readings is the correct one. Indeed, our disposition to assent to the instances of the equivalence schema seems at least partly to be *a function of* how the 'iff' operator is to be interpreted, implying *a fortiori* they cannot be *constitutive* of its interpretation. For example, if the 'iff' operator in the equivalence schema only represents extensional equivalence, we will be more disposed to assent to the propositions that instantiate it. If, on the other hand, 'iff' represents cognitive equivalence (a stronger relation), we will be less disposed to assent to the propositions that instantiate the equivalence schema. In short, bare dispositional assent does not implicitly fix the meaning of the equivalence schema in view of the widely acknowledged fact that there are crucial aspects of its interpretation that bare dispositions tell us nothing about such as the question of how the 'iff' operator is to be read in it.[11] The same point applies *mutatis mutandis* to any other version of the T-schema.

Perhaps, then, since *bare* dispositions will not suffice to fix the meaning of the T-schema, the dispositionalist might appeal to *patterns* of sets of dispositions to assent to the instances of the T-schema to fix its meaning (e.g., to distinguish between different readings of the 'iff' operator). One question this proposal raises is: what are 'patterns of sets of dispositions'? Presumably, they are sets of dispositions with structural relations to one another.

---

[10].   There is no standard reading of 'iff' the proponent of WD can appeal to in this connection.

[11].   Aside from questions concerning the proper interpretation of the equivalence schema, another important question the Horwichian dispositionalist defense of the primitivist thesis leaves open is why we should have dispositions to assent to *true* statements and not *false* statements, or statements bearing some other kind of property.

The crucial difficulty with such a proposal is that *patterns of sets of dispositions* (or *dispositions with structural relations to one another*) is plainly not a natural kind: mere dispositions do not stand in explanatory or structural relations to one another. Indeed, as Collins has pointed out (2002, pp. 672-3), the only way to discern the structural relations between sets of dispositions is to appeal to an underlying theory of semantic competence — a theory of what it is to competently speak and understand a language. Plainly, though, in this case the theory of semantic competence would be doing the explanatory work.

In short, the problem is that the only proposal available to the defender of WD that I am aware of (and the only proposal on the table) involves justifying truth-property ascriptions via Horwich's criterion of *legitimately instantiating the T-schema*, and then claiming that the T-schema does not need to be explained (to satisfy the relevancy requirement) since it is in some sense explanatorily primitive. But it would seem that no version of the T-schema can be explanatorily primitive in the way this line of argument requires: its interpretation is not implicitly fixed by dispositional assent (either bare dispositional assent or patterns of sets of dispositional assent). The implication is that WD seems unable to forestall the requirement of explicitly explaining how the T-schema functions, in conjunction with actualities, to discriminate between true and false statements, and that it needs to invoke inflationary semantic notions to do so.

Sustained reflection thus bears out our intuitive supposition that the defender of WD seems unable elude the second horn of the dilemma: she cannot furnish us with a justification for truth-property ascriptions that satisfies basic relevancy requirements without recourse to inflationary *explanans*. The implication in this case is that WD's view of truth collapses into some form of inflationism.

6 Conclusion

In the end, then, WD is on shaky ground: it appears to be fundamentally unstable. The problem is that WD's ontological thesis about truth — that it is both a property and deflated — generates a truth-property ascription dilemma the proponent of WD is at pains to circumvent. In attempting to elude the first horn of the dilemma, WD's view of truth collapses into that of strong deflationism. In attempting to elude the second horn of the dilemma, WD's view of truth collapses into that of some form of inflationism.

Importantly, the deficiency in WD's ontological thesis seems to issue from its failure to conform to a view regarding the nature of property-ascription justifications that is difficult to dismiss. Intuition and sustained analysis seem to bear out that the justification of property ascriptions are governed by (POP) and its natural extension (POP*): to be justified in ascribing some property P to an entity or set of entities E relevant grounds G must be available to license the ascription by explaining why E bears P, *where in order for G to explain why E bears P, G must make reference to one or more sortals of E (by virtue of which E bears P)*. (POP*)'s legitimacy implies the folly of taking the ontological middle road when it comes to alethic theorizing — of claiming that truth is both a property and deflated. More generally, though, (POP*)'s legitimacy implies the folly of taking the ontological middle road when it comes to theorizing about properties in general (of claiming that *anything* can both be a property and deflated), and for this reason, has enormous bearing on the prospects for deflationary metaphysical positions.

References

Ayer, A.J. (1936) *Language, Truth, and Logic*, Penguin, London.

Boghossian, P. (1990) 'The Status of Content', *Philosophical Review* 99, pp. 157-184.

Collins, J. (2002) 'On the Proposed Exhaustion of Truth', *Dialogue* 41, pp. 653-679.

Davidson, D. (1990) 'The Structure and Content of Truth', *Journal of Philosophy* 87, pp. 279-328.

Field, H. (1994) 'Deflationist Views of Meaning and Content', *Mind* 103, pp. 249-258.

Grover, D.L., Camp, J.L., and Belnap, N.D. (1975) 'A Prosentential Theory of Truth', *Philosophical Studies* 27, pp. 73-125.

Gupta, A. (1993) 'Minimalism,' *Philosophical Perspectives* 7, pp. 359-369.

Horwich, P.G. (1998a) *Truth*, second edition, Clarendon Press, Oxford.

———— (1998b) *Meaning*, Clarendon Press, Oxford.

———— (1999) 'The Minimalist Conception of Truth', in S. Blackburn and K. Simmons (eds.), Truth, Oxford University Press, Oxford, pp. 239-263.

Mellor, D. H. and Oliver, A. (eds.) (1997) *Properties*, Oxford University Press, New York.

Ramsey, F.P. (1927) 'Facts and Propositions', *Proceedings of the Aristotelian Society* 7.

Soames, S. (1997) 'The Truth About Deflationism', in E. Villaneuva (ed.), *Philosophical Issues* 8, pp. 79-94.

———— (1999) *Understanding Truth*, Oxford University Press, New York.

Sosa, E. (1993) 'The Truth of Modest Realism', in E. Villaneuva (ed.), *Philosophical Issues* 3, pp. 77-95.

Tarski, A. (1983) *Logic, Semantics, Metamathematics*, 2nd ed., Hackett, Indianapolis.

Wright, C. (1992) *Truth and Objectivity*, Harvard University Press, Cambridge.

———— (1996) 'Response to Commentators', *Philosophy and Phenomenological Research* 56, pp. 863-868, 911-941.

————————————————————

Glen Hoffmann
Department of Philosophy
Ryerson University
<ghoffman@ryerson.ca>

THE SORITES Charter
· Version 2.2 ·
Copyright © 1995-2006 by the SORITES team

PREAMBLE

This Charter is an updated and amended version of the original «Copyright Notice and Legal Disclaimer» appeared in Issue # 01 of *SORITES* (April 1995). It is a covenant between the **SORITES** team, as a publisher, and its contributors and readers, prescribing their respective duties and rights, as well as the terms and conditions under which those rights have to be exercised.

Persons dealing with **SORITES** in either capacity, as contributors or as a readers (or users), bind themselves to abiding by these regulations, as does the **SORITES** team itself.

This Charter aims to safeguard all legitimate interests in a balanced way, so as to ensure maximum access to, and use of, the materials we produce, to serve the scholarly public at large, to enhance the prestige of our contributors and the influence of the journal and its publisher, while preventing dishonourable behaviour and academic fraud.

This Charter bears a family resemblance to various schemes: the GNU Free Documentation, Creative Commons, Creative Archive and Open Content licenses. Moreover, the current version borrows a few notions and phrases from them. Nevertheless this Charter stands on its own and must not be interpreted as adjusted to any of those models. Although all those approaches share a strong leaning towards open and free content-divulgation, a distinctive view prevails in each case.

Openness and freedom are broad and loose concepts admitting of sundry conceptions. 'Free content' has been commonly agreed upon to stand for content which readers can and may honestly use without hindrance or threat of reprisals — rather than as content they necessarily get for free — but there are several views as to the justification and implementation of those liberties. A thorough elucidation of those problems goes beyond the scope of this Preamble, though.

In order for the current scheme to stand out as a differentiated dispensation, it will be called '**the allodial approach to intellectual property**'. Four features make it unique:

 — The first one is that, while it allots to everybody an unhampered access to intellectual goods **without discrimination**, it also recognizes the right to private intellectual ownership, construed — according to Civil-Law systems — as twofold: an entailed, perpetual, untransferable and non-renounceable entitlement to authorship-recognition and to the integrity of the work; and an alienable right to the exclusive use and profit from that work.

 — The second one is that it regards a compilation (such as a journal Issue) as a collective work, and hence as an intellectual entity jointly authored and owned by the Editor and the contributors taken together as a group, while each article, by itself, is an intellectual entity belonging to its author alone.

 — The third one is that it resorts to the legal notion of *servitude* or *easement*, which enables us to reconcile all those rights and liberties, in the same way as happens with land freeholds: certain positive or negative interests, franchises or privileges (*servitudes*) may be conferred on non-owners over an estate for certain legitimate reasons and purposes.

 — The fourth one is that it rules out any registration requirement, as contrary to the rights to privacy, unobstructed access and uncontrolled dissemination.

* * *

The **SORITES** team will publish revised versions of the **SORITES** Charter, similar in spirit to the present version, but different in detail, to address new problems or concerns. Authors and readers having obtained certain rights in virtue of the current Charter will keep those rights and will never be subject to restrictions brought in by later versions.

Everybody is free to copy this Charter, and to modify or attune it to their own needs, provided appropriate authorship credit is given to the **SORITES** team.

DEFINITIONS

[D01] A **separate material** is an intellectual product which is released, broadcasted or published as an intangible entity standing on its own rather than as a section or a division of another intellectual product.

Parts of an aggregated work — such as essays gathered into a collection — are not separate materials, since they are not released as intellectual products on their own.

[D02] **An Issue of the *Sᴏʀɪᴛᴇs* journal as a whole** is either:

(i) a single document or file (be it its official version or any unofficial rendering also released by the Sᴏʀɪᴛᴇs team as an undivided file, such as the PDF file or a print-out thereof); or

(ii) a set of files produced by slicing one of the entire-file versions in order to facilitate handling, browsing or downloading (in which case the conveyor is bound to distribute the whole collection taken together at the same time).

[D03] A reader of Sᴏʀɪᴛᴇs is an individual who, abiding by all the terms laid down in this Charter, accesses the Sᴏʀɪᴛᴇs site, or a mirror thereof, or receives or uses any file or document directly or indirectly proceeding from that site.

[D04] As a rule, institutions and organizations are not readers. They are considered users only when this Charter explicitly confers a certain right on them.

[D05] Private use is personal use by an individual. Sharing downloaded or stored files within an organization is a divulgation, not private use.

[D06] An Issue of the journal *Sᴏʀɪᴛᴇs* is an aggregated work; the articles it contains are not separate materials.

[D07] A copy of a material is either a print-out on paper or a reproduced file or set of files.

[D08] A reproduced file, or a set of files, is either (i) a duplication of the original file, or set of files, as released by the Sᴏʀɪᴛᴇs team; or (ii) any result of converting the downloaded file or files to a different format by applying conversion tools, provided the content is not garbled or curtailed.

[D09] A different format is any modification which leaves the content intact and unaltered.

[D10] The phrase 'copy or partial copy' covers any print-out, extended quotation, electronic posting or display as well as any other reproduction.

————————————————————

ARTICLES

Art.01. — Sᴏʀɪᴛᴇs is both a resource site, <http://sorites.org> (or set of sites, such as <sorites.info>, <sorites.net>, etc), and a digital journal, *Sᴏʀɪᴛᴇs*, consisting in a series of issues. Both of them belong to the publisher, the Sᴏʀɪᴛᴇs team.

Art.02. — Each Issue of the *Sᴏʀɪᴛᴇs* journal as a whole is a collection, an aggregated work, jointly owned by the contributors and the publisher.

Art.03. — Individual papers remain the private property of their respective authors (the contributors) encumbered with the following servitudes:

(i) No part of any such paper may be printed or displayed elsewhere, or incorporated into a publication of any sort, unless *Sᴏʀɪᴛᴇs* is therein clearly and explicitly acknowledged and cited as the original source of publication.

(ii) In particular, when a contributor posts a version of his or her paper on a home page, institutional web-site or subject repository, a link must be made to the version displayed at the Sᴏʀɪᴛᴇs web-site.

(iii) When contributors merge a recast version of a substantial portion of an article published in *Sᴏʀɪᴛᴇs* into another paper, they must also give due credit to the *Sᴏʀɪᴛᴇs* source and clearly inform their audience how to access the Sᴏʀɪᴛᴇs site (including a reference to the official web-site, **http://sorites.org**).

(iv) Contributors adhere to the other terms and provisions contained in this Charter, grant Sᴏʀɪᴛᴇs readers and users the rights provided by Art.06 and consent to abide by the publisher's claims and commitments (arts. 04, 13 and 14).

(v) When contributors republish their papers, they are free to apply any other License, provided clauses (i), (ii) and (iii) of this article are honoured.

(vi) Any individual who signs a co-authored article is a contributor who accedes to the above burdens being laid on the paper. Arrangements between co-authors cannot stand in the way of that undertaking.

Art.04. — By submitting a manuscript to *Sᴏʀɪᴛᴇs*, contributors declare: that it is an output of their personal research; that it has not been published yet (in the sense of appearing under an ISSN or an ISBN); that it is not currently submitted for publication elsewhere; and that it complies with established Copyright laws. Acceptance of a manuscript by the Sᴏʀɪᴛᴇs team is done in good faith under the assumption the declaration is truthful.

Art.05. — The authors of the included papers and the publisher — whether jointly or separately, as the case may be — hereby reserve all rights not conferred to other parties in this Charter.

Art.06. — Subject to honouring the conditions set forth in this Charter, readers are hereby accorded the following rights:

[i] The right to freely access, view, read, search, store, and link to, any material offered at the **SORITES** site, especially any Issue of the *SORITES* journal, any paper therein included, and any other file or document displayed or supplied at the **SORITES** site.

[ii] The right to duplicate single detached articles for a personal, private and fair use done in good faith. Fair use encompasses teaching, study, criticism and review.

[iii] The right to reproduce and distribute on any medium *verbatim* copies of any **separate material** released at the **SORITES** site, provided the material is reproduced **in its entirety** and preserved as an Invariant Document, unaltered in its text, titles and headings.

Art.07. — When a copy of a separate material published by **SORITES** is incorporated into a volume or a package containing other materials, the content obtained from the **SORITES** site must be printed on separate sheets, clearly demarcated, by page breaks, from those accompaniments. No text or logo from other sources must be mingled with, or interpolated into, the **SORITES** material.

Art.08. — The freedom awarded to readers in Art.06 means that:

(1) No fee will be charged or requested by the **SORITES** team.

(2) No registration will be demanded.

(3) No one will be locked out of accessing the **SORITES** site.

(4) As regards issues of the *SORITES* journal and papers therein published, exercising those rights is not conditional upon reciprocation or remuneration of any sort, whether moral or material, or upon any consideration other than respect for intellectual property claims raised in this Charter.

(5) As regards other materials offered at the **SORITES** site, moral consideration may be demanded, such as not using those materials for any unethical purpose or activity.

Art.09. —

[1] Private business companies are not entitled to incorporate a copy of an Issue of the *SORITES* journal into any commercial material they sell for profit without the previous conformity of the **SORITES** team.

However they are hereby allowed to combine a whole Issue of the *SORITES* journal — or any other material offered as a separate unit at the **SORITES** site — into publications they circulate at no charge, provided they observe the other provisions of this Charter, especially Art.07.

[2] Learned societies and academic establishments are hereby authorized to sell integral copies of issues of the *SORITES* journal or any other material offered as a separate unit at the **SORITES** site.

[3] Search engines and indexing data-bases are entitled to store, reproduce and display integral or partial copies of any material published by the **SORITES** team, including issues of the *SORITES* journal, even of detached papers, provided due credit and links to the original **SORITES** site (or a mirror thereof) are ostensibly shown.

Art.10. —

[i] Except as provided by Art.09[3] or by fair use legal provisions, no copy or partial copy of a detached paper may be distributed without its author's acquiescence.

[ii] The **SORITES** team hereby authorizes any such transmission subject to the author's approval, provided the copyright box on top of the paper is literally incorporated into the copy or partial copy.

Art.11. —

[i] No part of any Issue of the *SORITES* journal may be delivered to a plurality of individuals — whether in writing or by any other means — unless the source is clearly and explicitly recognized. Furthermore, such a recognition must comply with the following stipulations.

[ii] No part of any Issue of the *SORITES* journal 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 the *SORITES* journal, the author's name and the paper's full title.

[iii] Whenever the quotation occurs within a publication, it is also mandatory to mention the date of publication and the official pages (as shown within the Copyright box on top of the paper), as well as the ISSN (1135-1349) and the official web-site: **http://www.sorites.org**, or a mirror thereof such as **http://www.sorites.info**.

[iv] When the quotation takes place in an electronic medium, a link to the **SORITES** site (or a mirror thereof) is also imperative.

Art.12. —

(1) Any use of copies or partial copies of an Issue of the *SORITES* journal, or of papers therein included, not permitted in this Charter will be a violation of the contributors' and the publisher's intellectual property.

(2) Distribution of copies of detached papers without the author's consent is hereby prohibited unless pursuant to Art.09[3].

(3) Derivative works are unlawful, except in so much as they are provided by Art.09[3].

(4) Shortening, abridgment or truncation is only licit as provided either by Art.09[3] or else by fair use terms in accordance with intellectual property legislation.

(5) In any case, the two following activities are absolutely forbidden: 1. change of words or any other semantic or syntactic alteration; 2. false, faulty, obscured or disingenuous attribution.

(6) The **SORITES** team reserves the right to sue malicious or reckless transgressors of this Charter as culprits of plagiarism, forgery or other misdemeanours, requesting both compensatory and punitive remedies. It also reserves the right to publicly denounce serious copyright violations.

Art.13. —

[i] Any Material displayed at the **SORITES** site, including issues of the *SORITES* journal, is provided *as is*, without any guarantee of any kind.

[ii] While the **SORITES** team tries to ensure the correctness of all the content it publishes, it makes no representation or warranty as to the accuracy, completeness or fitness for any purpose of such a content and — to the maximum extent allowed by law — disclaims any pledges or assurances, whether express or implied, to the effect that any published material is free of defects or non-infringing.

[iii] By furnishing those documents, the **SORITES** team does not underwrite any commitment except in so much as explicitly espoused in the present Charter.

[iv] Any views expressed in papers published in **SORITES** are the opinions of the contributors and not those of the **SORITES** team. The publisher disowns being answerable for the contributors' ideas or arguments.

Art.14. — Under no circumstances and under no legal theory shall the **SORITES** team, or any contributor or supplier of **SORITES**, be liable to any person for any indirect, special, incidental or consequential damages for loss of good will, mistaken belief, confusion, delusion or any and all other damages or losses suffered as a result of downloading, reading, or using any materials retrieved from the **SORITES** site or included in any Issue of the *SORITES* journal. The user assumes, at their sole risk, full responsibility for relying upon any opinion or information contained in any material published or displayed by **SORITES**.

Art.15. —

[i] By opening, accessing, downloading, storing, reading or in any other way using any material offered at the **SORITES** site or any part thereof, readers enter into a synallagmatic contract with the **SORITES** team binding themselves to full acceptance of the above stipulations. Breach of any of those provisions entails automatic forfeiture of all rights derived from this Charter and termination of the contract.

[ii] If, upon downloading or storing a file containing an Issue of the *SORITES* journal or a part thereof, or receiving an electronic or printed copy thereof, or any other material obtained from the **SORITES** site, a reader fails to assent to the conditions stipulated in this Charter, he or she is obliged to immediately discontinue reading, storing, sharing or using the material in any way.

[iii] In the event of such a termination, though, third-party rights derived from usage prior to the termination will survive.

[iv] Readers to whom previous versions of this Charter bestowed rights in excess of those granted here, or subject to less stringent conditions, may continue enjoying those rights.

[v] The **SORITES** team will revise this Charter at its sole discretion. Acquired rights will endure, though.

[vi] In the event that there exists any inconsistency between this Charter and legal commandments, the latter will take precedence.

[vii] If a reader cannot satisfy simultaneously the duties under this Charter and commitments stemming from custom, contract of other sources, they must either extricate themselves from those commitments or else refrain from accessing **SORITES** and from using any material released by the **SORITES** site.

[viii] This Charter is the complete and exclusive agreement between the **SORITES** team, its contributors, and its readers which supersedes any proposal or prior deal, oral or written, and any other communication relating to the subject matter of this Charter.

Art.16. — Rights granted in this Charter are contingent upon third-party intellectual-property claims.

Art.17. — In compliance with Spanish Law, all issues of the *SORITES* journal are legally registered, electronic copies being deposited with the competent authorities, namely the Legal-Deposit office of the Autonomous Community of Madrid (Registration: M 14867-1995.)

Art.18. — Although a mirror of **SORITES** is displayed at the internet host <www.ifs.csic.es/sorites/> belonging to the Spanish Academy of Sciences, or C.S.I.C., thanks to a permission graciously given by the system's administrators, the journal is not sponsored or endorsed by that institution, the only owner and publisher being the **SORITES** team.

––––––––––––––––––––––––––––––

Madrid. 2006-10-05 [V.2.2]
(Previous versions: 1995-04-10; 1996-12-31; 1998-04-27; 2002-10-03; 2004-12-31;
2005-01-18 [V.2.1])
Lorenzo Peña (The **SORITES** Team)
<http://sorites.org/copyrigh.htm>

Release Notice