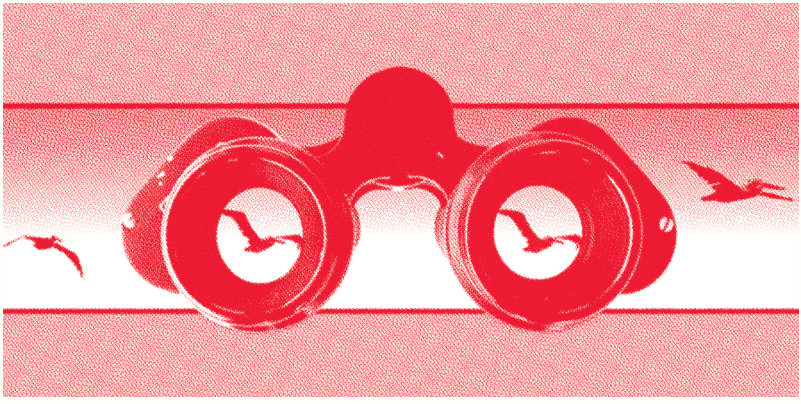
ORIGINALES

Time, Thought and Vulnerability

An Inquiry in Cognitive Dynamics

THE SECOND JUAN LARRETA LECTURES (2017)

PAULO ESTRELLA FARIA



Time, Thought and Vulnerability presents the results of an investigation of the conditions under which certain circumstances that are beyond the knowledge (and therefore, the control) of the agents can affect the validity of inferences whose correctness is supposed to be purely *a priori* evaluable. The discussion involves a careful examination of the controversy over the compatibility of anti-individualism and the "authority of the first person", within the background of the controversy over the transparency of the mental content and the accessibility of the logical form of inferences. A comparison with the debate on the vicissitudes of preservative memory in a temporalist semantics is articulated by examining the arguments presented by Mark Richard and Paul Boghossian, respectively, against temporalism and anti-individualism. Finally, the debate on the transparency of mental content is dissociated from those two theoretical frameworks (temporalism versus eternism; individualism versus anti-individualism) by engaging in a direct discussion of the postulate of the transparency of logical form.

Sociedad Argentina De Análisis Filosófico

Time, Thought, and Vulnerability

Time, Thought, and Vulnerability

An Inquiry in Cognitive Dynamics

The Juan Larreta Lectures

2017

Paulo Estrella Faria

SADAF

Faria, Paulo Estrella
Time, Thought, and Vulnerability : An Inquiry in Cognitive Dynamics. The Juan Larreta Lectures, 2017 / Paulo Estrella Faria . - 1a ed . - Ciudad Autónoma de Buenos Aires : SADAF, 2021.
Libro digital

ISBN 978-987-47781-3-0

1. Filosofía. 2. Epistemología. 3. Filosofía del Conocimiento. I. Título.
CDD 121

SADAF
www.sadaf.org.ar

Cover design: Iñaki Jankowski | www.jij.com.ar
Cover image: Luiza Estrella
Editing: Recursos Editoriales

isbn: 978-987-47781-3-0

Table of Contents

Foreword

Honoured by the Argentine Society for Philosophical Analysis (SADAF) with the invitation to deliver the 2017 Juan Larreta Lectures, I took the occasion to probe into a set of interrelated problems belonging to what David Kaplan termed 'cognitive dynamics'—a field of philosophical inquiry which deals with the conditions on which propositional contents are retained, reiterated, redeployed, modified, lost and (sometimes) retrieved. My overall aim was to try and bring together two strands in my recent work which had until then followed parallel paths in a seemingly Euclidean space, all but avoiding one another. These are *the philosophy of time* (concentrating on the temporalism vs. eternalism debate, hence on the very notion of a temporal proposition: a proposition whose truth-value changes with time), and *the epistemology of reasoning* (concentrating on the individualism vs. anti-individualism debate, hence on the very notion of a world-involving thought, specifically as that impinges upon what Paul Boghossian calls 'the apriority of our logical abilities'). That's why I called the first two lectures, respectively, 'Transience' and 'Preservation'. In these two lectures I claim, in particular, that Mark Richard's 1981 argument from belief retention against temporalism and Paul Boghossian's 1989 argument from preservative memory against anti-individualism share a common structure and a crucial pair of analogous premises; that they invite analogous responses (which have actually been put forward); *and* that such responses fall short of fully taking the sting off the original arguments, given precisely the way they are interrelated. Put it as follows. Suppose temporalism is true; then

we have a problem (Richard's problem) about content preservation. Suppose anti-individualism is true; then we have a problem (Boghossian's problem) about content preservation. Since I hold that both problems are real and stand unsolved (for good reason: they are, or so I will argue, strictly unsolvable), my defence of both temporalism and anti-individualism eventuates in an investigation of the varieties of conceptual (and other) losses which are the lot of creatures whose cognitive lives are such that both temporalism and anti-individualism are true of them—hence the title of the third and last lecture.

I am grateful to the audiences at SADAF for the lively discussions. Running the risk of forgetting someone (for which I apologise in advance), I single out Alberto Moretti, Diana Pérez, Eduardo Barrio, Eleonora Orlando, Federico Penelas and Sandra Lazzer, whose contributions were warmly appreciated. I hope the result does not disappoint them.

Porto Alegre, May 2018

1. Transience

Unfathomable Sea! whose waves are years,
Ocean of Time, whose waters of deep woe
Are brackish with the salt of human tears!
Thou shoreless flood, which in thy ebb and flow
Claspest the limits of mortality!

And sick of prey, yet howling on for more,
Vomitest thy wrecks on its inhospitable shore;
Treacherous in calm, and terrible in storm,
Who shall put forth on thee,
Unfathomable Sea?

Percy Bysshe Shelley, "Time" (1821)

My aim in these lectures is to bring together two strands in my recent work which up to now have travelled parallel paths, all but avoiding each other. These are *the philosophy of time* (concentrating on the temporalism vs. eternalism debate, hence on the very notion of a temporal proposition: a proposition whose truth-value changes with time), and *the epistemology of reasoning* (concentrating on the individualism vs. anti-individualism debate, hence on the very notion of a world-involving thought, specifically as that impinges upon what Paul Boghossian is wont to call "the apriority of our logical abilities").

Both strands converge in issues of cognitive dynamics as that philosophical discipline was defined by David Kaplan almost 30 years ago: cognitive dynamics is the study of the conditions on which propositional contents are retained, reiterated, redeployed, modified, lost and (sometimes) retrieved.[1]

Our starting point is an examination of the central question in the philosophy of time, namely, how to think of transience.

1 See Kaplan (1989a: 537).

Let's start with a couple of definitions.[2] *Eternalism* is the thesis that for every proposition *p*, and every bit of time information i_t needed for truth evaluation, i_t is specified in *p*. *Temporalism* is the thesis that there are exceptions to eternalism. That is, for some proposition *p*, and some bit of time information i_t needed for truth evaluation, i_t is unspecified in *p* (equivalently: *p* is neutral with respect to i_t).

According to eternalism, then, if a proposition is once true (or false), it is always true (or false). According to temporalism, some propositions have a changing truth-value. Notice that eternalism is the most ambitious thesis, as it holds that *every* proposition, if it has a truth-value at all, has an unchanging truth-value, while temporalism has it that *some* propositions have changing truth-values.

Our first question, then, is this: are there (as Aristotle, the Stoics and the Schoolmen thought) temporal propositions, namely, propositions whose truth-value is relative to some occasion (be it that of their utterance, that of their evaluation, or yet another), which, in consequence, can be reiterated, their identity preserved, even though their truth-value (relatively to each relevant occasion) is variable? Or should every sentence whose utterance only has a truth-value relatively to some occasion be construed (as Frege and Russell held) as the expression of a propositional function in which at least one free variable (usually unarticulated in the "surface grammar") takes as arguments instants or time-intervals?[3]

---

2 I borrow the wording from Schaffer (2012).

3 An interesting (and rather thorny) further question, which I am glad to leave for another occasion, is whether an affirmative answer to the *first* of the above questions implies the denial of the thesis (which usually prompts the affirmative answer to the *second* question) according to which the truth predicate does not admit adverbial modification. Alternatively: is it possible to treat tenses and time adverbials, by analogy with modal operators, in such a way that, for a designated time (as, in the modal case, for a designated world), the predicate 'true at *t*' (as its modal counterpart, 'true at *w*') is reducible to the predicate 'true' (*simpliciter*)? A decision about that will be of consequence to the issue of the possibility (if any there is) of a temporalist alternative to presentism—the thesis, championed by Arthur N. Prior, according to which

It is a remarkable fact that the affirmative reply to our first question has prevailed throughout the greatest part of the history of philosophy, and that only in modern times (starting, actually, in the 18th century) there has been a gradual articulation of what was to become the canonical view in Contemporary logic and philosophy: namely, the choice of the second of the previously mentioned alternatives.

Indeed, the assumption that the answer to the first question is affirmative, and that the answer is unproblematic, is distinctive of the manner in which the relations between logic and time were conceived of in Ancient and Medieval philosophy.

In *De Interpretatione,* Aristotle writes: "Every statement-making sentence must contain a verb or an inflexion of a verb. For even the definition of man is not yet a statement-making sentence—unless 'is', or 'will be' or 'was' or something of this sort is added" (17^a9ss). Such explicit mention of the verbal tenses is not accidental. As Hintikka writes, "for Aristotle the typical sentences used in expressing human knowledge or opinion are not among those Quine calls *eternal sentence*s (or even among *standing sentences*) but among those Quine calls *occasion sentences*. That is to say, they are not sentences to which we assent or from which we dissent once and for all. They are sentences to which we can subscribe or with which we must disagree on the basis of some feature or features of the occasion on which they are uttered (or written). In particular, the sentences Aristotle is apt to have in mind are *temporally indefinite*; they depend on the time of their utterance" (Hintikka 1973: 64).

The idea is not that the time of utterance supplies, as in Frege and his successors, the argument of a propositional function, so that "It is sunny in Buenos Aires" turns out to be, at the time I utter it, an incomplete expression of the proposition <*It is sunny in Buenos Aires now*> (that is, at 6 PM, November 7, 2017). Rather, the property that Hintikka calls temporal indefinition is a char-

---

the domain of variation of our most unrestricted quantifiers comprises only presently existing entities, or, in other words, 'present' and 'real' are synonyms (cf. Prior 1972).

acteristic of a *complete* proposition whose truth-value changes with time.

The idea of a proposition which is in *that* sense temporally indefinite—the idea of a temporal proposition—and the imbrication of logic and temporality which is required to account for such propositions, make possible, for instance, to raise the problem of future contingents the way Aristotle does it in Chapter IX of *De Interpretatione*. As Prior remarks, the idea that a complete proposition may have different truth-values at different times sheds light on Aristotle's conjecture that "There will be a sea-battle tomorrow" might be (because of the indeterminacy of the situation) "not yet" definitely true or definitely false: "That things might change to being true or false from not being definitely either, is certainly a more radical view than that they might change from being true to being false and vice versa, but it is not as far from this as it is from the view that the passage of time is quite irrelevant to the truth and falsehood of propositions" (Prior 1967a: 16).

The Master Argument of Diodorus Cronus is another example of the pertinence of temporal considerations for Ancient logic. Indeed, it can be understood as an attempt at clarifying, through the demonstration of the inconsistency of a set of propositions drawn from Aristotelian philosophy, the relations between time and modality. For our purposes, Diodorus' aim (a proof by reduction of fatalism) is less important than the fact that his argument relies on an explanation of modal notions in terms of temporal propositions: everything which is *past* and *true* is *necessary*; *possible* is what *is* or *will* be true.[4]

In 1949 Benson Mates published a paper entitled "Diodorean Implication", later to become a chapter of his book *Stoic Logic*. In

4 This is how Epictetus, our most reliable source, presents it: "The Master Argument seems to have been put forward on the basis of some such principles as the following. These three propositions are irreconcilable in so far as any two contradicts the one that is left over: (1) that everything that has come about in the past is necessarily the case, (2) that the impossible cannot follow from the possible, (3) that something can be possible that is not true at present nor ever will be in the future. Recognizing this contradiction, Diodorus relied on the plausibility of the first two propositions to establish that 'nothing is possible that neither is nor ever will be the case'" (*Discourses*, 2.19.1).

the attempt to formalise Diodorus' thought, Mates would help himself freely to expressions like '*p* at time *t*'. In the first chapter of *Past, Present, and Future*, dedicated to the precursors of tense-logic, A. N. Prior describes how Mates' attempt encouraged him to try, alternatively, to write *Fp* for "It will be the case that *p*", by analogy with the usual modal construction $\Diamond p$ ("It is possible that *p*"). The analogy raised, unavoidably, a problem for a logic which would treat tenses and temporal adverbs, by analogy with the formal regimentation of modalities, as operators whose operands are temporally (as in the modal case, modally) neutral propositions. The plurality of systems of modal logic prompted the unavoidable question: to which of these systems correspond the Diodorean definitions? This question would be the clue to the development, in the ensuing decade, of tense logic (cf. Prior 1967a: 20-31).

The privilege accorded to temporal propositions, and the interest in the study of their logical properties, is equally manifest in Stoic logic.[5] As Hintikka remarks: "Virtually all the examples of singular sentences that were used by the Stoics as examples and are preserved to us seem to be temporally indefinite. What is more important, such temporally indefinite sentences are put forward by the Stoics as examples of sentences that are taken to express a complete *λεκτόν* (*lekton*). These complete assertoric *lekta* or in short *ἀξιώματα* (*axiomata*) of the Stoics are in many respects reminiscent of the 'propositions' that many modern philosophers postulate as meanings of eternal assertoric sentences. However, *axiomata* differ from propositions in that they are temporally indefinite in the same way as occasion sentences. By saying 'writes' one does not express a complete *lekton*, we are told by the Stoics, because 'we want to know *who* [writes]'. Nevertheless, a sentence like 'Dion is walking' is said to express a complete *lekton*, in spite of the fact that it leaves room for the analogous question: '*When* is it that Dion is walking?'" (Hintikka 1973: 70-71).

The Stoics were thus able to talk freely about changes in the truth-value of a *lekton*. It comes as no surprise that, in the first attempt at a systematic reconstruction of Stoic logic by a modern

5 Cf. Mates (1953).

logician, Benson Mates called *lekta* "propositional functions with a temporal variable" (Mates 1953: 132), thus misrepresenting, *volens nolens*, the material he was expounding. For, if "Dion is walking" is a propositional function taking as arguments instants or time-intervals, then the proposition which, each time, the utterance of such a sentence expresses is an *atemporally* true or false proposition—the expression of an eternal truth (or falsehood), not the transient truth (or falsehood) posited by the Stoics.

The distortion, and the anachronism, were exposed by Geach in his review of Mates' book. "The Stoics neither had a pair of terms answering to the Peano-Russell distinction between a proposition and a propositional function, nor gave any example that could suitably be translated by an expression like 'Socrates dies at *t*'" (Geach 1955: 144). Introducing that distinction would jeopardise the examples of Stoic propositional logic which reached us. For the Stoics held that

> (1) If Dion is alive, then Dion is breathing; but Dion is alive; therefore, Dion is breathing

is of the form "if *p*, then *q*; but *p*; therefore *q*'". But this form is not to be found in

> (2) For any *t*, if Dion is alive in *t*, then Dion is breathing in *t*; but Dion is alive *now*; therefore, Dion is breathing *now*.

Hence Geach's question: "May not the Stoics well have thought that, though the truth-value of 'Dion is alive' changes at Dion's death, the sentence *still expresses the same complete meaning* (*lekton*)?" (Geach 1955: 144). As we saw, the answer to this question is affirmative.

Like Aristotle and the Stoics, Medieval logicians had no trouble in admitting transient truths, expressed by temporal propositions, and they probed systematically the logic of such propositions. In his brief overview of conceptions about time and truth in the history of logic, Prior (1957: 104) holds that the two main theses of Medieval logic concerning that were:

> (i) tense distinctions are a proper subject of logical reflection;
>
> (ii) what is true at one time is in many cases false at another time, and vice versa.

It comes as no surprise that Geach, credited by Prior with having called his attention to the relations between time and truth in Ancient and Medieval logic, should have expressed, in a review of Julius Weinberg's book about Nicolaus of Autrecourt, the same criticism he would address, some years later, to Mates about Stoic logic: "Such expressions as 'at *t*' (pp. 168, 172) are out of place in expounding scholastic views of time and motion. For a scholastic, *Socrates is sitting* is a complete proposition, *enuntiabile*, which is sometimes true, sometimes false; *not* an incomplete expression requiring a further phrase like 'at time *t*' to make it into an assertion" (Geach 1949: 244).

The idea of a temporal proposition—a proposition whose truth-value changes with time—is so natural in Ancient and Medieval philosophy that, from a historical point of view, what seems to require explanation is rather the emergence, as late as in the 17th century, in the work of Leibniz and other forerunners of mathematical logic, of an *atemporal* logic, and of the related idea that every proposition is atemporally (eternally) true or false.[6]

Frege is, as usual, exemplary in his effort to make explicit the assumptions underlying the understanding of logic which receives, in his work, its first systematic formulation. For Frege, a proposition cannot be true at a time and false at another: a temporal proposition—if, *per impossibile*, there were such a thing—could not be coherently assessable as true or false.[7] Missing a temporal indication, "It is sunny in Buenos Aires" could only be "true on" certain occasions and "false on" others. But what that means is that such a "proposition" is actually *incomplete*: "A thought is not true at one time and false at another, but it is either true or false, *tertium non datur*. The false appearance that a thought can be true at one time and

6 I will not risk an explanation, except to remark that the extrusion of time from logic chimes with the mathematisation of the scientific worldview: the laws that the new science of Galileo, Descartes and Newton seeks are eternal propositions expressible in mathematical equations.

7 This remark contains, *in nuce*, the gist of what would be Gareth Evans's criticism to the temporal calculi introduced by Prior (cf. Evans 1985), on which more below.

false at another time arises from an incomplete expression. A complete proposition or expression of a thought must also contain the time datum".[8]

Likewise, in the first of his *Logical Investigations*, Frege writes: "The present tense is used in two ways: first, in order to indicate a time; second, in order to eliminate any temporal restriction, where timelessness or eternity is part of the thought—consider for instance the laws of mathematics. Which of the two cases occurs is not expressed but must be divined (*erraten*). If a time-indication is conveyed by the present tense one must know when the sentence was uttered in order to grasp the thought correctly. Therefore the time of utterance is part of the expression of the thought" (Frege 1918: 331-2).

These passages by Frege articulate the orthodoxy which would prevail throughout most of the subsequent history of logic and analytic philosophy: what appears at the level of "surface grammar" as temporal indeterminacy or neutrality should be understood as expressive incompleteness, namely, the kind of incompleteness that characterises a propositional function as opposed to a complete proposition. According to this tradition, opposed by very few dissenters (McTaggart, Findlay, Geach, Prior, Kaplan), "It is raining" is, on every occasion of utterance, equivalent to "It is raining now". Hence, far from being the expression of a proposition which, its identity preserved, has different truth-values in the course of time, that sentence is, on each occasion of utterance, the expression of an eternal proposition: that which predicates, *of the time at which it is uttered*, that is satisfies (atemporally) the function *it is raining at t*.

As I remarked, McTaggart was one of the dissenting voices in that tradition. Indeed, a corollary of the First Part of his proof of the unreality of time (the sub-argument aimed at showing that there can only be change if the *A* series belongs to the objective reality of time) is that there can only be change if there are irreducibly temporal propositions. McTaggart's argument is that

8 Frege (1912: 338; Gareth Evans's translation in Evans 1985: 230).

change cannot consist, *contra* Russell, in the facts that an object *a*, which is *F* at *t*, is *non-F* at *t**, for both these are eternal facts—as Russell explicitly acknowledged.[9] Change can only consist in the fact that *a*'s *F*-ness, from having been future (at a time when *a* was not yet *F*), becomes successively present and past (past, that is, in times at which *a*, if it still exists, is, as the case may be, *F* or *non-F*): "It follows from what we have said that there can be no change unless some propositions are sometimes true and sometimes false. This is the case of propositions which deal with the place if anything in the *A* series—'The battle of Waterloo is in the past', 'It is now raining'. But it is not the case with any other propositions" (McTaggart 1927: 15).[10]

The idea of a tense calculus, such as it was to be developed by Prior in the years 1950s, was outlined by J.N. Findlay in the paper "Time: A treatment of some puzzles" (1941). There, Findlay held that "our conventions with regard to tenses are so well worked out that we have practically the materials in them for a formal calculus", and that "the calculus of tenses should have been in-

9 See Russell (1903: 471).

10 McTaggart's second example is, as we have seen, infelicitous: "It is now raining" is *not* a temporal proposition. But recognition of the non-equivalence of "It is raining" and "It is now raining", which evaded both Prior in "On Spurious Egocentricity" (Prior 1967b) and Gareth Evans in "Does tense logic test upon a mistake?" (Evans 1985), was slow to emerge: it was Hans Kamp's merit to first articulate it, while still a graduate student, in material distributed in 1967 to an UCLA seminar ("The treatment of 'now' as a 1-place sentential operator", mimeo), later incorporated in the paper "Formal properties of 'now'" (Kamp 1971). Prior adopted the distinction, giving due credit to Kamp, in "Now" (Prior 1968). Kaplan presented a new and important argument for the non-equivalence in a footnote to "Demonstratives" to which I will come back (cf. Kaplan 1989a: 503 fn. 28). This is how Nathan Salmon describes, in retrospect, Kamp's seminal contribution: "It has become well known since the middle of the 1970s that the phenomenon of tense cannot be fully assimilated to temporal indexicality, and that the presence of indexical temporal operators necessitates 'double indexing', i. e., relativisation of the extensions of expressions—the reference of a singular term, the truth-value of a sentence, the class of application of a predicate (or better, the semantic characteristic function of a predicate) etc.—to utterance times independently of the relativisation to times already required by the presence of tense or other temporal operators" (Salmon 1989: 356).

cluded in the modern development of modal logics". A tense calculus would include such "obvious" propositions as that

(3) *x* present = (*x* present) present

(4) *x* future = (*x* future) present = (*x* present) future

and also such comparatively recondite propositions as that

(5) (*x*).(*x* past) future

i.e., "all events, past, present and future, will be past" (Findlay 1941: 233).[11]

In Findlay's notation, the variables take events as arguments, and adverbs of time are regimented as predicates of events. That amounts to an extrusion of tenses, which get expressed, exclusively, by the predicates '*pastness*', '*presentness*' and '*futurity*', and their combinations. Findlay is here following McTaggart: in his notation, as in McTaggart's, "Queen Anne died" is rendered as "The death of Queen Anne is past".[12] In order to recover the usual mode of expression, one must extract from such nominalised phrases, which are the designators of events ("the death of Queen Anne", "the discovery of America", "the fall of the Berlin Wall"), the predications out of which they were fashioned ("Queen Anne dies", "America is discovered", "The Berlin Wall falls"); but then we can only express temporality through *tensing* or *adverbially modifying* the predication. The idea that the two modes of expression are equivalent underlies the tense calculi developed by Prior starting in the 1950s.

Prior himself explains this idea in "Changes in Events and Changes in Things": "Turning now to our main subject, I want to suggest that putting a verb into the past or future tense is exactly the same sort of thing as adding an adverb to the sentence. 'I *was* having my breakfast' is related to 'I am having my breakfast' in exactly the same way as 'I am *allegedly* having my breakfast is

11 A remark by Prior is apposite here: "The last law is unfortunately symbolised; the formula suggests that everything will have been the case (even permanent falsehoods); but it is easily enough amended to '((*x* present) or (*x* past) or (*x* future)) → (*x* past) future'" (Prior 1967a: 9).

12 Notice how the atemporal copula invites, all but unavoidably, the question "*When* is Queen Anne's death past?", which lies at the root of McTaggart's Paradox (cf. McTaggart 1927: 21).

related to it, and it is only an historical accident that we generally form the past tense by modifying the present tense, e.g. by changing 'am' to 'was', rather than by tacking an adverb" (Prior 1962: 13).[13]

Adverbs of time are thus regimented, in analogy with the modal operators, as sentential operators whose operands are *temporally neutral sentences*. What that means is that, in a tense-calculus, the operands of temporal operators are *temporal propositions* (propositions with variable truth-value); so 'F(a is G)' says that *a will be G*.

That reading would seem not to be mandatory. After all, it is *possible* to construe a sentence like "I am speaking" as the incomplete expression of an atemporally true (or false) proposition; and that's precisely what Frege (as we saw) recommended. But that alternative raises a problem for the interpretation of temporal operators.

Suppose I utter the sentence "I am speaking". As we have seen, eternalists and temporalists diverge about the proposition which that sentence, as uttered by me in the present context, expresses. According to eternalists, that is the proposition <*Paulo is speaking at 6 PM on November 07, 2017*>. According to temporalists, that is the proposition <*Paulo is speaking*>. If we embrace the eternalist reading, we are positing the presence of a hidden indexical in the sentence: we are assuming, in other words, that "I am speaking" = "I am speaking now".

The trouble with that reading was exposed, with surgical precision, by Kaplan: it makes temporal adverbial modification in the sentence vacuous: "Technically, we must note that intensional operators must, if they are not to be vacuous, operate on contexts which are neutral with respect to the feature of circumstance the operator is interested in. Thus, for example, if we take the content of *S* ["I am writing"] to be (i) [*David Kaplan is writing at 10 AM on 3/26/77*], the application of a temporal operator to such a content would have no effect; the operator would be vacuous" (Kaplan 1989a: 503-4).

13 See also the summary presentation of the project in Prior (2004).

If, following Prior, we render "I was writing" as "This was the case: I am writing", and, in that construction, we read "I am writing" as the expression of a temporal proposition, we catch the idea that the pastness of an event is its past presentness (that being past is *having been* present); and we do that because we make it possible that the operator 'This was the case' (in Prior's notation, 'P') modifies the operand proposition.[14] If, on the other hand, we embrace the eternalist reading, we must read "This was the case: I am speaking" as "This was the case: I am speaking now", and in that reading the operator is vacuous: no modification is being introduced in the content on which it is supposed to operate.

The non-equivalence of "It is sunny" and "It is sunny now" rests on the fact that *being sunny* is a temporary property while *being sunny now* is a permanent property. So McTaggart's and Kaplan's arguments converge.

What is at stake is, as McTaggart realised, the very possibility of change.

Look at it this way. If the *sentence* "Paulo is in Buenos Aires" expresses a true proposition when it is uttered on November 7, 2017, even if the *same* sentence, uttered on November 7, 2016, would express a false proposition, how are we to represent the truth of the proposition we express as we utter that sentence *today*? Three alternatives are available:

(6) "Paulo is in Buenos Aires" is true on November 7, 2017. [Aristotle, the Stoics, the Scholastics.]

(7) "Paulo is in Buenos Aires <u>on November 7, 2017"</u> is true. [Frege, Russell.]

---

14 "English speakers find it hard to see these things quite clearly; for in English sentences the point of view of the speaker dominates even subordinate clauses. When an English speaker, for example, wants to say on Tuesday that someone complained on Monday of a sickness that he had that day, the correct form of words will be 'He said he *was* sick', although the man was in fact complaining not of then-past but of a then-present sickness, and his words would have been 'I *am* sick' (...) on the few occasions on which we use phrases like 'It was the case that', in English, they are not followed by the present but the past; we say 'It was the case that he *was* sick', not 'It was the case that he *is* sick', thus hiding from ourselves the fact that it is the past presentness of his being ill, not its past pastness, to which we are alluding" (Prior 1967a: 14).

(8) "Paulo-on-November-7, 2017 is in Buenos Aires" is true.

[Quine, David Lewis.]

In (7) and (8) the copula has been detensed, and what Frege called a 'time datum' has been introduced, respectively, in the predicate (7) and in the subject (8). The result was to turn a temporary property into a permanent one: while (6) represents a temporal proposition, (7) and (8) are representations of timelessly true (eternal) propositions.

In "Does Tense Logic Rest upon a Mistake?", Gareth Evans claims that tense logic awaits a "semantic foundation" in that the concept of a proposition whose truth-value changes with time has not been made clear by its practitioners. Evans examines three interpretations of 'proposition whose truth value changes with time' and finds them all wanting. It is remarkable, though, that in two of these interpretations the operands of temporal operators are *not* temporal propositions. In the system Evans calls T_2 they are propositional functions; and then in T_3 they are eternal propositions. The only rendering of 'truth at a time' which gives us temporal propositions is one on which no stable evaluation is possible. According to the T_1 rendering, "It is raining" is true if it is raining at the time of evaluation. "On this interpretation, a proper appreciation of the semantic functioning of tense requires us to abandon the idea that particular historical utterances of tensed sentences are assessable, once and for all, as correct or incorrect. Rather, we must acknowledge that the evaluation of particular utterances must change as the world changes. Suppose *x* uttered 'It is raining' at *t*, when it is raining. Then, at *t*, his utterance was correct, but now it has ceased to rain, it has become incorrect" (Evans 1985: 327).

Given that uncomfortable result, Evans argues that the only way to bar retroactive reassessment, thereby ensuring stable evaluations, is to introduce a time reference (like 'now'), which amounts to turning the proposition into the expression of an eternal truth.

In the same vein, Mark Richard argued that temporalism is unable to provide a coherent account of content preservation. Consider the following sequence:

(9) a. Mary believed that Nixon was president.
b. Mary still believes everything she once believed.
c. Mary believes that Nixon is president.

The temporalist is allegedly committed to treat that as a valid argument. She holds that the object of Mary's belief in (9a) is the time-neutral proposition <*Nixon is president*>. If she retains her belief in that proposition (if she still believes now what she once believed), then she believes now that Nixon is president.

The eternalist has no such problem. For, on the eternalist reading, the object of Mary's belief in the past was the proposition <*Nixon is president now*>. If she retains her belief in that proposition (if she still believes now what she once believed), then she believes now that Nixon was president *then*.

As with Evans's argument, the conclusion is that the only way to ensure content preservation is to let the time information needed for truth-evaluation be a constituent of the proposition.

The temporalist reply to Richard's argument challenges his account of belief retention. The gist of the reply can be found in Frege's remark that "If someone wants to say today what he expressed yesterday using the word 'today', he will replace this word with 'yesterday'. Although the thought is the same its verbal expression must be different in order that the change of sense which would otherwise be effected by the differing times may be cancelled out" (Frege 1918: 332). Frege is thinking, of course, of eternal contents, but the remark makes a general point which is also available to the temporalist: *cognitive decentring is the key to content preservation.* In general, *to retain a belief is not to believe the same time-neutral proposition, but another proposition appropriately related to the original one.*

Richard has considered that reply and raised a problem for it. Keeping in mind that Lyndon Johnson was President of the USA from 1963 to 1969, that he acceded to the Presidency upon John F. Kennedy's assassination in 1963, that he was re-elected in 1964 and that Herbert Humphrey was the democratic candidate in the 1968 elections, consider the following scenario. In 1966 Mary had the belief she expressed using

(10) It will be the case that Johnson is re-elected.

By 1968, in view of the results of the Democratic National Convention, Mary gave up that belief. However, Mary may well have had, early in 1964, a belief she would correctly express using (10). Now, she may have retained that (true) belief while not retaining the (false) belief she expressed in 1966 using (10).

On the face of it, temporalism is unable to account for Mary's change of mind. For (10), on the temporalist account, would express the same proposition in 1964 as it did in 1966. Thus, Mary believes the past tense counterpart of the proposition expressed by (10) in 1964 if she believes the past tense counterpart of the proposition expressed by (10) in 1966. We are led to the conclusion that Mary retains her belief from 1964 if and only if she retains her belief from 1966. The eternalist has no such problem, since on her account the (eternal) proposition expressed by (10) in 1964 is not the same as that expressed by (10) in 1966.

This criticism, however, misses the fact that tense operators in natural language are not restricted to temporal prefixes such as 'it was the case that' or 'it will be the case that'. As Berit Brogaard remarks: "Basic tense operators, such as 'it was the case that', 'it will be the case that', can combine with time adverbials, such as 'yesterday', 'now', 'two weeks ago', 'in 1981', 'during World War II', 'during Bush's first term', 'when my students handed in their papers', and so on. Thus, the tense operators of English, if such there are, include 'it was the case during World War II that', 'it was the case last Friday that', 'it will be the case when my students have handed their papers that', and so on" (Brogaard 2012: 48).

This fact about the grammar of tense operators is captured by Prior's metric tense logics. Let n be a unit of time measurement, say one year. Then:

(11) $P_n q$ = q n years ago

So:

(12) P_{106} (Alan Turing is born) = Alan Turing was born 106 years ago.

(13) F_{12} (Paulo retires) = In 12 years from now Paulo will retire.

And that disposes of Richard's objection.

Or so it seems. For it will work only to the extent that predicative content is kept invariant through decentring. And that's how the dual of temporalism, anti-individualism, will pose a new threat to content preservation, as we will see in the next lecture.

2. Preservation

Recalled from the shades to be a seeing being,
From absence to be on display,
Without a name or history I wake
Between my body and the day.

W. H. Auden, *Horae Canonicae*,"1. Prime" (1949)

The sense of danger must not disappear:
The way is certainly both short and steep,
However gradual it looks from here;
Look if you like, but you will have to leap.
...
A solitude ten thousand fathoms deep
Sustains the bed on which we lie, my dear:
Although I love you, you will have to leap;
Our dream of safety has to disappear.

W. H. Auden, "Leap Before You Look" (1940)

This lecture is an exercise in the epistemology of reasoning. Its topic is a *prima facie* improbable one: the role of luck in the recognition of deductive validity. Let me begin by explaining what I have in mind.

It is one thing to know that a pattern of inference, say *modus ponens*, is truth-preserving; it is quite another to recognise a particular inference as valid. When I speak here of 'recognition of deductive validity' what I have in mind is the latter: it is knowledge of the logical properties of particular inferences that is my concern here. I don't mean the kind of reflective knowledge that a logically sophisticated reasoner would express by making use of such terms of art as 'valid inference' or 'logical consequence'. I mean the ability to correctly take a set of beliefs as providing sufficient reason for a further belief, or a change of mind.

I take my lead from David Kaplan, who writes in the "Afterthoughts" to *Demonstratives:* "Logic and semantics are concerned

not with the vagaries of actions but with the verities of meanings" (Kaplan 1989b: 584-5). In the section of that work to which this sentence belongs, Kaplan is explaining why his investigation of the semantics of demonstratives and other indexicals concentrates on *occurrences of expressions in a context* rather than on the pragmatic notion of an *utterance of an expression by an agent in a context*. He writes: "Utterances take time, and are produced one at a time: this will not do for the analysis of validity. By the time an agent finished uttering a very, very long true premise and began uttering the conclusion, the premise may have gone false. Thus, even the most trivial inference, *P* therefore *P*, may appear invalid" (Kaplan 1989b: 584). For the purposes of logical regimentation, such accidents of context-dependency will be circumvented by referencing, say, all occurrences of 'now' to a single instant, all occurrences of 'here' to a single place, and so on. Likewise, all occurrences of the same demonstrative, say 'this', will be referenced to a single object by assuming (what Kaplan calls) a single *directing intention*. We are, that is, invited to think of two different occurrences of 'this' as driven by a single intention aiming twice at the same object, an intention we would presumably express through such behaviour as pointing, staring fixedly at the object, and so son. That should do the trick as far as logic is concerned. "But", asks Kaplan, "does it leave our logic vulnerable to a charge of misrepresentation? What is it that we hope to learn from such a logic? [...] To assume that one intention can drive two occurrences of a demonstrative seems to me more falsification than idealisation" (Kaplan 1989b: 590).

The distinction between the vagaries of actions and the verities of meanings is supposed to help with these perplexities; yet, Kaplan remarks, at the close of his discussion: "There is something I'm not understanding here, and it may be something very fundamental about the subject matter of logic" (*Ibid.*).

I will eventually submit a conjecture as to what it is that Kaplan may not be understanding here; also, whether it is indeed something fundamental about the subject matter of logic. But first things first.

My main question is: can we know by reflection alone whether an inference is valid? And then, however the main question is

answered, can rightness of reasoning be, if on occasion, a matter of luck—of one's being, as it were, in the right place at the right time? Can it be that whether we succeed in inferring correctly depends on circumstances beyond our ken? And if so, and then turning full circle, how can we know by reflection alone, if indeed we do, whether an inference is valid?

I will answer affirmatively, though not unqualifiedly, to each of the first three questions. That is, I will argue that to a *very large* extent it is by reflection alone that we know whether an inference is valid. But I will also argue that rightness of reasoning can be a matter of luck; that whether we succeed in inferring correctly may well depend on circumstances evading our knowledge; *and* that this is no hindrance to our being, more often than not, able to know by reflection alone whether an inference is valid.

That will leave us with a picture on which *the contingencies of external individuation of thought contents* (no matter how widespread the phenomenon should turn out to be; no matter, in other words, whether anti-individualism is true) *will not put in jeopardy reflection as a mode of access to logical form.* But it is also part of the picture that *reflection may not be enough: that we may sometimes have to count on what John McDowell calls "a courtesy by the world".*

The main question (that of what Paul Boghossian described as "the apriority of our logical abilities" (1992), has been usefully—if, to my mind, unsatisfyingly—discussed in the framework of the still ongoing debate about the compatibility of anti-individualism and first-person authority. It is easy to see why: if many, perhaps most, representational mental states and events depend for being the specific states and events that they are on non-representational relations between the individual and a wider environment, then even fully conceptualised contents (those we ascribe in oblique, *de dicto*, clauses) would seem to fail the intuitively plausible requirement of *transparency*, thus defined:

> (14) The content *c* of a propositional attitude *a* of *S* is transparent to *S* if for every content *c** *S* is in a position to tell by reflection alone whether $c = c^*$.

So construed, transparency requires both identity and difference of contents to be accessible to reflection. The external indi-

viduation of content would seem to jeopardise that access, paving the way for failure to recognise either identity or difference of content. The predicament of Kripke's Pierre is an instance of the former failure; that of the slow-switched denizens of Twin Earth, like Boghossian's Peter, an instance of the latter.

The threat to transparency of validity is now manifest. Here is the scheme of a putatively valid reasoning:

(15) a. Fa
b. Ga
c. $\exists x (Fx \wedge Gx)$

Now add subscripts to your non-logical constants, if only to mark that they are different tokens of the same type, hence not *logically* guaranteed to be co-referential (we are supposed to be dealing here with some real-life inference, not with logical idealisation):

(16) a. Fa_1
b. Ga_2
c. $\exists x (Fx \wedge Gx)$

And there we are: in *Environment 1*, $d(\text{`}a_1\text{'}) = d(\text{`}a_2\text{'})$, and the inference is valid. Switch to *Environment 2*, where $d(\text{`}a_1\text{'}) \neq d(\text{`}a_2\text{'})$, and the inference is a fallacy of equivocation. And now suppose you either raise in Environment 1 the possibility that you are in Environment 2, and accordingly refrain from drawing the conclusion, or else you go through the steps (a) to (c), taking yourself to be in Environment 1 when, in fact, you are in Environment 2. Either way, your reasoning abilities are crippled.

The second alternative (you take yourself to be in Environment 1 when, in fact, you are in Environment 2) is, of course, the template exploited in the "slow switching" thought experiments, and in a good deal of the extant literature about recognition of validity. The first one (you raise in Environment 1 the *possibility* that you are in Environment 2, and refrain from drawing the conclusion) will resurface nearing the end of my discussion.

Now, I want nothing with slow switching and what Boghossian (1994) called "the semantics of travel". I think these scenarios of metaphysical victimisation (you go to bed at home and, while sound asleep, are stealthily taken overnight from Earth to

Twin Earth without ever realising it) have had a corrupting influence on the minds of those who have written on knowledge of validity. They have, in particular, made all but invisible the ways in which, in real life as opposed to philosophical fiction, the vagaries of actions are apt to affect the verities of meanings.

Look at the matter this way: given the task of assessing the inferential behaviour of someone who performs the inference (a)-(c) in Environment 2, wholly unaware of being in such an unfriendly environment, you are faced with a trade-off between what Joseph Camp calls "doxastic" and "inferential charity" (Camp 2002: 38-39). That is, you will have to choose between maximising the subject's set of true beliefs at the expense of her rationality or else maximising her rationality at the expense of her set of true beliefs. Other things being equal, faced with someone who performs the inference (a)-(c) in Environment 2, you go for inferential charity and look for a (possibly tacit) false premise.

Why is that so? Because you are aware of the unavailability, for the subject whose reasoning abilities you are assessing, of the relevant information about her environment. She has just done her best given the available information.

And that's what I find so unsatisfying about slow switching scenarios and the way the debate about "the apriority of our logical abilities" has been carried on: the switched subject is *stipulated* not to have cognitive access to the shifts in environment which are apt to affect the validity of her inferences. Here is Tyler Burge, and bear with me to have him quoted at some length: "Suppose that one underwent a series of switches back and forth between actual Earth and actual Twin Earth so that one remained in each situation long enough to acquire concepts and perceptions appropriate to that situation. Suppose occasions where one is definitely thinking one thought, and other occasions where one is definitely thinking its twin. *Suppose also that the switches are carried out so that one is not aware that a switch is occurring. The continuity of one's life is not obviously disrupted.* So, for example, one goes to sleep one night at home and wakes up in twin home in twin bed—and so on. (Your standard California fantasy.) Now suppose that, af-

ter decades of such switches, one is told about them and asked to identify when the switches take place. The idea is that one could not, by making comparisons, pick out the twin periods from the 'home' periods" (Burge 1988: 58; *my emphasis*).

I emphasise the last sentence: for it brings to the fore the limits of the canonical strategy, as put forward by Tyler Burge, Sidney Shoemaker, or Donald Davidson, to reconcile anti-individualism and first-person authority.

According to these authors, the propositional content of the "first- order" thought *p* is *integrally* absorbed in the self-ascription (the "second-order" thought) *I think that p —whatever* that first-order content may be.[15] Courtesy of that integral absorption, such second-order thoughts ("*cogito*-like thoughts", as Burge calls them) enjoy the property of *self-verification*: my *thinking that I am thinking that p* makes it the case, *eo ipso*, that *I am thinking that p*.

Such absorption of the "first-order" thought (*this glass is full of water*) into the self-ascription (the "second-order" thought *I am thinking that this glass is full of water*) is a *constitutive* relation, presumably immune to the vagaries of mental causation (*contra* Heil 1988). As both Burge, Davidson and Shoemaker variously stress, first-person authority is *normatively* grounded—in particular, it is grounded in the constitutive role of self-knowledge in rationality.[16] The key idea here is that ascribing rationality to an agent amounts necessarily to ascribing her a *prima facie* privileged access to the contents of her own propositional attitudes: the capacity to critically assess one's own judgments is a constitutive feature of rationality; now the exercise of that capacity requires, of a subject who thinks of an object whatever, that she knows it is *that* object, and not another thing, she is thinking of.[17]

---

15 Cf. Davidson (1984, 1987), Burge (1988), Shoemaker (1988), Heil (1988).

16 Cf. in particular (in addition to the papers cited in the preceding note) Burge (1996).

17 Since ascription of privileged access is, as I stressed, *prima facie* defeasible under a variety of circumstances loosely describable as lapses of rationality, Burge speaks of an *entitlement* to self-knowledge (see, in particular, Burge 1996). The vicissitudes of rationality (self-deception, dissociation, weakness of the will) constitute one of the main subjects of the essays gathered together in Davidson's posthumously published *Problems of Rationality* (Davidson 2004).

Hence, knowledge of one's own propositional contents (first-person authority; "basic self-knowledge", in Burge's phrase) is not grounded in anything like a *reliable process*, like those which on many contemporary approaches to cognition account for the crucial difference between knowledge and (mere) true belief.[18] As Kevin Falvey puts it, should first- person authority have to be explained by some reliable process courtesy of which, say, the first-order thought *p* would *cause* the second-order thought *I think that p,* the propositional attitudes we typically submit to assessment when engaging in critical reasoning "would be treated as *objects* of investigation, in such a way that the point of view of the reflective, critical reviewer and the point of view of the reviewed attitudes would be no more unified than that of one person and another. This is inconsistent with the rational *immediacy* with which it follows that I must change my first-order attitude when it becomes apparent to me that it is not adequately supported" (Falvey 2003: 234). A rational subject, in other words, cannot coherently be treated as a *spectator* of her own propositional attitudes and contents.

Yet ascription of authority is, to begin with, *prima facie* ascription, subject to revision on the face of defeating evidence —as is the case, generally, with all ascriptions of rational capacities and attitudes. Then, and even more crucially, the constitutive relation alleged by Burge, Davidson, Shoemaker *et alii* to hold between the "first-order" thought and the corresponding self-ascription is restricted (this is an essential feature of every "*cogito*-like thought") to *present* tense —more specifically, to thoughts characteristically expressed by sentences in the *first* person *present* tense *indicative* mood in its *assertoric* use. Only in this privileged case, a judgment of the form "S thinks that *p*" is a sufficient con-

18 I have in mind the "reliabilist" analysis of knowledge, introduced by Frank Ramsey in 1929, substantially refined by Alvin Goldman and others from the mid-seventies on, and embraced by many among the most influential contemporary epistemologists. According to that analysis, knowledge is true belief *brought about by a reliable belief-forming process*—typically, by the exercise in normal conditions of such cognitive capacities as perception, memory or inference; and the acceptance of trustworthy testimony.

dition of its own truth. In *all* other cases, the difference between perspectives (personal, temporal, modal) makes for a wide range of possibilities of misascription.[19]

In other words, the solution required a complete coincidence (same agent, same time, same possible world) between the perspective from which a subject would think the first-order thought *this glass is full of water* and that from which the subject would self-ascribe that first-order thought, thinking *I (hereby) am thinking that this glass is full of water.* In all other cases, the Cartesian property of self-verification is conspicuously absent.

Burge couldn't be more explicit about that: "Of course, the person may learn about the switches and ask 'Was I thinking yesterday about water or twater?'—and yet not know the answer" (Burge 1996: 64).

That admission sets the stage for Boghossian's Memory Argument. As he writes in "Content and Self-Knowledge":

> These remarks strike me as puzzling. They amount to saying that, although S will not know tomorrow what he is thinking now, he does know right now what he is thinking right now. For any given moment in the present, say t_1, S is in a position to think a self-verifying judgment about what he is thinking at t_1. By Burge's criteria, therefore, he counts as having direct and authoritative knowledge at t_1 of what he is thinking at that time. But it is quite clear that tomorrow he won't know what he thought at t_1. No self-verifying judgment concerning his thought at t_1 will be available to him then. Nor, it is perfectly clear, can he know by other non-inferential means. To know what he thought at t_1 he must discover what environment he was in at that time and how long he had been there. But there is a mystery here. For the following would appear to be a platitude about memory and

19 As Davidson writes: "Neither speaker nor hearer knows in a special or mysterious way what the speaker's words mean; and both can be wrong. But there is a difference. The speaker, after bending whatever knowledge and craft he can to the task of saying what his words mean, cannot improve on the following sort of statement: 'My utterance of *Wagner died happy* is true if and only if Wagner died happy'. An interpreter has no reason to assume this will be *his* best way of stating the truth conditions of the speaker's utterance" (1984: 13).

> knowledge: if S knows that p at t_1, and if (at some time later) t_2, S remembers everything S knew at t_1, then S knows that p at t_2. Now, let us ask: *why* does S not know today whether yesterday's thought was a *water* thought or a *twater* thought? The platitude insists that there are only two possible explanations: either S has forgotten or he *never* knew. But surely memory failure is not to the point. In discussing the epistemology of relationally individuated content, we ought to be able to exclude memory failure by stipulation. It is not as if thoughts with widely individuated contents might be easily known but difficult to remember. The only explanation, I venture to suggest, for why S will not know tomorrow what he is said to know today, is not that he has forgotten but that he never knew (1989: 157-8).

Here is the argument, as reconstructed by Ludlow (1995):

(17) a. If S forgets nothing, then what S knows at t_1, S knows at t_2.
b. S forgot nothing.
c. S does not know that *p* at t_2.
Therefore,
d. S did not know that *p* at t_1.

In other words, Burgean "basic self-knowledge" is not self-knowledge enough. As Boghossian writes elsewhere: "[...] the assurance that this sort of proposal provides about the compatibility of externalism with authoritative self-knowledge is [...] *hollow*: it carries with it none of the usual *consequences* of first-person authority about thought content" (1992: 15). Self-knowledge should be a cognitive achievement, and that is not the case with Burge's "basic self-knowledge", which comes for free and is so thin as not to discriminate between water thoughts and twater thoughts.

Nor is this all. For, as we should expect, slow switching will have a bearing on "any reasoning that takes place over time, hence any reasoning" (Burge 1998: 98). For one thing, the subject's ability to assess rightness of inference would seem to be jeopardised by his unawareness that he's been switched.

That was the problem raised by Boghossian in "Externalism and Inference": "[anti-individualism] is inconsistent with the thesis that our thought contents are *epistemically transpar-*

ent to us [...] this is true in a sense that falsifies another important and traditionally held view—that we can detect *a priori* whether our inferences are logically valid or not" (Boghossian 1992: 13).

And, to be sure, that would be very disturbing indeed. After all, the main interest (and, just possibly, the main promise) of anti-individualism lies in its acknowledgement of the impact of exposure to changing contexts on the constitution of thought contents. That's precisely what explains the interest aroused, in the literature about anti-individualism, by cases of context-switch, often illustrated with such elaborate fantasies about space-travelling from Earth to Twin Earth and back, interplanetary abductions and like exercises in science fiction. At the end of the day, such exercises should have been just a device to graphically describe a range of much less extraordinary phenomena which, if anti-individualism is right, take place in a variety of situations prompted by exposure to differences between the contexts in which the rational capacities of a single subject must be exercised.

Anyway, that is the setting of Boghossian's problem about anti-individualism and inference (Boghossian 1992). Boghossian's argument has the form of a *reductio ad absurdum*: the truth of anti-individualism implies the possibility of undetectable errors in reasoning, due to unperceived shifts in propositional content. That possibility clashes with the transparency of mental content; therefore, anti-individualism is false.

Suppose that, having enjoyed a happy childhood on Earth, I am someday carried away to Twin Earth. And suppose further, as Burge asks us to, that "the switches are carried out so that one is not aware that a switch is occurring. The continuity of one's life is not obviously disrupted" (Burge 1988: 58). I just wake up on Twin Earth in twin bed and everything looks exactly as before (remember, that was built into the thought experiment). In due time, according to anti-individualistic common wisdom, my usage of the term 'water' comes to mean what the linguistic community to which I now belong uses it to mean: namely, *XYZ*. And here I am, inferring from the conjunction of true premises

(18) a. I enjoyed playing in water (= H_2O) when I was a kid. [from memory]

b. This glass is full of water (= *XYZ*). [from current perception]

the false conclusion

c. This glass is full of the same liquid I enjoyed playing in when I was a kid.

To make things worse, the fallacy I fall prey to is not comparable to the usual fallacy of equivocation, in which an ambiguity is neglected, and the reasoner is in principle in a position to detect and rectify, on a wholly *a priori* basis, the flaw in her reasoning. In the slow switching scenarios, there is just nothing the subject can do to prevent or fix the irrationality, short of undertaking an empirical investigation of the environment, and of her own personal history.

Still worse, given the conditions built into the slow switching stories, it is (to put it mildly) unclear what such an "empirical investigation" could possibly look like. It's not as if there might be traces, like a flight ticket from Earth to Twin Earth in the inside pocket of your jacket, or custom papers attached to your passport, or a message from home in the answering machine. The two worlds were stipulated to be indiscernible, exact duplicates (apart from the single "external" difference): were it not so, we would have learnt nothing from the fictions.

The perception that there's not much that the victims of slow switching can do to detect the external sources of their possible logical shortcomings—a perception which I find to be widespread (if mostly tacit) in the literature—answers for the surprising willingness, displayed by friends and foes of anti-individualism alike, to devise exculpating moves as a response to those scenarios of logical misfortune.

After all, the *only* difference between the lucky and the unlucky reasoners lies wholly beyond their ken. I suppose Boghossian speaks for most writers in the field when he writes: "It seems to me that there is an immediately recognisable sense in which there can be no difference *in respect of rationality* between *[the reasoner in Possible World 1 and that in Possible World 2]*. It seems im-

plausible in the extreme to say that they differ in their capacity to *reason"* (Boghossian 1992: 27).

That is of course an eminently plausible appraisal, given what the differences between the two contrasting worlds are supposed to be. Small wonder, then, that the choices on offer are (with the remarkable exceptions of Sorensen 1998, Williamson 2000 and Millikan 2009) a variety of exculpating moves, designed to shield the rationality of the possibly unlucky reasoner against the contingencies of context-shifting. (See, e.g. Schiffer 1992, Burge 1998, Ludlow 2004, Sosa 2005, Collins 2008, Recanati 2012).

The *individualistic* exculpating move will predictably recoil from the broader and unsafer landscape to an inner domain, a.k.a. narrow content, sealed off from the contingencies of external causation, accident, and luck. Here is how Boghossian introduces it: "If, then, it is also true that there is an important sense in which [the reasoner's] behaviour *makes sense from his point of view*, we would appear to have here an argument for the existence of a level of intentional description which conserves that sense" (Boghossian 1992: 28). The move bears comparison with Kant's forceful "shrinking" of the proper domain of moral assessment to the inner realm where a pure will operates by itself, sealed off alike from the vagaries of causation, contingency and luck.[20]

A bit more surprising are the *anti-individualistic* exculpations, paramount among which is the Schiffer-Burge "anaphoric" view of content preservation (see Schiffer 1992, Burge 1998).

The main idea is that the reiteration, in an occurrent episode of thinking, of the content of a *past* thought is made possible by a dependency relation comparable to that which holds between relative pronouns, and other anaphoric expressions, and their antecedents in the linguistic constructions in which they feature. In "Laura was confident that she would get the prize", the pronoun 'she' designates Laura: its semantic value is determined by the anaphoric antecedent which is the proper name—as the value

20 See Bernard Williams's perceptive discussion in "Moral Luck" (Williams 1976).

of a bound variable in first-order quantification is determined by the quantifier which is its anaphoric antecedent.[21] Analogously, in "Galileo said that the Earth moves; although that was true, it all but costed his life", the expressions 'that' and 'his' hark back, respectively, to the proposition <*The Earth moves*> and to Galileo (the man who risked his life by asserting that proposition); and the structure of such back-reference is, again, that of anaphora.

Hence, the unhappy reasoning in (18a-c) would get reinterpreted as:

(19) a. I enjoyed playing in water when I was a kid.
b. This glass is full of water↑.
c. This glass is full of the same liquid I enjoyed playing in when I was a kid.

Here I resort to 'water↑' to mark the anaphoric dependence of 'water', as tokened by the reasoner in premise (19b), on its occurrence in premise (19a). And what we have as a result is, small wonder, a valid argument with a false premise: as 'water' in (19a) denotes H_2O (the thought content here being supplied by preservative memory), premise (19b) amounts to the false judgment that the Twin Earth glass is full of H_2O.

Burge's thesis is that when I think *today* that sometime in the past I entertained a certain thought, the content of my self-ascription (the "second-order" thought) *I thought that p* stands to the content of the ("first-order") thought *p* in a dependency relation *analogous* to that which holds between a pronoun or relative clause and its anaphoric antecedent. According to Burge, then, when I remember on Twin Earth that as a child I enjoyed playing in water, I am thinking of *water* (H_2O), not of *twater* (*XYZ*), because in the very act of recollecting I *defer* the determination of that term's extension to the competent user *I myself* was in the past (as the user of a proper name defers the fixation of its referent to his predecessors in the historical chain described by Kripke). When now I connect that preserved propositional con-

21 The analysis of quantification as formal regimentation of relative clauses (and, in particular, of the bound variable as the formal counterpart of the relative pronoun) is a recurring theme in Quine's philosophy of logic: see, *e.g.*, Quine (1960: 135-7).

tent to my *occurrent* judgment about the content of the glass I have before me, in the context of the inference (17a-c), my intention is to reuse, in the second premise, the *same* concept which is a constituent of the first one. Quite generally, *within* an inferential chain, the conceptual contents which are constitutive of each inferential step usually stand in anaphoric dependency relations: a paraphrase substituting relative clauses for the separate premises is apt to make these relations manifest.[22]

By now it should be apparent that we are dealing with a way more problematic extension of the canonical compatibilist thesis. To begin with, Burge's program obviously stands or falls with the heuristic fecundity, so far very insufficiently probed, of the analogy between anaphora and preservative memory.[23] But there's trouble of another, possibly more serious, kind lurking around. Burge is adamant, in his reply to Boghossian as in other writings, that the kind of memory on which inference depends does not work through *discrimination* (as is the case with episodic memory) but through *preservation* of content. That's why his main target in Boghossian's argument is the tacit assumption that knowledge of one's own propositional contents should be *discriminative* knowledge.[24] Such a requirement, Burge claims, misses what is most distinctive about "first-person authority": the ability to critically appraise one's own judgments, which presupposes direct access to their contents.

But that's precisely what raises the difficulty I, for one, find myself in. After all, even granting to Burge, Davidson, Shoemaker *et alii* that first-person authority is grounded in normative prin-

---

22 That is the core of Stephen Schiffer's reply to Boghossian in Schiffer (1992).

23 Commenting on the analogy, Burge warns: "The analogy must be used with caution. I do not model preservative memory on pronominal back-reference. I believe that preservative memory is more basic (both ontogenetically and in explanations of epistemology and rationality) than anaphora in language. In fact, it seems to me that a linguistic theory of anaphora has to be able to account for anaphora supported by preservative memory" (Burge 1998: 93).

24 "In preserving knowledge, S (or S's memory) need not be in the third-person position of solving the problem of whether yesterday's knowledge had one content rather than another. That would be to take the past thought as an object of identification" (Burge 1998: 96).

ciples, rather than in some reliable process through which intentional contents would be identified and re-identified, the fact remains that content preservation *in memory* is itself dependent on causal processes whose reliability we are, in the best of possible worlds, entitled to take for granted (as is the case with that of any other kind of process) *ceteris paribus*. This is something Burge himself occasionally gestures at—to give a single example: "Given appropriate reliance upon preservative memory, and given the existence of causal memory chains going back to the states which carried intentional content, preservative memory takes up the 'antecedent' content automatically, without having to identify it" (Burge 1998: 94). The problem here is to devise a decent analysis—decent enough to supersede unspecified appeals to "normal conditions"—the circumstances in which reliance on preservative memory is appropriate; and, in particular, what is involved in "the existence of causal memory chains going back to the states which carried intentional content". At any rate, the crucial point implicitly granted by Burge is that content preservation in memory is vulnerable to the vicissitudes of its *empirical* bases.[25]

Then —and here we meet again, I'm afraid, what remains the hardest nut to crack in this whole story—, even in those cases where the conditions for "appropriate reliance" on preservative memory (whatever they may be) *are* satisfied, the fact remains that, like the "basic self-knowledge" which finds expression in "*cogito*-like thoughts", preservative memory as Burge describes it delivers propositional contents of which it is not improper to say that we preserve and retrieve them *whatever they are*—not very much unlike a blind man who would pick from a box one out of a set of marbles differing exclusively in their *color*. No wonder Burge and his allies make such a point of rejecting the imposi-

25 In particular, Burge explicitly acknowledges that after-the-fact discovery of switching may precipitate the *loss* of the ability to preserve past thoughts. The problem has an obvious counterpart for so-called 'collective memory' (on which a wee bit more below). For a wide, if not up-to-date, interdisciplinary panorama of empirical work on the vicissitudes of preservation, the reader is referred to Schacter (1995).

tion of a discriminating requirement upon knowledge of one's own propositional attitudes and contents; no wonder, either, Burge calls *episodic* memory (by contrast with preservative memory as he understands it) "substantive" memory: the choice of words sufficiently indicates what is *not* preserved on the model we are considering.

It appears, from what we've seen, that there are at least *two* very different senses of "preservation" at play here; also, that in the more important one (the preservation which deserves to be called, using Burge's terminology, "substantive") "self-knowledge" of the Burge-Davidson-Shoemaker kind, and accordingly preservative memory as it is depicted in the analogy with anaphoric back-reference, are not after all *content*-preservative in any plausible sense. That's one of the chief reasons, may I add in passing, why anti-realism about the past is something more than an unmotivated metaphysical fantasy.[26] In a quite determinate sense, the past is (let no one take offense) *reconstructed* as the conceptual content of memory—like that of testimony, in the case of what is often called "collective memory"—undergoes changes, whether noticed or unnoticed.[27]

Which brings us to another, slightly more surprising anti-individualistic path to exculpation: the one provided by Peter Ludlow's "Orwellian" theory of content preservation. On Ludlow's theory, "it is not the job of memory to record contents, but rather to provide information about past episodes relative to current environmental conditions" (Ludlow 1996: 316).[28] Likewise, Sven Bernecker asks: "What would be the evolutionary utility of designing memory in such a way as to make it respond to differences that, introspectively, don't make a difference? What I am suggesting is that the job of memory, rather than replay previously recorded contents, is to provide information about

26 For a perceptive examination of the relations between memory's vulnerability and anti-realism about the past see Wright (1986).

27 The significance of such remaking of the past for *personal identity* is the subject of a vast clinical psychiatric literature, examined in historical perspective in Hacking (1995), to which I will come back in the last lecture.

28 The theory is further articulated in Ludlow (2004).

past states relative to the present environmental conditions. The transfer of contents and concepts across time might be a sufficient condition for memory but it falls short of being a necessary condition" (Bernecker 1998: 241).[29]

As in Orwell's *Nineteen Eighty-Four*, the past is rewritten from the standpoint and priorities of the present. Hence, when I, on Twin Earth, recall having played in water as a kid, the content of my "Orwellian" memory is the false judgment that I played in twater (*XYZ*). The outcome is, again, a valid argument with a false premise:

(20) a. I enjoyed playing in water (*XYZ*) when I was a kid. [False]
b. This glass is full of water (*XYZ*). [True]
c. This glass is full of the same liquid I enjoyed playing in when I was a kid. [False]

I will come back to "Orwellian" memory in my last lecture. For the time being, let me just enter the suspicion that we are faced with a problem of material adequacy here: namely that there is something seriously inadequate about calling *that* "remembering".

Now there's no question that all these different construals somehow manage to take into account the fact that, as Boghossian says, the reasoner's behaviour "makes sense from his point of view". What is not so conspicuous is the sheer absence, *built into the very terms of the slow switching thought experiments*, of any other point of view against which the subject might try and assess the correctness of her reasonings. And that is my complaint against the freewheeling use of so many thought experiments in Contemporary analytic philosophy: we end up losing some of our grip on what things look like in some real life, down to earth surroundings.

The shared assumption underlying all the exculpating moves we have considered is explicitly stated by David Sosa: "*Ignorance is insufficient for incoherence*: inferring subjects are in principle in a position to avoid invalidity, no matter what their state of knowl-

29 In fairness to Bernecker I should add that he later changed his mind: see Bernecker (2010).

edge (indeed, no matter what the truth of their premiss beliefs)" (Sosa 2005: 219).

As the attentive reader will not have failed to notice, there is a further assumption at work here: namely, that ignorance is always excusable. Which, as I stressed, makes perfect sense in the contrived setting of the slow switching thought experiments. *There* was indeed virtually nothing that the switched subjects could do to prevent the fallacies of equivocation they were prone to —hence, the appeal of the exculpating moves we briefly reviewed.

Against that tide, I want to avoid the pressure towards exculpation, and the felt urge to protect the subject's rationality at the expense of her true beliefs. But then the first thing is to get rid of science fiction and follow Wittgenstein's advice, bringing words back home.

What happens, then, if we detach the examination of the main question from that framework? To begin with, we will stay on Earth and take into account the possibility that the information that the subject actually lacks about her environment *is*, after all, available; moreover, that the subject would be apprised of it had she only cared to know.

No "semantics of travel", then: uncontroversial examples of external individuation of content—beginning, foreseeably enough, with singular thoughts—should suffice for our purposes. If anti-individualism is right, then the opacity of validity will be just more widespread than we need to assume for the purposes at hand.

Suppose then as I'm coming home in the afternoon I notice this beautiful Golden Retriever dog playing around in my neighbour's front yard. I stop for a while to pet my new acquaintance, who turns out to be very amiable. As I walk home I think "That's a very friendly dog".

Now, a couple of days later, same scene—or so it seems. Here's the nice front yard with its blooming bushes and this beautiful golden dog running around. Again I stop, hoping to attract the dog's attention, resorting perhaps to whistling or finger clapping, yet this time to no avail: the dog keeps running nonstop

around the yard, heedless of my inviting moves, barking up every other tree. Maybe he's spotted a cat, who knows. I walk home thinking: "That's a very restless dog".

Am I now entitled to infer that there is a dog in my neighborhood who is both friendly and restless, as in (21), namely, the previous (15)?

(21) a. Fa
b. Ga
c. $\exists x (Fx \wedge Gx)$

Well, suppose my neighbour is a breeder of Golden Retrievers and what I successively spotted on those two occasions were a pair of siblings from the same litter —call them Argos the Friendly and Targos the Restless. As things go, Argos is not excitable at all, while Targos is of a rather unfriendly disposition. Suppose further there are no other dogs in the neighbourhood. So my conclusion is just false, and my reasoning is invalid—a plain fallacy of equivocation. Its form is not the one in (21) but rather the one in (22):

(22) a. Fa
b. Gb
c. $\exists x (Fx \wedge Gx)$

And the trouble lies in the way my mistaken empirical assumption (that there was one single dog which I encountered twice) impinges on my grasp of the logical form of the inference I performed—specifically my taking the inference to be of the form in (21).

For it's not as if I inferred validly, except that my inference relied on a tacit (and false) identity premise (namely that 'that dog_1' = 'that dog_2')—so that my reasoning was really an enthymeme:

(23)a. Fa
b. Gb
c. a = b
d. $\exists x (Fx \wedge Gx)$

That will not do, for at least two reasons. First, it's not as if I would be thinking of each dog as, say, 'the dog I saw on Tuesday' and 'the dog I saw on Friday' and then raising and answering an

identity question. On both occasions of meeting with the dogs, the demonstrative concept *this dog* did all the job of referring.

Sure, I could any time introduce two separate descriptions, say 'the dog I spotted on Tuesday' and 'the dog I spotted on Friday', and raise an identity problem; but why would I do that when I'm utterly heedless of the possibility that I met with two different dogs?

But there's worse. Suppose, to raise the problem in a wholly general setting, at t_1 I see object a_1 and think "This is *F*". At t_2, I see object a_2 and think "This is *G*". Then I draw the conclusion: "Something is both *F* and *G*". Am I entitled to that conclusion? Well, of course, provided $a_1 = a_2$. But was that a tacit premise, and my inference an enthymeme?

Kaplan considers the possibility with respect to occurrences of pure indexicals. He remarks that "You stay. Therefore, it is not the case that you do not stay" is not as it stands an instance of the law of Double Negation (as the references of the two free-standing occurrences of 'you' are left indeterminate); he then wonders whether that could be fixed by an effort to have both occurrences of 'you' refer to the same individual—fixing one's attention and trying not to blink in the meantime, as he says. But if so, the form of the argument would really be "You_1 stay. Therefore, it is not the case that you_2 do not stay", which is not valid. "Perhaps", he suggests, "we should give up on Double Negation, and claim that the argument is a valid enthymeme with the implicit premise '$\text{You}_1 = \text{you}_2$', the premise we strove to make true by fixing our attention. 'All right', said the Tortoise to Achilles, 'repeat the argument and this time remember to utter the additional premise'" (Kaplan 1989b: 589).

Now in the original Lewis Carroll scenario, Achilles had to face a regress of *inference rules*: the additional premise was at each step the statement, in the form of an ever more complicated conditional, of the rule authorising the detachment of the conclusion in the preceding step. Here, by contrast, we face a regress of *empirical assumptions*: what is at stake is not which inference rule we are supposed to follow but whether it applies to the case at hand.

Suppose then that

(24) Fa_1

(25) Ga_2

are not enough for you to infer '$\exists x (Fx \wedge Gx)$'. After all, you need to make sure that 'a_1' and 'a_2' are co-referential. That is, you need the further premise:

(26) $a_1 = a_2$

All right, said the Tortoise to Achilles: that's not going to do, either. For now you have to make sure that 'a_1' as it occurs in (24) and 'a_1' as it occurs in (26) are also co-referential; and *ditto* for 'a_2' as it occurs in (25) and 'a_2' as it occurs in (26). At which point it is manifest that, as Kaplan implies, you are embarked on a vicious regress.[30]

The regress is only stopped if at some point you can just *take for granted* that two tokens of the same type have the same semantic value. But if you are to be entitled to do that at some point into the regress, then you may as well be entitled to do it from the very beginning, so that no regress arises in the first place.

That does not mean, let me hasten to add, that identity of reference will thereby be secured, only that it will be taken for granted (or, as John Campbell says, *traded upon*) rather than taken as a premise, whether tacit or explicit, in the argument. Indeed, the very point of the regress argument lies in drawing the line between tacit premises in enthymematic reasoning and the kind of background empirical assumptions which we are bound to *riskily* take for granted in assessing deductive validity.

An identity statement will be taken as a premise, whether tacit or explicit, only if there is a difference *recognisable from the first-person perspective* (in other words, by reflection alone) between the ways the object is thought of in each of a pair of premises. That's why, should I reason from "The dog I met on Tuesday is very friendly" and "The dog I met on Friday is very restless" I would

30 The full-fledged regress argument is due to John Campbell in a paper antedating by two years Kaplan's "Afterthoughts" (see Campbell 1987; and, for a brief restatement, Campbell 1994: 75-6).

be helping myself to the identity "The dog I met on Tuesday = The dog I met on Friday" as a premise.

Mark that it's not as if, by contrast, 'this dog', as tokened in the presence of Argos, would be an exercise of the *same* concept as that exercised in a tokening of 'this dog' in the presence of Targos. Having a different extension is indeed a sufficient condition for the two tokens of 'this dog' being exercises of two different demonstrative concepts or modes of presentation. But the difference, most certainly here anyway, ain't in the head. If there are two Fregean senses at play here, they are *de re* senses, their identity partially fixed by the subject's non-representational relation to his environment. (And, again, if anti-individualism is right, then the problem is just more widespread than, for the purposes at hand, I am assuming here.)

That is what makes logical appraisal, if on occasion, a matter of luck —of one's being in the right place at the right time. It's not just what we *aim at* referring to; it's what we *succeed* in so doing; and this is not wholly up to us. That's why Kaplan's appeal to directing intentions is ultimately frustrating in just the way he suggests it is. Sure I must aim at a particular dog when I think "This dog is friendly"; but what dog (if any) I succeed in meaning is a matter of how things are in my environment, and that may well lie beyond my ken.

What then do I know by reflection alone about, say, a putative case of universal instantiation, "Everything must perish, therefore Turandot must perish"? What I *do* know is that, provided 'Turandot' refers, the inference is valid. I'm not supposed to know by reflection alone whether 'Turandot' refers: rather, when inferring I either know it on independent grounds or *riskily* take it for granted. Likewise with my staple example about the two dogs. Call that *unsafe reasoning*, as in Faria (2009): then the point I'm wanting to drive home here is that reasoning is bound to be always, to some extent, an unsafe business; also, that this is something we are apt not to understand about the subject matter of logic.

The basic idea here (minus the allusion to risk which matters to me) was encapsulated in Stephen Schiffer's response to

Paul Boghossian's paper "Externalism and Inference" back in 1992. Rejecting the suggestion that appeal to narrow content was mandatory in order to take into account the fact that the ostensibly faulty reasoning of slow-switched Peter made sense from his (Peter's) point of view, Schiffer asked: "Can't we explain Peter's being epistemically justified by observing that his belief was produced by reasoning of a form guaranteed to be valid but for undetectable externalist contingencies?" (Schiffer 1992: 37).

I think Schiffer's suggestion may be expanded into an account of our entitlement to take for granted, *as we all do most of the time*, that what he calls "undetectable externalist contingencies" are not going on. We need such an account, at any rate, if we are to make good the claim that the external individuation of thought contents is no hindrance to our being able, more often than not, to know by reflection alone whether an inference is deductively valid.

Now it would be nice to have a principled way of sorting out the relevant kinds of cases here. At a bare minimum, it seems to me that two main sorts of cases should be distinguished: those in which a rational subject will be expected to *know* that the relevant empirical facts obtain, and those in which we are entitled to *take for granted* that the relevant empirical facts obtain.

The line is rather thin here, but I have in mind the contrast between what we are able to *come* to know (a proper subset of which will comprise facts we cannot afford *not* to know, the domain of epistemic obligation), and what lies beyond the reach of cognitive achievement. It is open to me to come to know that my neighbour is a breeder of Golden Retrievers, not that objects around me are usually stable and are not switched every time I blink. It is open to me to come to know that here is a hand and here is another, not that there is an external world.

If that distinction is on the right track, then a place must be secured, in the epistemology of reasoning, as in epistemology *tout court,* for the notion of inexcusable ignorance.[31] We will accordingly take into account the possibility that, unlike what hap-

31 I come back to that idea in the last lecture.

pens with the exiles on Twin Earth (as the story is usually told), a reasoner will often be in a position to acquire the relevant information about her circumstances and environment, and thereby raise and settle explicit questions about existence, permanence, identity or difference of the objects she reasons about; also, that failure to do that may be a failure to comply with an epistemic duty. I may, as a matter of fact, disregard the possibility that the dog I saw on Tuesday is not the dog I saw on Friday, but that may be a simple case of epistemic negligence given antecedently available information.

But there is nothing I can do to ultimately satisfy myself that objects around me are usually stable and are not switched while I blink. So much has to be simply taken for granted, *even as it may, on occasion, turn out false.* Sometimes life imitates art: in 2009 two white donkeys artfully painted so as to look like zebras were on display at the Marah Land Zoo in Gaza. It is not known that the Zoo keepers ever read Dretske. It is safe to say, on the other hand, that their contrivance failed to undo the philosophical verdict that, by and large, being a cleverly disguised donkey is not a relevant alternative to being a zebra.

The same holds for our background assumptions about existence, permanence, identity, or causal properties of objects around us. To the extent that we are bound to take for granted (as opposed to come to know) that these assumptions hold, we may be said to be *a priori*, even though defeasibly, entitled to them.

Now whether we are so entitled would seem to depend, as my appeal to the notion of relevant alternatives was meant to suggest, on what is *normal*—on what is usually the case; on what Wittgenstein was wont to call *natural history*. And here an objection may be raised, to the effect that it is not up to philosophy to settle, from the armchair as it were, what is normal. Here's David Sosa again, commenting on Tyler Burge's claim that if there are any cases of equivocation due to slow switching (which Burge goes to great lengths to hold that there aren't), "they are marginal" (Burge 1998: 102). Writes Sosa: "I find Burge's approach here troubling: one might have thought that whether such cases are

typical or marginal—how often they occur, that kind of thing—is not for philosophers to judge" (Sosa 2005: 224).

I want to resist that thought. I claim that we must be entitled to our background empirical assumptions *most of the time* if reasoning is to be possible *at all*.

Can I prove that? I think John Campbell showed how to do it (see, again, Campbell 1987). Suppose you are holding a ball which feels soft and looks red. You think "This ball is soft" and "This ball is red". Since you take both judgments to be true and don't even dream of their not being about the same object, you are ready to draw the conclusion "Something is both soft and red".

But what if, instead, you pause to do a bit of epistemology? Unleash your imagination; and suppose the visual image of a ball is being conveyed to you through a set of mirrors from a rigid red ball in the adjacent room while your tactile sensations are of a soft green ball which you are at present unable to see.

The supposition will be wild but notice that just *thinking* about it is enough to deprive you of your inferential innocence. For, having raised the possibility, you have *ipso facto* introduced two new ways of thinking about what you had previously thought of as 'this ball'. You now have 'the ball I'm touching' and 'the ball I'm seeing', and an identity problem.

In Campbell's lingo, you ceased to *trade upon* the identity of the seen and the touched object; instead, you introduced a divide in the information coming to you from the object, so that your premises are now of the form "a is F", "b is G". And now, of course, it will be wise of you to refrain from drawing any conclusion from this couple of logically independent judgments.

What I want to stress is that *this was your own deed*: no Evil Genius needs to have pulled the old switcheroo so as to have you take this for that; no overnight travel from Earth to Twin Earth must have taken place. You have, by your own unassisted means, undone whatever entitlement you may have had to draw the conclusion.

Or so it seems. For whether you have indeed managed to lose your entitlement is a more delicate question than meets

the eye—one which I don't have the space to discuss here, and accordingly will content myself with sketching.[32] What I have in mind is the contrast between, on the one hand, actually envisaging, in full seriousness, the possibility that your assumption of uniqueness is really false (which amounts to take seriously the skeptical scenario in which you are in perceptual contact with two different balls) and, on the other hand, going through the motions of doing that for the sake of what I described as "doing a bit of epistemology".

Incidentally, I think the perception that this contrast is a real and important one underlies Hume's depiction, at the end of Book I of the *Treatise on Human Nature*, of the contrast between philosophical doubts and the certainties of what he calls "the natural man"; also, that it provides the main motivation for Contemporary epistemological contextualism. So David Lewis writes, for instance: "Maybe epistemology is the culprit. Maybe this extraordinary pastime robs us of our knowledge. Maybe we know a lot in daily life; but maybe when we look hard at our knowledge, it goes away. But only when we look at it harder than the sane ever do in daily life; only when we let our paranoid fantasies rip" (Lewis 1996: 550). My suggestion is that what Lewis is describing here as "looking harder" may be better understood as *making believe* (as we are wont to do when engaged in "doing a bit of epistemology") that a possibility is being seriously considered. The suggestion draws on Myles Burnyeat's invaluable discussion of what he calls "insulation" in Burnyeat (1984), as well as on Thompson Clarke's discussion of the contrast between "the plain" and "the philosophical" in Clarke (1972).

However that may be, compare that (real or simulated) outcome to the symmetrically opposite case: there actually are two objects, a rigid red ball and a soft green ball, but you cannot track the fact that your information comes from these two distinct sources, so that you keep thinking of both as 'this ball'. *Now* you are prone to draw a false conclusion from a set of true

32 I thank James Conant and Alexandre Machado for urging me to enter this qualification.

premises, exactly like those unlucky exiles on Twin Earth, Boghossian's Peter and his kin.

The conclusion I'm aiming at is now at hand. In section 401 of *On Certainty* (1969), Wittgenstein writes: "I want to say: propositions of the form of empirical propositions, and not only propositions of logic, form the foundation of all our operating with thoughts (with language)".

I think that is a difficult idea, not because it is recondite, but because it goes against the grain of a deeply entrenched picture of what Wittgenstein here describes as "operating with thoughts". On that picture, thought evolves in a frictionless medium, sealed off from the vagaries of actions, the hazards of causation and the accidents of history.

And that's what brings me to my last topic: conceptual loss and its consequences.

3. Loss

What though the radiance which was once so bright
Be now for ever taken from my sight,
Though nothing can bring back the hour
Of splendour in the grass, of glory in the flower;
We will grieve not, rather find
Strength in what remains behind.

William Wordsworth, "Ode: Intimations of Immortality" (1807)

What has been achieved, if anything, up to this point?

Richard's argument raised a problem about content preservation for temporalism. We saw how the problem is circumvented, but the solution raised a further problem. According to temporalism, retaining a belief is not believing the same temporal proposition but believing a proposition related by the truth-value link to the original proposition. Hence, if Mary believed that Obama was doing a good job in the White House and keeps believing everything she once believed, what Mary believes now is not that Obama *is* doing a good job in the White House, but that Obama *was* doing a good job in the White House. Likewise, if I believe that today is sunny I will have retained my belief tomorrow if I then believe the past tense counterpart "Yesterday it was sunny".

On this account, what is preserved is the time-neutral kernel "It is sunny", which is the operand of the past tense operator 'this was the case'. That kernel is not a propositional function as on Recanati's account, or on Evans's T_2. It is a complete time-neutral proposition. And that was the gist of the proposed solution of Richard's problem about content preservation.

But then it is vital that the time-neutral kernel itself is kept invariant. And anti-individualism, as we saw, threatens that invariance. This is in particular the case in episodes of conceptual change, which may eventuate in the unavailability of thought contents which the subject cannot retrieve even as she forgets

nothing. Hence the inadequacy of the extant compatibilist replies to Boghossian's Memory Argument with its corollary about anti-individualism and inference.

The Schiffer-Burge anaphoric account of preservative memory suffers from two limitations which seriously impair its force as a solution to Boghossian's problem. First, the (highly idealised) model is predicated on the satisfaction of empirical conditions of which we know no more than that they involve what Burge describes as "the existence of causal memory chains going back to the states which carried intentional content" (Burge 1998: 94). And then the kind of knowledge of content allegedly preserved through such anaphoric chains as posited in the model falls short of being discriminative knowledge of content: it provides nothing in the way of telling water thoughts from twater thoughts.

The "Orwellian" solution put forward by Peter Ludlow and Sven Bernecker faced a problem of material adequacy (there's something seriously inadequate about calling such systematic re-conceptions of the past *remembering*), while raising the problem which I want to discuss in this last lecture.

Let's start, uncontentiously I hope, from a metaphysical truism.

The unchangeability of the past is a metaphysical axiom if anything deserves to be so called. It would seem to be the merest common sense as well: what is done is done and cannot be undone, it's no use crying over spilled milk. Poet and metaphysician alike side with the common man: "nothing can bring back the hour of splendour in the grass"—not even, we are assured, God Almighty. So Aquinas deems it contradictory that God should undo what has once been done (*Summa Theologica* I, Q. 25 art. 4). Jewish orthodox theologians would seem to concur with the Angelical Doctor, adding the further twist that the man is sinful who, wishing for some evil *not to have* happened (say, for his missing son to be still alive), prays to God that it be so: "It is blasphemous", they'd say, "to pray that something should *have* happened, for, although there are no limits to God's power, He cannot do what is logically impossible; it is logically impossible to alter the past, so to utter a retrospective prayer is to mock God by asking Him to perform an impossible task" (Dummett 1964: 335).

That almost universal consensus has been challenged, we are often reminded of nowadays, by advances in theoretical physics. Gödel models of Relativity Theory are a case in point. In "A Remark About the Relationship between Relativity Theory and Idealistic Philosophy" Gödel writes: "By making a roundtrip on a rocket ship in a sufficiently wide curve, it is possible in these worlds to travel into any region of the past, present, and future, and back again, exactly as it is possible in other worlds to travel to distant parts of space" (Gödel 1949: 205).

Gödel's models raise issues which I am not prepared to discuss here, except perhaps to warn against what Wittgenstein called the 'prose' which often surrounds a mathematical achievement. An example of that kind of prose is Palle Yourgrau's book *A World without Time*, in which it is claimed that there has been in the scientific community "a conspiracy of silence", "one of the intellectual scandals of the past century" aimed at hiding the fact that, rather than explaining time, what Einstein did was to explain it away. It would then have been a scientific discovery, which the whole scientific community is guilty of keeping secret, that there is no time and McTaggart was right all along (see Yourgrau 2005).

Whatever we are to think of Gödel's foray into theoretical physics, the issue is of no consequence for the problem I want to discuss. For there is another sense in which the past would seem to be liable to change even should there be no logically possible world (let alone a physically possible one) in which what was once done should be undone; and that is my current concern.

The envisaged possibility is for something which has happened (granted, for the sake of the argument if you like, that it could not possibly, once happened, ever *cease to have* happened) that it should later *come to have been* what, at the time it happened, it was not. *The past would be changed, that is, not through retroactive undoing but through retroactive becoming.*

That this might look like a real possibility was brought to the fore by Donald Davidson, to whom we owe, to the best of my knowledge, the first published statement of a paradox that has since plagued philosophical treatments of actions and events.

Davidson's puzzle, as I will call it, made its appearance in the context of an examination of purported principles of event identity—principles, that is, providing criteria for the truth of statements of the form "*A = B*" in the range of cases where '*A*' and '*B*' are terms (names or descriptions) standing for events (Davidson 1969).[33]

It would seem to be a truism that numerically identical events will take the same amount of time to come about. Yet, this is easily shown to lead to paradox. We learn from the spectre's disclosure to the Prince of Denmark that Claudius killed King Hamlet by pouring poison in his ear. There is, then, in the story told in Shakespeare's play (and then retold in the play-inside-the-play staged by Prince Hamlet), a killing that was a poisoning. Yet, there are as well two clearly distinguishable events: Claudius's pouring of the poison and the King's death. One precedes the other and causes it. Now where does Claudius's killing of the King come in? Current wisdom in the philosophy of action agrees with common sense in taking Claudius's killing of the King to be identical with his poisoning him. Are we to conclude that Claudius killed King Hamlet before the latter died? (Davidson 1969: 177).

Understandably enough, many a philosopher has taken the puzzle to show that Claudius's pouring of the poison and his killing of King Hamlet could not possibly be the same event, albeit under two different descriptions.[34] I won't rehearse this line of response, which I take to have been conclusively shown (by Anscombe, Davidson, and others) to get quickly driven into worse trouble. If (a) *Claudius's pouring of the poison* was not the same event as (b) *his killing of King Hamlet*, then neither was it the same as (c) *his poisoning of the King*; so we'd already have three distinct events, of which (b) would presumably have come about later than (c), which in turn would have taken place later than (a).

---

33 The puzzle is, however, detachable from the framework where it first made its appearance—a point of some consequence for the discussion below. So, we find Elizabeth Anscombe, who denies that we are in need of any theory of event identity, in the grip of the same problem in Anscombe (1979).

34 See, *e.g.*, Thomson (1971).

But then, by the same token, neither would (d) *Claudius's turning the uncorked vial* (so as to let the poison out) be the same as his pouring of the poison, as the latter takes longer to reach completion. At which point a blind man can see that this is not going to end soon: for, given continuity, any event will be endlessly divisible in temporal parts, any aggregate of which will have to count, by parity of reasoning, as a distinct event.[35] The right thing to say is rather that, as Claudius turned the uncorked vial, he *thereby* poured poison in the King's ear, *thereby* poisoning him, *thereby* killing him; and that here we have not four different acts which Claudius performed but a single one: for turning the vial *was*, in these circumstances, pouring poison; and, in these circumstances, it *was* poisoning the King; and, in these circumstances, it *was* killing him, as there's nothing else which (in these circumstances) Claudius had to do in order to kill King Hamlet.[36] What Claudius did (the action he performed, the event which *was* the action) was already, and wholly, done once he finished pouring the poison.

Yet, at the time he was done with the pouring, the King was presumably still, if for a short while, alive. Should you be bothered by the "short while", feel free to change the example so as to have the whole story take as long as you fancy: think, say, of a terrorist who sets a clock bomb to explode long after every member of his organisation will have left the country. I say it makes no difference: killing will take time, however little. So, again: are we to say that in such cases an event took place which, as things turned out, proved to *have been* a killing even though it was over before anyone died?

Davidson's own view, which I share, is that this is indeed what we should say, so we better find a way to reconcile ourselves to the seeming paradox.[37] Yet Davidson's first stab at how to get

35 This is not the whole story, but is as much as is required for our current purposes. For further details see Anscombe (1979).

36 See, for the *locus classicus* of this trimming, with the blade of Ockam's razor, of a vicious multiplication of events, Anscombe (1957).

37 An alternative path, whose attractions have been pressed upon me by some good willing friends, is to take Davidson's puzzle as a *reductio* of the view that

there is apt to sound as a nonstarter. We are reminded that we may often be in a position to know that an event is a pouring of poison without knowing that it is a killing, "just as we may know that an event is the death of Scott without knowing that it is the death of the author of *Waverley*" (Davidson 1969: 177). Now, it is unlikely, to say the least, that the old leaf from Russell's book should strike many a reader as the high road to easiness about retroactive becoming. For one thing, general failure of substitutivity in intensional contexts is insensitive to the modal difference which, here, would make all the difference: I mean that between, on one hand, cases where it is *possible that you don't know* that *the G* is *F* even though you know that *the H* is *F* and, as a matter of fact, *the G* = *the H*; and, on the other, cases where it would be *impossible that you know* that *the G* is *F*—indeed, logically impossible as, much as you might know the G "by any other name", you would still be debarred from knowing it *as the G* for the plain reason that something further would still have to happen so that it *came about* that it was (after all) a G.[38]

Now that would hardly be news to Davidson (or to Anscombe, for that matter). The puzzle was not supposed to be a matter of what the event is thought or said or justifiedly believed to be, but of what it is. But then we are entitled to ask for some principle linking the availability of a "description under which" something happened and some description-independent matter of fact. Otherwise, retroactive becoming boils down to retrospective re-description, a possibility unlikely to arise much excitement beyond the precincts of metaphysical anti-realism.[39]

Such a linking principle is, I take it, what Davidson is gesturing at when he writes, in the immediate sequel to his quoted remark on 'knowing-that' clauses: "To describe an event as a

---

actions are events. After much thinking, I still don't know that I am prepared to embrace the offer—and anyway the problem about events would remain untackled.

38 If that's how the story should be told—whether so is what this is all about, indeed.

39 Which are, to be sure, wide enough to deserve some attention, as we will see in a while. See, for a start, Dummett (1969).

killing is to describe it as an event (here an action) that caused a death, and we are not apt to describe an action as one that caused a death before the death occurs; yet it may be such an action before the death occurs".[40]

On the face of it, this is just the paradoxical conclusion restated. Yet the focus on causation shifts the emphasis from what we may not know (as a matter of mere ignorance) to what would arguably *not be there* to be known: namely, that some currently unfolding event (a pouring of poison) will actually cause something else (a death) to come about, *thereby*, and only then, making it right to describe it as a killing. For, in the best of possible cases, what we will know is that, barring interference or prevention, the currently unfolding event will turn out to have been a killing.

I say "barring interference or prevention": the adverbial clause is there to stress the importance, and the elusiveness, of what Davidson calls "directness of causal connection". After all, we would think, it's not all settled, not that often anyway, once and for good. Here's how Davidson puts it: "To describe the pouring as a killing is to describe it as the causing of a death; such a description loses cogency as the causal relation is attenuated. In general, the longer it takes for the effect to be registered, the more room there is for a slip, which is another way of saying, the less justification there is for calling the action alone the cause" (Davidson 1969: 177). Now slips come in all shapes and sizes, and the point bears on delicate issues in the philosophy of causation. For all that, having or lacking justification is a matter of epistemic entitlement, not of "attenuation" (a bad word at any rate) of causal relations; and, for all Davidson says, there may well be a matter of fact whether the pouring was a killing, however long it may have taken for the death to come about, however complicated the whole story may have gotten, and whether we

40 Davidson (1969: 177). Compare: "If it is not true to say it is between t and t' that A has killed B, that will be because B isn't dead until t'. So much is agreed. But it merely means that, although that act has occurred which, as things turn out, will prove to be an act of killing, things have not yet turned out so" (Anscombe 1979: 214).

are ever justified in believing it or not. Seen in that light, retroactive becoming now starts to look like less of a paradox: indeed, the very idea that it was a form of change in the past loses a good deal of its grip.

Not once and for all, however. Think of a long dead man who becomes the father or grandfather of a newborn child; and then of a person who, under therapy, comes to think of herself as having been sexually abused in early childhood. These are different cases, apt to elicit different responses. Now here's what Anscombe writes about the former: "Only of a (then) living man do we say that he became a father or a grandfather; but it can *come about* that someone was a father or a grandfather even though he is dead" (Anscombe 1979: 215). No qualm about it coming about, yet mark the tense: he *was* a father or a grandfather; but he was already dead when the child was born, so he was a father or a grandfather before the child was born. And here's Hacking about the latter case: "Child abuse is a new kind that has changed the past of many people, and so changed their very sense of who they are and how they have come to be" (Hacking 1992: 230). Again, no worry about people achieving a changed "sense of who they are and how they have come to be", yet mark the conjunction 'so' in "so changed their very sense of who they are": if they managed to achieve it, that was because the newly acquired description matches a new kind (of behaviour) whose emergence "has changed the past of many people". Here's how Hacking puts it elsewhere: "If a description did not exist, or was not available, at an earlier time, then at that time one could not act intentionally under that description. Only later did it become true that, at that time, one performed an action under that description" (Hacking 1995: 243).

Notice the absence of anything like Davidson's linking principle: hence, of any trace of the modal distinction we found to be crucial to tell retroactive becoming from mere retrospective re-description.

Now, let there be no doubt about it, this is a liability of episodic memory, one which psychologists study under the label of memory distortions. Descriptions previously unavailable (in the

most interesting cases, to the agent himself) may be brought to bear on the action, and that's how the agent *becomes* retroactively a child abuser or a multiple personality; or else, a description under which the agent may have thought of his action may become unavailable to the describer, and a deserter in World War I becomes retroactively a victim of post-traumatic stress disorder (Hacking 1995: 241).

The latter case is particularly interesting in that the acquisition of a new description under which the action is now held to have been performed ("post-traumatic stress disorder") is accompanied by the loss of the description under which the action *was* performed: the result is that the agent's behaviour is emptied of its ethico-political significance, which becomes unavailable to the agent himself. Yet, it would be a mistake to describe that as a case of forgetting. *What kind of loss is undergone by the subject is a matter of what kind of good a concept is* —a topic to which I will come back presently.

Notice, in the meantime, that we are now in a position to reassess Boghossian's alleged truism about memory —namely that if S forgets nothing, then what S knows at t_1, S knows at t_2. *The sort of conceptual loss we are dealing with amounts to the unavailability of a previous thought which comes about through no forgetting.*

That's how Anscombe's account of intentionality —that an action can only be reckoned intentional, or unintentional, under one or another of its possible descriptions —features at the heart of Hacking's "dynamic nominalism": what started a nominal kind becomes, through "the looping effect of human kinds", an entrenched, that is a real kind.[41]

Here is the looping effect explained by Hacking in "Making Up People": "What is curious about human action is that by and

41 As a couple of by-products, Hacking manages to give a plausible sense to Nelson Goodman's otherwise unadulteratedly nominalistic claim that the entrenchment of the extension of a predicate "derives from the use of language" (see Goodman 1954: 95) and a precise sense to Michel Foucault's *prima facie* outrageous contention that until the 19th Century there were neither mentally ill people (although there have always been plenty of crazy people) nor homosexuals (nor, for that matter, heterosexuals either, although there has always been plenty of sexual practices of all flavours)—no small feat either, to be sure.

large what I am deliberately doing depends on the possibilities of description [...] this is a tautological inference from what is now a philosopher's commonplace, that all intentional acts are acts under a description. Hence, if new modes of description come into being, new possibilities for action come into being in consequence" (Hacking 1986: 108).

And again in "World-Making by Kind-Making: Child Abuse as an Example": "Trivial enough: only people can understand what they are called and how they are described, so only people can react to being named and sorted. But it becomes an important difference in kinds when we realise that entities—people and their acts—of a kind can change in response to being so grouped, that the group thereby changes, and hence our characterisation of the group itself has to be revised. In this way human kinds have feedback. A looping effect unknown in the inhuman world" (Hacking 1992: 190).

And that's how we come to have "an indeterminacy in the past": "I am not about to address that banal topic, the indeterminacy of memory. I mean an indeterminacy about what people actually did, not about what we remember them doing. I mean an indeterminacy about past human actions, where it is something about our actions, not our memories of them, that is indeterminate" (Hacking 1995: 234).

What are we to make of these claims? I propose to approach them from sideways on, starting from well-known examples of conceptual loss, specifically as they feature in the history of science.

Loss of conceptual contents, and the problems facing any attempt at retrieving them lie, to give one single example, at the very heart of Thomas Kuhn's philosophy of science. Ever since *The Copernican Revolution* (1957), Kuhn was busy identifying and exposing—with the richness of detail that only the *empirical* study of concrete examples of conceptual change can provide—the fallacious inferences pervading, most of the time tacitly, the historiography of science, and marring the philosophical understanding of the history of sciences. In a brutally simplified example, we might feel prone to reason thus:

(27) a. Aristotle said that the Earth is not a planet.
b. Galileo said that the Earth is a planet.
Ergo,
c. Galileo contradicted Aristotle.

That, as Kuhn showed for a wealth of much more sophisticated examples, is a fallacy because the word 'planet', in its two occurrences, expresses two distinct concepts, with different, only partially intersecting extensions.

In (27a) 'planet' means something like *heavenly body moving along an independent path, with respect to that of the stellar sphere, around the unmoving center of the universe (sc. the Earth)*, and its extension is the set {Sun, Moon, Mercury, Venus, Mars, Jupiter, Saturn}: in *that* sense "The Earth is not a planet" is an "analytic" proposition (one reason why, by the way, it's also something which, as far as I know, Aristotle never bothered to say). In (27b) 'planet' means something like *non-luminous heavenly body which revolves around a star*, and its extension is the presumably *infinite* set of which {Mercury, Venus, Mars, Jupiter, Saturn} is a tiny subset; and in that sense "The Earth is a planet" is an *empirical* (and not an "analytic") truth.

That was the true source of incommensurability, which raised such a scandal among the "rationalists" in the philosophy of science, and which has been so badly abused, on the other hand, by relativists, deconstructionists, and bad anthropologists. The phenomenon I've been describing may be redescribed, from the larger perspective afforded by the comparison between the vicissitudes of preservation in memory and in tradition, as the incommensurability *with oneself* which is the outcome of an *intra-subjective* "paradigm change". As Campbell remarks in another context: "In making an earlier judgement and in making the current judgement, you were aiming at truth both times; and your success in the earlier enterprise affects your prospects of success in the later enterprise. But if your ways of understanding the proposition at different times were incommensurable, there would be no way in which the two enterprises could be connected" (Campbell 2001: 179).

In the Afterword to *Black-Body Theory and the Quantum Discontinuity, 1894-1912*, Kuhn writes: "Entry into a discoverer's culture

often proves acutely uncomfortable, especially for scientists, and sophisticated resistance to such entry ordinarily begins with the discoverer's own retrospects and continues in perpetuity [...] Systematics distortions of memory, both the discoverer's memory and the memory of many of his contemporaries, are a first manifestation of resistance" (Kuhn 1978: 364).

Kuhn illustrates the point with two examples drawn from his interviews with the surviving protagonists of the quantum revolution, in this case Otto Stern and Niels Bohr. The details of the examples matter less than the common pattern they exhibit: a revolutionary scientist is unable to make sense of some of his own earlier work through imposing upon it the new conceptual framework in which he has been working since the heyday of the revolution. Kuhn sums up: "Not always but quite usually, scientists will strenuously resist recognising that their discoveries were the products of beliefs and theories incompatible with those to which the discoveries themselves gave rise. Similar resistance is encountered among later generations, but memory and its distortion are no longer involved" (1978: 366).

That was the epistemological obstacle described in *The Structure of Scientific Revolutions* as "the invisibility of revolutions". There Kuhn wrote: "Inevitably those remarks will suggest that the member of a mature scientific community is, like the typical character of Orwell's *1984*, the victim of a history rewritten by the powers that be. Furthermore, that suggestion is not altogether inappropriate. There are losses as well as gains in scientific revolutions, and scientists tend to be peculiarly blind to the former" (Kuhn 1962: 167).

Well, here is Orwell himself: "The really frightening thing about totalitarianism is not that it commits atrocities but that it attacks the concept of objective truth: it claims to control the past as well as the future" (Orwell 1944: 88). "Nazi theory indeed specifically denies that such a thing as *the truth* exists. There is, for instance, no such thing as *science*. There is only *German science, Jewish science,* etc. The implied objective of this line of thought is a nightmare world in which the Leader, or some ruling clique, controls not only the future but *the past*. If the Leader

says of such an event, *It never happened* —well, it never happened. If he says that two and two are five —well, two and two are five" (Orwell 1943: 258-9).

I said that assessing what exactly is lost when concepts are lost is a matter of what kind of good a concept is. The topic is an elusive one because, with the exception of Wittgenstein and the philosophers he influenced, the analytic tradition in Contemporary philosophy has inherited the empiricist picture of a concept as a device for classifying. On that construal, the loss of a concept amounts to the loss of a discriminating capacity, much as acquired colour-blindness, if there were such a thing, would amount to the loss of the capacity to tell red from green things. Concepts are instruments of description, and description is in the service of classification.

Against that tide, Wittgenstein has put forward a view of concept possession on which, apart from and actually underlying recognitional capacities, *possessing a concept implies participating in a system of reactions, attitudes, dispositions (what Wittgenstein called a form of life)*.

The adoptive father of the robot boy David, in Steven Spielberg's film *Artificial Intelligence*, knows pretty well what he is doing when he presses on his wife the fact that David is just *a toy*: if you don't want the consequences (and "the consequences" here are *all* the consequences) of treating David as a person, the best you can do is to avoid using this concept when thinking and speaking about David.

As this example clearly illustrates, it is not just the *logical* structure of our concepts, call it their inferential role ("If David is a person, then...") which is at stake. For it is not just a matter of our rational commitments: being in the possession of a concept implies, as Wittgenstein stressed, participating in a system of reactions, attitudes, dispositions: we don't treat alike a toy and a human being, we don't react alike to their behaviour. Our *feelings* are just as essentially tied to concept-possession as our recognitional capacities. In a word, what we can *think* depends also on what we can *feel*.

Cora Diamond discusses, from a Wittgensteinian perspective, two telling examples about the use of the concept of a *human*

being and how that concept differs from that of *a member of the species Homo Sapiens.*

In Tolstoy's *War and Peace* there is a scene in which Pierre Bezukhov's life is saved. Pierre is brought, a prisoner, before General Davoût, who, when he first looks up from the papers on his writing table, sees Pierre, who is standing before him, only as the present prisoner, the present circumstance to be dealt with; but something in Pierre's voice makes him look at him intently. At that moment, says the narrator, "an immense number of things passed dimly through both their minds". Tolstoy says nothing of what things are these; but in that look human relations between the two men are established; and it is that look which saves Pierre's life.

Now contrast that with a scene in Primo Levi's *Se questo è un uomo.* Levi, the prisoner in Auschwitz, is brought before Panwitz, a chemist in a supervisory position at the Buna factory where some Auschwitz prisoners worked. If Levi's credentials as a chemist are accepted, he will be assigned to work at Buna, he will not be destroyed by horrendous physical labor in the cold, he will be far more likely to survive the "selections". Panwitz sits at his writing table, Levi stands before him, and the look that passes between them "was not one between two men".

Diamond comments: "Those two scenes by those two writers show us what there can be in a look, what sense of the sharedness of human life, what denial of that solidarity, what the depth can be of recognition and of its denial" (Diamond 1988: 265).

Briefly, we will have lost the concept of a human being if we come to the point of identifying it with *member of the species Homo Sapiens.*

The topic has been, understandably enough, much discussed in moral philosophy, at least since Elizabeth Anscombe's (1958) influential paper "Modern Moral Philosophy". In that paper she argued that the concept of moral obligation presupposed that of a moral legislation and that, on its turn, was only intelligible against the background of belief in a divine lawgiver. The decline of that belief in modern times had the consequence that the phrase "moral obligation" fails to express any concept, and that those, be they philosophers or laymen, who use such phrase,

go through the motions of leading a moral life which became unavailable to them.

Anscombe's criticism of deontological ethics was taken up and generalised by Alasdair McIntyre in *After Virtue*. It is McIntyre's main thesis that key moral notions still in use no longer have the kind of context essential for their significance. As he writes, "the language and the appearances of morality persist even though the integral substance of morality has to a large degree been fragmented and then in part destroyed" (McIntyre 1981: 5).

Whatever we make of McIntyre's thesis, or any of the examples just discussed, they bring to the fore the extent to which conceptual loss goes beyond loss of recognitional capacities, and amounts to an impoverishment of experience. They also enable us to make sense of Kuhn's much-discussed claim that after a revolution scientists live in a different world. Specifically, they make available for us a reading of that claim on which it does not imply the anti-realism Kuhn is often saddled with. Here is a selection of passages from *The Structure of Scientific Revolutions:*

> The very ease and rapidity with which astronomers say new things when looking at old objects with old instruments make us wish to say that, after Copernicus, astronomers lived in a different world. In any case they responded as though that was the case (Kuhn 1962: 117).
>
> At the very least, as a result of discovering oxygen, Lavoisier saw nature differently. And in absence of some recourse to that hypothetical fixed nature that he "saw differently", the principle of economy will urge us to say that after discovering oxygen Lavoisier worked in a different world (Kuhn 1962: 118).
>
> When [the chemical revolution] was done, even the percentage composition of well-known compounds was different. The data themselves had changed. That is the last of the senses in which we may want to say that after a revolution scientists work in a different world (Kuhn 1962: 135).

The post-revolutionary world contains the same individuals, but not the same kinds.[42] That's the gist of Kuhn's criticism of

42 See the excellent discussion in Hacking (1993).

Putnam's and Kripke's thesis that natural kind words are rigid designators, as it is articulated in "Possible Worlds in History of Science" (Kuhn 1989) and "Dubbing and Redubbing: The Vulnerability of Rigid Designation" (Kuhn 1990). As Kuhn puts it in the discussion published as an Appendix to *The Road Since 'Structure'*: "You can trace the individual planets, Mars, heavenly bodies through the Copernican revolution—what you can't trace through it is 'planets'. Planets are just a different collection before and after the revolution" (Kuhn 2000: 312).

That was a source of incommensurability; it was also, by the same token, a source of the invisibility of revolutions discussed by Kuhn.

It remains to be seen, by way of conclusion, where that leaves us. I want to bring this lecture to a close by taking a look at a widespread assumption about the connection between responsibility and control—for it is the main lesson I want to draw of this inquiry in cognitive dynamics.

This is how the story goes: you should not be held responsible for what evades your control. It is not for you to stop the autumn leaves from falling. "Ought", we are told, implies "can".

The story has been challenged. Moral luck has been a lively issue in ethics since the exchange between Bernard Williams and Thomas Nagel which sparked the debate over thirty years ago (see Williams 1976, Nagel 1976). The absent-minded driver who passes a red light may be lucky enough that no pedestrian is crossing the street; but, red lights or no, if a child comes running after a ball and gets hit, that will make all the difference. The *moral* difference: "We feel sorry for the driver, but that sentiment co-exists with, indeed presupposes, that there is something special about his relation to the happening, something which cannot merely be eliminated by the consideration that it was not his fault" (Williams 1976: 28).

Jurisprudence, a realm where consequences *always* matter, has long known the doctrine of strict liability, which features in judicial settlements of torts and criminal accountability. In the English case of Rylands v. Fletcher (1868), the defendants' underground water reservoir caused an old mine shaft owned

by the plaintiff to collapse; although the court found that the defendants were not negligent, they were still strictly liable for damages.[43]

It's not so much that "ought" will not imply "can", just like that. Rather, there will be more to "can" than meets the eye. Specifically, there are some tricky questions to be faced in each case concerning how actual control (or lack thereof) relates to control in what we keep calling, for want of better words, close enough possible worlds. The unlucky driver who hit the child had actually no control over the outcome; but *it might have been otherwise*: that's how issues of negligence bear on such matters.

Now, moral luck itself is not my present concern; but the background assumption about the connection between responsibility and control lies at the heart of what I have been after all along.

My aim is, in conclusion, to bring into sharper focus what Roy Sorensen, in a highly perceptive (if, sad to say, largely unheeded) paper, proposed to call "logical luck" (Sorensen 1998). It was Sorensen's original insight that current debates on content externalism (specifically as they bear on the apriority of our logical abilities) are fruitfully illuminated when set against the framework provided by comparison with the *prima facie* unrelated topic of moral luck. I would, given time to concentrate on the topic, have some reservation to voice about Sorensen's unabashedly consequentialist approach to blameworthiness; unlike Sorensen, moreover, I would make essential use of a distinction between excusable and inexcusable ignorance: something which the prevailing approach to the "externalism and inference" debate (relying, as it does, on the slow switching thought experiments introduced by Burge) has made all but invisible.

But the essential thing is that, as Timothy Williamson puts it, "one may be rationally required to do something even though one is not on a position to know that one is rationally required

43 Compare Article 927, single paragraph, of the Brazilian Civil Code: "The defendant will be liable for damages, *independently of guilt*, in such cases as are stipulated in statute, or when the defendant's usual activity implies, by its own nature, risk for the rights of others" (my emphasis).

to do it. If we imagine that some candidate criterion of rationality is perfectly accessible, then we are always likely to prefer that criterion; but once we recognise that perfect accessibility is quite generally an unattainable ideal, we can learn to live with an imperfectly accessible criterion. We have nothing else to live with" (Williamson 2000: 15-16).

The legal philosopher H.L.A. Hart opens a paper on criminal responsibility with the following imagined dialogue: "I didn't *mean* to do it: I just didn't think". "But you ought to have thought" (Hart 1961: 136).

"Ought" *does* imply "can". It is not, in such cases as Hart discusses, as if there was *nothing* the defendants could possibly have done to avoid the unfortunate outcome. Rather: what *actually* evades the control of the subject *would* be under her full control in close enough (and, moreover, *epistemically accessible*) possible worlds. The owner of a rundown house may, as a matter of fact, ignore that the ceiling is about to collapse. *Yet, that would be a manifest case of inexcusable ignorance.* The relevant information was fully available, had she only cared to go after it.

In such down to earth cases, the information which the subject actually lacks *is* available: the subject would be apprised of it if only she cared enough to know. Not so on the slow switching scenarios—hence the exculpating moves of which I was complaining in yesterday's lecture.

In what is likely to be one of the most gnomic elucidations of the concept of inexcusable ignorance, Wittgenstein writes: "That I am a man and not a woman can be verified, but if I were to say I was a woman, and then tried to explain the error by saying I hadn't checked the statement, the explanation would not be accepted" (Wittgenstein 1969: § 79).

Epistemology, including the epistemology of reasoning, has been, for most of its history, this cantankerous discipline in which we keep asking whether and how we possibly know anything. Much of what I was arguing for may be summed up in the idea that it is high time for a change of gear. Here is Robert Louis Stevenson, and let him have the last word: "I told him I was not much afraid of such accidents; and at any rate judged it unwise

to dwell upon alarms or consider small perils in the arrangement of life. Life itself, I submitted, was a far too risky business as a whole to make each additional particular of danger worth regard" (Stevenson 1876: 53).

Bibliographical References

Anscombe, G.E.M. (1957). *Intention*, second edition (Ithaca & New York: Cornell University Press, 1963).
—. (1958). "Modern Moral Philosophy", *Philosophy* 38: 1–19, reprinted in *Collected Philosophical Papers, Volume III: Ethics, Religion and Politics* (Oxford: Basil Blackwell, 1981): 26-42.
—. (1979). "'Under a Description'", *Noûs* 13: 219-233, reprinted in *The Collected Philosophical Papers of G. E. M. Anscombe, volume II: Metaphysics and the Philosophy of Mind* (Oxford: Basil Blackwell, 1981): 208-219.

Aristotle. *Categories* and *De Interpretatione*, translated with notes and a glossary by J.L. Ackrill (Oxford: Clarendon Press, 1993)

Bernecker, Sven (1998). "Self-Knowledge and Closure", in Peter Ludlow & Nora Martin (eds.), *Externalism and Self-Knowledge* (Stanford: CSLI Publications, 1998): 333-347.
—. (2010). *Memory: A Philosophical Study* (Oxford: Oxford University Press).

Boghossian, Paul A. (1989). "Content and Self-Knowledge", *Philosophical Topics* 17: 5-26, reprinted in *Content and Justification* (Oxford: Clarendon Press, 2008): 139-158.
—. (1992). "Externalism and Inference", *Philosophical Issues* 2: 11-28.
—. (1994). "The Transparency of Mental Content", *Philosophical Perspectives* 8: 33-50, reprinted in *Content and Justification* (Oxford: Clarendon Press, 2008): 159-176.

Brogaard, Berit (2012). *Transient Truths: An Essay in the Metaphysics of Propositions* (Oxford: Oxford University Press).

Burge, Tyler (1988). "Individualism and Self-Knowledge", *Journal of Philosophy* 85: 649-663, reprinted in *Cognition through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection (Philosophical Essays, Volume 3)* (Oxford: Oxford University Press, 2013): 55-67.
—. (1996). "Our Entitlement to Self-Knowledge", *Proceedings of the Aristotelian Society* 96: 91-116, reprinted in *Cognition through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection (Philosophical Essays, Volume 3)* (Oxford: Oxford University Press, 2013): 68-87.

—. (1998). "Memory and Self-Knowledge", in Peter Ludlow & Nora Martin (eds.), *Externalism and Self-Knowledge* (Stanford: CSLI Publications, 1998): 351-370, reprinted in *Cognition through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection (Philosophical Essays, Volume 3)* (Oxford: Oxford University Press, 2013): 88-103.

Burnyeat, Myles (1984). "The Sceptic in His Place and Time" in Richard Rorty, J.B. Schneewind & Quentin Skinner (eds.), *Philosophy in History* (Cambridge: Cambridge University Press): 225-254, reprinted in revised and expanded version in Myles Burnyeat & Michael Frede (eds.), *The Original Sceptics: a Controversy* (Indianapolis, IN: Hackett Publishing Company): 92-126.

Camp, Jr., Joseph L. (1987). "Is Sense Transparent?", *Proceedings of the Aristotelian Society* 88: 273-92.
—. (2002). *Confusion: A Study in the Theory of Knowledge* (Cambridge, MA: Harvard University Press).

Campbell, John (1994). *Past, Space, and Self* (Cambridge, MA: The MIT Press).
—. (2001). "Memory Demonstratives", in Christoph Hoerl & Theresa McCormack (eds.), *Time and Memory: Issues in Philosophy and Psychology* (Oxford: Clarendon Press).

Clarke, Thompson (1972). "The Legacy of Skepticism", *The Journal of Philosophy* 69: 754-769.

Collins, John, (2008). "Content Externalism and Brute Logical Error", *Canadian Journal of Philosophy* 38: 549-574.

Davidson, Donald (1969). "The Individuation of Events", in Nicholas Rescher (ed.), *Essays in Honor of Carl G. Hempel* (Dordrecht: D. Reidel: 216-34, reprinted in *Essays on Actions and Events*, 2nd edition (Oxford: Oxford University Press, 2001): 163-179.
—. (1984). "First-Person Authority", *Dialectica* 38: 101-111, reprinted in *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press, 2001): 3-14.
—. (1987). "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association 1987*: 441-58, reprinted in *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press, 2001): 15-38.
—. (2004). *Problems of Rationality* (Oxford: Clarendon Press).

Diamond, Cora (1988). "Losing Your Concepts", *Ethics* 98: 255-277.

Dummett, Michael (1964): "Bringing About the Past", *Philosophical Review* 73: 338-59, reprinted in *Truth and Other Enigmas* (Cambridge, MA: Harvard University Press, 1978): 333-350.
—. (1969) "The Reality of the Past", *Proceedings of the Aristotelian Society* 69: 239-258, reprinted in *Truth and Other Enigmas* (Cambridge, MA: Harvard University Press, 1978): 358-374.

Epictetus. *Discourses*, in *Discourses, Fragments, Handbook*, translated by Robin Hard (Oxford: Oxford University Press, 2014): 1-277.

Evans, Gareth (1985). "Does Tense Logic Rest upon a Mistake?", in *Collected Papers* (Oxford: Clarendon Press): 343-363.

Falvey, Kevin (2003). "Memory and Knowledge of Content", in Susana Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, MA & London: The MIT Press): 219-240.

Faria, Paulo (2009). "Unsafe Reasoning: A Survey", *Dois Pontos* 6: 185-201.

Findlay, J. N. (1941). "Time: A Treatment of Some Puzzles", *Australasian Journal of Psychology and Philosophy* 19: 216-235.

Frege, Gottlob (1912). Anmerkungen Freges zu: Philip E. B. Jourdain, *The Development of the Theories of Mathematical Logic and the Principles of Mathematics*, in Ignacio Angelelli (ed.), *Kleine Schriften* (Darmstadt: Wissenschaftliche Buchgesellschaft & Hildesheim: Olms, 1967): 327-328.
—. (1918). "Der Gedanke: eine logische Untersuchung" in Günther Patzig (ed.), *Logische Untersuchungen* (Göttingen: Vandenhoeck & Ruprecht, 1966): 30-53.

Geach, Peter T. (1949). Critical notice of *Nicolaus of Autrecourt. A Study in 14th Century Thought*, by Julius Rudolph Weinberg, *Mind* 58: 238-45.
—. (1955). Review of *Stoic Logic* by Benson Mates, *The Philosophical Review* 64: 143-5.

Gödel, Kurt (1949). "A Remark about the Relationship between Relativity Theory and Idealistic Philosophy", in P.A. Schilpp (ed.), *Albert Einstein – Philosopher-Scientist* (La Salle, IL: Open Court): 557-562, reprinted in *Collected Works, Volume II: Publications 1938-1974,* Solomon Feferman, Stephen C. Kleene, John W. Dawson, Jr., Gregory H. Moore, Robert M. Solovay & Jean van Heijenoort (eds.) (Oxford: Oxford University Press, 1990): 202-206.

Goodman, Nelson (1954) *Fact, Fiction, and Forecast*, fourth edition (Cambridge, MA & London: Harvard University Press).

Hacking, Ian (1986). "Making Up People", in Thomas C. Heller *et alii* (eds.), *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought* (Stanford: Stanford University Press), reprinted in *Historical Ontology* (Cambridge, MA & London: Harvard University Press, 2002): 99-114.
—. (1992). "World-Making by Kind-Making: Child Abuse as an Example", in Mary Douglas & David Hull (eds.), *How Classification Works: Nelson Goodman among the Social Sciences* (Edinburgh: Edinburgh University Press): 180-238.
—. (1993). "Working in a New World: The Taxonomic Solution", in Paul Horwich (ed.), *World Changes: Thomas Kuhn and the Nature of Science* (Cambridge, MA & London: the M.I.T Press).
—. (1995). *Rewriting the Soul: Multiple Personality and the Sciences of Memory* (Princeton: Princeton University Press).

Hart, H.L.A. (1961). "Negligence, *Mens Rea* and Criminal Responsibility", in A.G. Guest (ed.), *Oxford Essays in Jurisprudence* (Oxford: Clarendon Press), reprinted in *Punishment and Responsibility: Essays in the Philosophy of Law* (Oxford: Clarendon Press, 1968): 136-157.

Heil, John (1988). "Privileged Access", *Mind* 98: 238-251.

Hintikka, Jaakko (1973). "Time, Truth, and Knowledge in Ancient Greek Philosophy", in *Time and Necessity: Studies in Aristotle's Theory of Modality* (Oxford: Clarendon Press): 62-92.

Kamp, Hans (1971). "Formal properties of 'now'", *Theoria* 37: 227-73.

Kaplan, David (1989a). "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals", in J. Almog, J. Perry & H. Wettstein (eds.), *Themes from Kaplan* (New York & Oxford: Oxford University Press): 481-563.
—. (1989b). "Afterthoughts", in J. Almog, J. Perry & H. Wettstein (eds.), *Themes from Kaplan* (New York & Oxford: Oxford University Press): 565-614.

Kuhn, Thomas S. (1957). *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought* (Cambridge, MA & London: Harvard University Press).
—. (1962). *The Structure of Scientific Revolutions*, third edition (Chicago & London: University of Chicago Press, 1996).
—. (1978). *Black-Body Theory and the Quantum Discontinuity, 1894-1912* (Chicago & London: University of Chicago Press).
—. (1989). "Possible Worlds in History of Science", in Sture Allén, (ed.), *Possible Worlds in Humanities, Arts and Sciences: Proceedings of the Nobel*

Symposium 65, *Research in Text Theory* 14 (Berlin: Walter de Gruyter): 9-32, reprinted in *The Road Since 'Structure': Philosophical Essays, 1970-1993, with an Autobiographical Interview*, edited by James Conant & John Haugeland (Chicago & London: University of Chicago Press, 2000): 58-89.
—. (1990). "Dubbing and Redubbing: The Vulnerability of Rigid Designation", in C. Wade Savage (ed.), *Scientific Theories, Minnesota Studies in the Philosophy of Science* 14 (Minneapolis: University of Minnesota Press): 298-318.
—. (2000). "A Discussion with Thomas S. Kuhn", in *The Road Since 'Structure': Philosophical Essays, 1970-1993, with an Autobiographical Interview*, edited by James Conant & John Haugeland (Chicago & London: University of Chicago Press, 2000): 255-323.

Lewis, David (1996). "Elusive Knowledge", *Australasian Journal of Philosophy* 74: 549-567.

Ludlow, Peter (1995). "Social Externalism, Self-Knowledge, and Memory", *Analysis* 55: 1578-159, reprinted in Peter Ludlow & Nora Martin (eds.), *Externalism and Self-Knowledge* (Stanford: CSLI Publications, 1998): 307-310.
—. (1996). "Social Externalism and Memory: A Problem?", *Acta Analytica* 14, reprinted in Peter Ludlow & Nora Martin (eds.), *Externalism and Self-Knowledge* (Stanford: CSLI Publications, 1998): 311-317.
—. (2004). "What Was I Thinking? Social Externalism and Shifting Memory Targets", in R. Shantz (ed.), *The Externalist Challenge* (Berlin: De Gruyter): 419-426.

Mates, Benson (1953). *Stoic Logic* (Berkeley & Los Angeles: University of California Press).

McIntyre, Alasdair (1981). *After Virtue: A Study in Moral Theory* (Notre Dame: University of Notre Dame Press).

McTaggart, John Ellis (1927). *The Nature of Existence,* volume 2, ed. C.D. Broad (Cambridge: Cambridge University Press).

Millikan, Ruth G. (2009). "Embedded Rationality", in Philip Robbins & Murat Aydede (eds.), *The Cambridge Handbook of Situated Cognition* (New York: Cambridge University Press): 171-181.

Nagel, Thomas (1976). "Moral Luck", *Proceedings of the Aristotelian Society*, supplementary volume 50, reprinted in *Mortal Questions* (Cambridge: Cambridge University Press, 1979): 24-38.

Orwell, George (1943). "Looking Back on the Spanish War", *New Road*, 1943, reprinted in Sonia Orwell & Ian Angus (eds.), *The Collected Essays, Journalism and Letters*, volume 2 (Boston: Nonpareil Books, 2000): 249-267.
—. (1944). "As I Please", *Tribune*, 04 February 1944, reprinted in Sonia Orwell & Ian Angus (eds.), *The Collected Essays, Journalism and Letters*, volume 3 (Boston: Nonpareil Books, 2000): 87-9.

Prior, Arthur N. (1957). *Time and Modality* (Oxford: Oxford University Press).
—. (1962). "Changes in Events and Changes in Things" (Lawrence: University of Kansas Press), reprinted in *Papers on Time and Tense* (Oxford: Oxford University Press, 2003): 6-19.
—. (1967a). *Past, Present, and Future* (Oxford: Oxford University Press).
—. (1967b). "On spurious egocentricity", *Philosophy* 42: 326-35, reprinted in *Papers on Time and Tense* (Oxford: Oxford University Press, 2003): 27-37.
—. (1968). "'Now'", *Noûs* 2: 101-119, reprinted in *Papers on Time and Tense* (Oxford: Oxford University Press, 2003):171-93.
—. (1972). "The Notion of the Present", in J.T. Fraser, F.C. Haber & G.H. Müller (eds.), *The Study of Time: Proceedings of the First Conference of the International Society for the Study of Time, Oberwolfach (Black Forest) – West Germany* (Berlin, Heidelberg, New York: Springer Verlag): 320-323.
—. (2004). "A Statement of Temporal Realism", in Jack Copeland (ed.), *Logic and Reality: Studies on the Legacy of Arthur Prior* (Oxford: Clarendon Press): 45-6.

Quine, W.V. (1960) *Word and Object* (Cambridge MA: The M.I.T Press).

Recanati, François (2012). *Mental Files* (Oxford: Oxford University Press).

Richard, Mark (1981). 'Temporalism and Eternalism', *Philosophical Studies* 39: 1-13

Russell, Bertrand (1903). *The Principles of Mathematics* (London: George Allen & Unwin, 1964).

Salmon, Nathan (1989). "Tense and Singular Propositions", in J. Almog, J. Perry & H. Wettstein (eds.), *Themes from Kaplan* (New York: Oxford University Press): 331-92.

Schacter, Daniel L. (1995). *Memory Distortion: How Minds, Brains and Societies Reconstruct the Past* (Cambridge, MA: Harvard University Press).

Schaffer, Jonathan (2012). "Necessitarian Propositions", *Synthese* 189: 119-162.

Schiffer, Stephen (1992). 'Boghossian on Externalism and Inference', *Philosophical Issues* 2: 29-38

Shoemaker, Sidney (1988). "On Kowing One's Own Mind", *Philosophical Perspectives* 4: 187-214.

Sorensen, Roy A. (1998). "Logical Luck", *The Philosophical Quarterly* 48: 319-334

Sosa, David (2005). "The Inference that Leaves Something to Chance", in Sanford Goldberg (ed.), *Internalism and Externalism in Semantics and Epistemology* (Oxford: Oxford University Press): 219-234.

Stevenson, Robert Louis (1879). *Travels with a Donkey in the Cevennes* (Teddington, Middlesex: The Echo Library, 2006).

Thomson, Judith Jarvis (1971). "The Time of a Killing", *Journal of Philosophy* 68: 115-132.

Williams, Bernard (1976). "Moral Luck", *Proceedings of the Aristotelian Society* supplementary volume 50: 115-135, reprinted in *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981): 20-39.

Williamson, Timothy (2000). *Knowledge and Its Limits* (Oxford: Oxford University Press).

Wittgenstein, Ludwig (1969). *On Certainty*, ed. G.E.M. Anscombe & G.H. von Wright, translated by Denis Paul & G.E.M. Anscombe (Oxford: Blackwell).

Wright, Crispin (1986). "Anti-Realism, Timeless Truth and *Nineteen Eighty-Four*", in *Realism, Meaning and Truth* (Oxford: Basil Blackwell): 176-203.

Yourgrau, Palle (2005). *A World Without Time: The Forgotten Legacy of Gödel and Einstein* (New York: Basic Books).