LOGICAL EXPLORATIONS IN THE PHILOSOPHY OF MIND

Essays on Aristotle's psychology, Wittgenstein's *Tractatus*, and Lévy-Bruhl's Logical Relativism

Logische verkenningen in de filosofie van de geest
Essays over de psychologie van Aristoteles, de *Tractatus* van Wittgenstein en het logisch relativisme van Lévy-Bruhl

Proefschrift ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus Prof. dr C. J. Rijnvos
en volgens besluit van het College van Dekanen.
De openbare verdediging zal plaatsvinden op
donderdag 19 maart 1992 om 16.00 uur

door

Gerrit Jan Cornelis Lokhorst

geboren te Gouda

Promotie-commissie

Promotores:

- Prof. dr J. Sperna Weiland
- Prof. dr G. Sundholm

Overige leden:

- Prof. S. J. Doorman, M. Sc.
- Prof. dr T. A. F. Kuipers

The original thesis was written in WordPerfect. This version was prepared with the LaTeX2e package. As a result, the page numbers are completely different.

Voor mijn vader en moeder
en voor Marjolein en Sebastiaan

Contents

Introduction

1. Topics to be discussed

In the essays collected here we have tried to shed some light on the following topics from the philosophy of mind:

1. Aristotle's account of "reflective awareness", the phenomenon which occurs when we "perceive" (as Aristotle put it) that we perceive that something is the case. (Chapter 1.)
2. Wittgenstein's early views on mental representation. (Chapters 2–4.)
   (a) Wittgenstein's "language of thought" theory, according to which the "empirical" self, subject or soul is a collection of sentence-like entities. (Chapters 2 and 4.)
   (b) Wittgenstein's analysis of propositional attitude ascriptions[1], according to which such ascriptions are similar to ascriptions of senses to sentences[2], but do not have the *Unsinnigkeit* of the latter: they are ordinary truth-functional sentences of the language instead. (Chapters 2 and 3.)
   (c) Wittgenstein's claim that the "metaphysical" or "transcendental" self, subject or soul (as distinct from the "psychological" or "empirical" self, subject or soul) does *not* represent and does not do anything else either. (Chapter 4.)
3. Lévy-Bruhl's "logical relativism", i.e., the thesis that people from different cultures may adhere to different "logics", may have their own idiosyncratic standards of "rationality", and may live in different "cognizable worlds". (Chapter 5.)

We have added a short technical paper (Chapter 6) in which two logical systems are studied which are similar to the systems from Chapter 5. Like the latter ones, they are paraconsistent—inconsistency-tolerating—systems of modal logic.

2. Why these topics?

The subjects we have mentioned are interesting and important and have accordingly received a great deal of attention in the past. However, all previous discussions of them are unsatisfactory in one way or another.

First, the texts of the philosophers raise several exegetical problems which have never been solved.

1. Aristotle's discussion of reflective awareness in *De anima* III.2 is generally acknowledged to be "very difficult".[3] First, it is not entirely clear which phenomenon Aristotle is discussing. It has usually been asserted that his remarks are concerned with reflective awareness, "the awareness on our

[1] Propositional attitude ascriptions (the term is Russell's) are expressions like "A believes that p", "A knows that p", "A thinks that p", "A perceives that p", "A fears that p", and so on. (Here and in the following, "A" is the name of an agent and "p" is an abbreviation of a sentence.)

[2] An example: "'the sun is shining' says that the sun is shining".

[3] Schiller (1975, p. 294.).

parts that we are, when seeing or hearing, for example, seeing or hearing".[4] However, this interpretation has recently been called into question.[5] Secondly, Aristotle's arguments concerning his topic (whatever it is) are "neither straightforward nor clear".[6] At least one commentator has concluded that "there seems no way of making the argument coherent".[7] Thirdly, there seems to be a glaring conflict between *De anima* III.2 and *De somno* II. Nobody has been able to reconcile the two accounts. Eminent Aristotelians like Aquinas and Brentano took the easiest way out and simply ignored one of the two passages. (They did not ignore the same one, though.) Others have tried to reconcile them but failed. The remaining commentators have tried to conceal their lack of understanding behind vagueness.[8]

2. Wittgenstein made some notoriously unclear remarks about mental representation in his early writings. These remarks have even been judged to be of "almost impenetrable obscurity".[9] However, the following seems certain. Wittgenstein saw the "empirical self, soul or subject" as a collection of thoughts which have the same structure as sentences. When making a propositional attitude ascription such as "A thinks that p", we assert that one of the thoughts belonging to A's "empirical self" says that p (i.e., expresses the proposition that p). Because a thought is similar to a sentence, the latter part of this assertion (thought t says that p) is similar to a statement of the form "sentence 'p' says that p".

   It is at this point that the problems begin. For according to Wittgenstein, "'p' says that p" is *unsinnig*. One cannot *say* that "p" says that p; this can only be *shown*. "'p' says that p" is a would-be sentence. It is not well-formed in the language in which "p" is stated ("the only language there is") and does not have a truth-value. Does the same apply to "A thinks that p"? Most commentators have assumed so.[10] But the resulting view is counterintuitive; and it would be rather surprising that Wittgenstein did not explicitly mention such a striking consequence of his views, if it did indeed follow from them. As a result, this interpretation is not really plausible. A theory which would explain why "A thinks that p" is similar to "'p' says that p" but which would not imply that propositional attitude ascriptions are *unsinnig* would be more attractive, but no one has come up with such a theory thus far. The main stumbling-block is Wittgenstein's principle of truth-functionality, which states that the truth-values of all sentences are completely determined by the truth-values of the "elementary sentences" which describe the world. It is hard to see how this thesis could apply to propositional attitude ascriptions.

   The topic with which Chapter 4 is concerned (the "metaphysical" subject) has never been properly understood either. The catalogue of conflicting interpretations at the beginning of this chapter shows that it has only caused a great deal of confusion.

3. Aristotle and Wittgenstein have a reputation for unclarity. As is to be expected, Lévy-Bruhl's writings pose fewer problems of a purely exegetical nature. Nevertheless, it is not altogether obvious what he may have meant

[4] This formulation is due to Kosman (1975, p. 499).
[5] Kosman (1975), Osborne (1983).
[6] Kosman (1975, p. 500).
[7] Hamlyn (1968, pp. 121–123).
[8] See Chapter 1 for precise references.
[9] Urmson (1956, p. 133).
[10] Including myself: I defended this interpretation in the (fortunately unpublished) Lokhorst (1985 *b*), although I had already rejected it in Lokhorst (1985 *a*).

by saying that "other people follow different logics", and it may be doubted whether everyone interprets this thesis in the same way.

Secondly, most previous expositions of the topics we have mentioned are unsatisfactory because they are vague and informal.

1. Most commentaries on Aristotle consist of paraphrases and quotations instead of lucid explanations.
2. Stegmüller once described a certain commentary on Wittgenstein's *Tractatus* as "eine Ansammlung von sehr undeutlichen Sätzen, die ihrerseits erst expliziert werden müßten", and this observation may be generalized to most writings on that work.[11]
3. Lévy-Bruhl's theory has never been formulated in a truly precise way either.

Thirdly, it is seldom pointed out that the theories which we have mentioned are as interesting and important today as when they were first proposed. (The biographical fact that both Wittgenstein and Lévy-Bruhl repudiated their own early views later on in their lives cannot, of course, be seen as a *refutation* of these views, nor does it make them less interesting.)

1. Aristotle's theory was adopted by Brentano and many philosophers and psychologists in his wake, but the fallacy he warned against is still often being made today: many people (e.g., Richard Dawkins, Keith Gunderson and Rudy Rucker, to mention but a few) claim that unlimited self-consciousness is out of reach for finite beings because it necessarily involves an infinite regress of mental phenomena.
2. It is not often recognized that Wittgenstein held a "language of thought" theory of the type which is so popular today. His analysis of propositional attitude ascriptions is, in fact, the same as the "sentence storage model" (or "belief box model") of such ascriptions, which Stalnaker has recently described as "perhaps common enough to be called the received view" today.[12] It has not been widely noticed either that Wittgenstein's truth-functional theory of propositional attitude ascriptions amounts to a ("semantic") *supervenience* theory of the sort which is receiving so much attention nowadays.[13]
3. Lévy-Bruhl's thesis of "logical relativism" has, of course, not been forgotten; there is still much controversy about it today.

The just-mentioned three circumstances have made us decide to study the above three topics afresh. We think we have been able to make some progress. We have solved the previously unsolved exegetical problems we have mentioned; we have expressed our ideas in the most precise way which is humanly possible (and perhaps non-humanly as well), namely in the form of formal axiomatic theories; and we have continually applied the ancient philosophical texts to modern discussions in the philosophy of mind (and sometimes even in the field of *artificial* intelligence).

Actually, we have done more than just elucidating the topics we have mentioned. Aristotle's remarks on reflective awareness cannot be understood in isolation from the rest of his psychological views, so we have discussed and formalized

[11] Stegmüller (1966, p. 195).

[12] Stalnaker (1991, p. 430). Stalnaker explains the belief box model as follows (p. 431):

> When we say that x explicitly believes that P, we say (on the storage model) that x stores some sentence that says that P. So a sentence of the form "x believes that P" makes an existential claim about x's beliefs: that there exists a sentence in x's belief box that says that P.

When "belief box" is replaced by "empirical self", and "belief" by "thought", we obtain the Wittgensteinian theory which has been described at the beginning of the present Section. The theory which is proposed in Chapters 2–4 corresponds, to be more precise, to a belief box model with a "coarse-grained conception of content" in the sense of Stalnaker (1991, p. 432).

[13] See, e.g., Lokhorst (1988 *b*).

the latter too, at least in so far as this was necessary. Similarly, Wittgenstein's views on mental representation cannot be understood in isolation from his basic ontological and semantic doctrines, so we have presented formal reconstructions of these doctrines too.[14]

3. Our approach to our subjects

Our studies differ from most previous commentaries in that we have continually tried to transform the insights of the philosophers we have been studying into formal axiomatic theories. The results are up to all standards of clarity and exactness which are common in the field of modern mathematical logic. We have taken this trouble for several reasons.

First, formalization is good for one's mental discipline. It is all too easy to make logical blunders and to be unaware of one's tacit assumptions as long as one sticks to the messy apparatus of ordinary language. We will repeatedly have occasion to point this out in the following chapters.

Secondly, formalization leads to a degree of objectivity which cannot be achieved in any other way. All informal theories ultimately require more or less *einfühlendes Verstehen* (empathetic "understanding") on the part of the reader. This makes it difficult to compare different theories, to check them against the textual evidence, to apply them to fields for which they were not originally intended, to determine whether they are consistent, and so on. In short, one does not usually arrive at theories which are able to stand on their own feet. Formalization can never have these defects: it always leads to theories whose consequences anyone can calculate for himself.

Thirdly, formalization makes one's theories accessible to a vast audience. The language of formal logic is the *lingua franca* of disciplines as diverse as mathematics, computer science, linguistics and philosophy. It is conceivable that workers from these disciplines may still profit from the insights our philosophers had once these have been reformulated in a language which they can understand. This is more than a theoretical possibility: both Chapter 1 and Chapter 5 contain some references to logical systems which are used in the design of "artificially intelligent" expert systems. If we had expressed our theories in, say, Dutch or Swahili—or even in some artificial language like Volapük or Solrésol—we would never have spotted the connection.[15]

In short, when one wants to blow the dust off ancient theories and tries to polish them up until they shine again with the same splendour which they had when they were first proposed, formalization is the best way to go. Formalization is not always possible, to be sure: some theories are so hopelessly "illogical" that one does not even know where to start.[16] But whenever it does look possible, one should certainly try it out. Even Lévy-Bruhl's theory—which may look like the very antithesis of logic at first sight—turns out to become clearer upon logical analysis.

[14] On the other hand, we have, in a sense, also done less than has just been announced, since not all chapters contain formalizations of the theories they are concerned with. Chapters 3 and 4 are less formal than the others because they are extensions of Chapter 2 and presuppose the formal account put forward there. Chapter 5 does not, strictly speaking, present a formalization of Lévy-Bruhl's thesis. It rather gives an infinite number of illustrations of the phenomena adumbrated by Lévy-Bruhl.

[15] Lincos (language for cosmic intercourse) (Freudenthal 1960) might be even more "objective" than the language of logic. If we had wanted to send Aristotle's, Wittgenstein's and Lévy-Bruhl's theories to the stars we would have used it, but logic is more widely known on Earth.

[16] As I have argued in Lokhorst (1989), it is impossible to formalize Hegelian dialectics. Even the inconsistency-tolerating systems from Chapters 5 and 6 cannot be used for this task.

The logical approach to philosophical texts is, in fact, the same as the scientific approach towards nature. Scientists try to devise simple theories which account for the facts, which make definite predictions, and which may therefore be refuted on the basis of further empirical data. In the same way, we try to devise simple theories which account for the statements of the philosophers, which have easily calculable consequences, and which may therefore be falsified by textual evidence we have not considered. We do not have to assume that the philosophers themselves were fully aware of these theories, just as we do not have to assume that nature—or the Author of the Book of Nature, if any—knows the laws which it obeys. (This takes care of the accusation of anachronism which may possibly be levelled against our analyses.)[17]

The formal approach to philosophical texts has another benefit: it gives us an insight into the strengths and weaknesses of the theories which have so far been developed within the field of philosophical logic. As soon as one starts using these theories, their capacities and shortcomings emerge—with the result that one may decide to improve on them in specific ways. The following chapters contain numerous examples of this phenomenon.

Thus, philosophy is not the only discipline which profits from the application of logic to philosophy: this activity is stimulating for logic as well. This phenomenon is familiar enough from mathematics, which too has often benefitted from efforts to apply it. Both philosophy and logic suffer when they ignore each other, and it should not be thought that logic suffers less: as Urquhart has recently noted, "the contemporary scene in philosophical logic often looks like a procession of monkeys holding each others' tails".[18] Urquhart took his inspiration from Russell's writings in order to break out of the circle. We have studied thinkers who are even farther removed from logic to reach the same goal.

4. Preview of the logical systems

Which kind of logical systems are we going to use? Which branch of logic is particularly relevant to the philosophy of mind? The philosophical theories we will examine below are either stated in terms of propositional attitude ascriptions or theories about such ascriptions, so it should come as no surprise that we will mainly be concerned with the logic of propositional attitude ascriptions. Even within this restricted domain of logic, many approaches exist today. Our analyses are in the tradition of the "modal" approach to the propositional attitudes which started with Hintikka's *Logic and Belief*, now exactly thirty years ago.[19]

Hintikka noted that certain propositional attitude ascriptions have the same logical properties as certain modal sentences. For example, modal logicians have generally assumed that if it is necessary that p and q is a logical consequence of p, then it is also necessary that q. If "it is necessary that" is replaced by "A believes that", we obtain a plausible principle for belief sentences—provided, that is, that A is fully rational ("logically omniscient") or, alternatively, that we are using a concept of fully rational belief. Other expressions which are used to indicate this concept are "implicit belief" or "commitment to believe".[20] If a rational agent believes that p and q follows logically from p, then this agent will also believe that

[17] Nevertheless, the philosophers whom we will discuss sometimes held views which are really closer to our formalizations than their informal writings would suggest. In Chapter 2, we will reconstruct Wittgenstein's situation ontology as a Boolean algebra of parts and wholes. Wittgenstein seems to have had the same idea himself, even though he did not explicitly say so. (See the Appendix to Chapter 2.)

[18] Urquhart (1991).

[19] Hintikka (1962); Lenzen (1980) is a more recent introduction to this area.

[20] See Stalnaker (1991) for a more explicit discussion of these concepts.

q; an agent who does not explicitly have the latter belief is, in any case, implicitly committed to it.

There are many principles apart from the just-mentioned one that might conceivably be added to both systems of modality and belief; for example, "if $\Box$ (p and q), then both $\Box\ p$ and $\Box\ q$", where "$\Box$" stands either for "it is necessary that" or "A believes that". The cases of modality and belief might, of course, also diverge at certain points; furthermore, it might turn out that some principles are acceptable for one of the propositional attitudes (e.g., rational thinking), but not for others (e.g., perceiving). Principles involving quantification are particularly controversial in all cases. To give but one example: is "if $\forall\ x\ \Box\ Px$, then $\Box\ \forall\ x\ Px$", where "$\forall$" stands for "for all", and "P" for any predicate, generally acceptable or is it not? (It is not.)

No matter which principles one might find acceptable, Hintikka's modal view of propositional attitude ascriptions made it at any rate possible to think much more clearly about the issues. This was particularly due to the fact that the standard "possible worlds" semantic analysis of modal logic could now be applied to these ascriptions. As a result, they could now be examined from a point of view which had not been available before. Principles like the just-mentioned quantificational one turned out to correspond with definite semantic conditions, whose plausibility or implausibility could be examined in their own right; the standard technical instruments of modal logic could be used to axiomatize notions of validity; effects of various model-theoretic assumptions could be calculated through; it became possible to check whether theories stated in terms of propositional attitude ascriptions are consistent, and so on. In short, "a whole dimension of systematicity was, as it were, imported for free".[21]

Some theories in the philosophy of mind are expressible in terms of nothing but propositional attitude ascriptions, and all of a sudden these theories became amenable to formal study too. Interesting areas of the philosophy of mind could now be given a mathematical appearance with comparatively little effort.[22] It is theories of the latter type which are studied in the following essays. We will sometimes focus our attention on the question whether certain principles are acceptable (for example, in Chapter 1), whereas we will follow the semantic route at other times (for example, in Chapter 2).

Although our general approach follows the long and venerable tradition which has just been sketched, the particular systems we propose are new. The first chapter presents several systems which are not only "non-normal", but also allow quantification over agents (the agents being sensory organs).[23] The second chapter introduces non-normal systems in which there is no distinction between names and predicates (predicates are names too) and in which there are countably infinite disjunctions and conjunctions. Chapter 5 illustrates Lévy-Bruhl's ideas by means of infinitely many non-classical multi-modal systems. Chapter 6 discusses two related non-classical modal systems.

The systems in Chapters 5 and 6 probably represent the most drastic departures from Hintikka's approach.[24] In these chapters, we relativize the basic assumption of rationality ("logical omniscience") which is built into Hintikka's conception. What is a rational agent? To what does one commit oneself when one believes that p? It all depends on the logical standards one applies. Hintikka considered just one

[21] To adapt an expression of Fodor's paraphrased by Churchland *née* Smith (1986, p. 467).

[22] See, for example, Bacon's formalization of Berkeley's theory of perception (Bacon 1979), and Hintikka's discussion of sense data theories (Hintikka 1969).

[23] Precise definitions of "normal" and "classical" systems are to be found in Chellas (1980).

[24] Chapter 6 is, strictly speaking, not concerned with doxastic logic (the logic of belief), but in view of what has been said above, it is easy to apply it to this area.

logic, classical logic, but many alternatives to it have been proposed during the last hundred years. We have accordingly found it desirable to study alternative types of rational belief. In Chapters 5 and 6, we discuss systems which are based on so-called "paraconsistent" or "inconsistency-tolerating" logics. Since the chapters presuppose some familiarity with such systems, let us briefly describe them here.

Many theories we encounter both in our daily lives and in our theoretical work are inconsistent. Various kinds of collections of data about the real world are manifestly inconsistent, Frege's set theory is inconsistent in a less conspicuous way, and Hegel even wrote several books in praise of inconsistency. (I do not know whether he was consistent himself.) Classical logic makes us helpless in the face of inconsistent theories, because it licenses us to infer any sentence "q" whatever from a set of premises of the form {"p", "not p"}. We are not so helpless in practice (we may, for example, decide to ignore one of the premises), and it would be desirable if "expert systems" reasoning about the world were able to handle inconsistencies in some sensible way too. Many logicians have therefore tried to devise "paraconsistent" systems which are free from the just-mentioned rule.

There are countless ways in which one may proceed in order to render (some or all) inconsistencies harmless (without giving up all logical principles). For example, one may introduce a third truth-value (standing for being both true and false), and revise the truth-tables for the connectives in an appropriate way. This is the strategy which has been followed in the systems discussed in Chapter 6. Alternatively, one may retain bivalence but alter the truth-conditions for negation; this is the approach adopted in the Da Costa systems examined in Chapter 5. The basic idea on which these systems are based is that although "not p" should certainly be true if "p" is false, it may as well be true as false if "p" is true. The truth-conditions for all other connectives remain the same. It is clear that this has the consequence that it may be the case that both "p" and "not p" are true whereas some other sentence "q" is nevertheless false. In other words, contradictions no longer lead to triviality (provability of all sentences in the language). Apart from the two approaches which have just been sketched, there are many more possibilities.[25]

The just-mentioned alternatives to classical logic do not contain modal or propositional attitude operators. When we add such operators and specify their truth-conditions, several questions arise. Some of them are of a technical nature. For example, what is the modal status of provable (valid) contradictions in the new modal systems? How should one axiomatize the new notions of validity? Chapter 6 presents some examples of answers to such questions.

On the other hand, there are new philosophical applications in sight. For example, a logic which does not have the theorem "if A believes (p and not p), then A believes that q" may conceivably be used to study the beliefs of agents who are inconsistent but yet do not believe everything. Such a system would describe the beliefs of an agent who is not ideally rational according to classical logic, but according to the paraconsistent logic under consideration; or to put it differently, it would capture an alternative, non-classical notion of "implicit" belief or "commitment to believe". In Chapter 5, we study systems in which there is an infinite number of such alternative operators occurring side by side. The systems represent situations in which believers of many logical denominations try to describe each others' beliefs as sympathetically as possible. Each agent A describes each agent B's beliefs "in B's own terms": if B believes that p, and q follows from p according to A's logic but not according to B's, then the belief that q is not imputed to B. Nor is it being said that B has a commitment to believe that q; he

[25] See da Costa & Marconi (1987) for a fairly recent survey of paraconsistent logic; see also Lokhorst (1989).

has not committed himself to the latter belief by his own lights. (Are intuitionists committed to believe that p or not p whenever they believe that p implies p? Not according to themselves.) By showing that it is logically possible to put oneself in someone else's "logical shoes" (no matter how different the other agent's logic might be), we are, as it were, taking the sting out of logical relativism: the situations the logical relativists have in mind are describable within logic itself.

The worlds which are used in the semantic study of "deviant belief" are no longer logically possible according to classical logic: they are "impossible" or "imaginary" instead (at least according to classical logic). However, they are still possible according to the alternative logics we are considering. It is possible to deviate even more from the Hintikka tradition: one may introduce thoroughly anomalous worlds in which the truth-values of sentences are no longer determined by the truth-values of their subsentences. When one does so, notions of belief may be studied which are no longer idealized in any way. Systems in which this is possible have received some attention of late.[26]

We have had no need to stray so far from the well-trodden path since we have only been interested in idealized propositional attitudes. Chapter 1 is concerned with "perfect perception", Chapters 2 to 4 with rational thinking, Chapters 5 with perfect adherence to different logics, and Chapter 6 with systems which are interpretable as antinomic logics of rational belief. Thus, the criticism which has most frequently been made of Hintikka's approach, namely that it is unduly idealized, does not affect our enterprises in the least: it is precisely this idealization which makes it ideally suited to our purposes. Or to put it the other way round: we have taken care to select topics which demand just this kind of idealization.

May this suffice as a warming-up. The proof of the pudding is in the eating: let us quickly proceed with the real work.

[26] See, e.g., Fagin & Halpern (1988), Wansing (1990).

CHAPTER 1

Aristotle on "Perceiving that we See and Hear": A Study in the Logic of Perception

Abstract

In *De anima* III.2 and *De somno* II, Aristotle gives two intriguing—but apparently conflicting—descriptions of "reflective awareness", the phenomenon which occurs when we "perceive" that we perceive. In this paper, we show (1) that it is possible to reconcile both accounts with one another; (2) that the resulting theory is still interesting from a modern point of view; (3) that this theory can be axiomatized as a certain provably consistent system of perceptual logic. In sum, Aristotle's theory is less muddle-headed than many modern commentators suppose.

0. Introduction

In chapter III.2 of Aristotle's *De anima* (*DA* 425b12 ff.), we encounter the following intriguing remarks about the phenomenon that "we perceive that we see and hear":

(*a*) Since we perceive that we see and hear [*aisthanometha hoti horōmen kai akouomen*], it must either be by sight [*tēi opsei*] that one perceives that one sees or by another sense [*heterai*].

(*b*) But in that case there will be the same sense for sight and for the colour which is the subject for sight. So that either there will be two senses for the same or the sense itself will be the one for itself.

(*c*) Again, if the sense [*aisthēsis*] concerned with sight were indeed different from sight, either there will be an infinite regress, or [again] there will be some sense which is concerned with itself.

(*d*) So that we had best admit this of the first in the series [*tēs prōtēs*].

(*e*) But this presents a difficulty: for if (i) to perceive by sight is to see [*tēi opsei aisthanesthai estin horan*], and (ii) if one is to see colour or that which possesses colour, then (iii) if one is to see that which sees, then (iv) that which sees primarily [*to horōn*] will have colour.

(*f*) It is clear, then, that to perceive by sight is not a single thing; for even when it is not the case that we see, it is by sight that we judge both darkness and light, though not in the same way.

(*g*) Moreover, even that which sees [*to horōn*] is in a way coloured; for each sense organ [*aisthētērion*] is receptive of the object of perception without its matter.[1]

[1] Translation by Hamlyn (1968), with some minor stylistic changes. All translations we have consulted amount to the same, with the exception of the translation by J.A. Smith in Barnes (1984, vol. 1, p. 677). He reads (a) as "Since it is through sense that we are aware that we are seeing or perceiving ... ", and so on. We think this is a rather *ad hoc* translation of *aisthanometha*. But it is a good *interpretation*, which we will defend ourselves in Section 1 below.

The above passage is intriguing for many reasons. First, what exactly is the phenomenon Aristotle is talking about? Most commentators think that the passage is concerned with some sort of "reflective awareness" or "perceptual self-consciousness"; Ross, for example, has described it as "one of the earliest discussions, in any author, of the difficulties involved in self-consciousness".[2] However, this "standard" interpretation has recently come under attack.[3]

Secondly, it is not entirely clear what Aristotle wanted to say about the phenomenon he was considering (whatever it was). The *DA* passage gives the impression that Aristotle wanted to make the point that reflective awareness of seeing is provided by the sense of sight itself. It is, however, doubtful whether this interpretation is correct, for Aristotle seems to be denying this in the following passage from the *De somno* (*DS* 455a13 ff.):

> Each sense [*aisthēsis*] possesses something which is special [*idion*] and something which is common [*koinon*]. Special to sight, for example, is to see, special to hearing is to hear, and similarly for each of the others. But there is also a common power [*koinē dunamis*] which accompanies them all, in virtue of which one perceives that one sees or hears. For it is not by sight, after all, that one sees that one sees; nor is it by taste or by sight or by both that one judges, and is capable of judging, that sweet things are different from white ones; but it is by some part which is common to all the sense organs [*tini koinōi moriōi tōn aisthētēriōn hapantōn*].[4]

It is too easy to say that Aristotle changed his mind in the course of time; it is better to try to come up with a theory which reconciles both passages with one another.

Thirdly, Aristotle's suggestion in the *DA* passage that unlimited reflective awareness (total self-consciousness) does not necessarily involve an infinite regress of mental phenomena (if this is indeed what he wanted to suggest) is exciting in itself, for even today many people—ranging from biologists to mathematicians, as we will see—still suppose that it necessarily involves such a regress. As a result, it is either claimed that we *cannot* be totally self-conscious (our finiteness is the bottle-neck), or that we are infinite (since we are self-conscious to an unlimited extent). Aristotle's remarks point to a third possibility.

A considerable number of commentaries have already been written on the *DA* and *DS* passages we have quoted. The debate started in the fourth century (Themistius), continued in the middle ages (Aquinas), saw a revival in the nineteenth century (Brentano and others), and has never stopped since. Yet no commentator has ever given an interpretation which is satisfactory in all respects. In this paper, we will present an exegesis which (i) reconciles the *DA* and *DS* accounts with one another, (ii) gives a logical reconstruction of the arguments in both works, (iii) explains where previous commentators went astray, and (iv) shows how Aristotle's insights may be applied to present-day discussions.

We will present our exegetical proposals in a rather unusual way: in the course of our discussion we will gradually develop a formal axiomatic theory which captures the points we want to make. We think that this is both instructive in itself and leads to a much greater clarity than can be achieved when one is using ordinary language. Thus, the present paper is intended as a first step towards an axiomatization of an interesting part of Aristotelian psychology.

---

[2] Ross (1949, p. 141); Ross (1961, p. 35). The *DA* discussion is a reaction to Plato's *Charmides* (167-169), but Plato discusses the topic in less detail.

[3] Notably by Kosman (1975) and Osborne (1983).

[4] Translation by Kahn (1966, p. 59), with some minor stylistic changes.

1. What was Aristotle talking about?

Let us first try to determine which subject Aristotle was talking about. As we have indicated, the *DA* and *DS* passages have traditionally been regarded as discussions about the phenomenon of "reflective awareness". This phenomenon occurs in situations like the following. Suppose that you see that the moon is shining. Then you may realize that this is what you are doing—seeing that the moon is shining. This is "reflective awareness". The same phenomenon may occur in connection with the activity of the other senses. When you are hearing that the wind is blowing, you may be aware of the fact that you are hearing this, and so on. As Kosman amusingly points out,

> Such reflective awareness is not restricted to philosophers or psychologists. Tourists and pilgrims, for example, are often as aware of the fact *that* they are witnessing the goal of their peregrinations—the fact *that* they are, for example, seeing Philadelphia—as they are of the actual object of their vision—that is, Philadelphia. Expressions of the form, "See Philadelphia and die", lead us to value, so to speak, having the experience more than the experience, that is *having seen* Philadelphia more than seeing Philadelphia, and thus in the middle of the experience to become more aware of the fact *that* we are seeing Philadelphia than of the Philadelphia we are seeing. This is called being self-conscious.[5]

Kosman regards it as "strange" that Aristotle said that we "perceive" that we see and hear. As a result, he doubts whether Aristotle was really talking about reflective awareness at all. We think there is little reason for such doubt. Aristotle said that we "perceive" that we see because the Greek language of his time did not have special terms for general concepts like "consciousness", "awareness" and "introspection".[6] So he had no choice but to use a perceptual verb in a wide sense when wanting to describe reflective awareness. Given his limited vocabulary, he expressed himself as clearly as he could. Perceptual expressions are still used in the same way today. For example, Thomason's system of perceptual logic similarly contains the axiom "if you see that p, then you see that you see that p".[7] This at least shows that not everyone regards such expressions as odd.

Contrary to what many commentators have asserted,[8] Aristotle certainly did not unconditionally accept the claim that we perceive that we see whenever we see: in the *De insomniis*, dreams are unequivocally described as effects of perceptual phenomena (sensations) which were *not* noticed when awake.[9] Nevertheless, Aristotle did not just propose the claim for the sake of the argument: he was much too sympathetic to it for this to be the case, and certainly took it seriously. This is not only borne out by the *DA* and *DS* passages, but also by the following passages from the *Nicomachean ethics* and the *De sensu*:

> ... and if he who sees perceives that he sees, and he who hears, that he hears, and he who walks, that he walks, and in the case of all other activities similary there is something which perceives that we are active, so that if we perceive, we perceive that we perceive, and if

[5] Kosman (1975, p. 503).

[6] Kahn (1966, p. 70 ff.).

[7] Thomason (1973). Bacon also considers the axiom. He says that it "seems innocuous, but might conceivably be refuted by psychological experiments" (Bacon 1979, p. 280). We will see that there are indeed such experiments.

[8] Brentano (1924, book 2, chapter 2, §10) and Berger (1989) are two examples of philosophers who have mistakenly attributed this view to Aristotle.

[9] *De insomniis* 460b28 ff., 461b22 ff. Cf. Sorabji (1974, p. 71 note 20) and van der Eijk (1991, introduction, section A2).

> we think, that we think; and if to perceive that we perceive or think is to perceive that we exist (for existence was defined as perceiving or thinking); and if ... [10]
>
> For if it is impossible that a person should, while perceiving himself or anything else in continuous time, be at any instant unaware of his own existence, and if there is in the time-continuum a time so small as to be absolutely imperceptible, then it is clear that that a person would, during such time, by unaware of his own existence, as well as of his seeing and perceiving.[11]

Both passages are non-committal, but clearly show that Aristotle was not averse to the thesis that we are reflectively aware. We suggest that he regarded the thesis as an *idealization*. It does not describe actual perceivers, but idealized ones, who are never drowsy, overwhelmed by too many sensory impressions, and so on: it fails to apply to actual persons in so far as these fall short of being such idealized agents. Or to put it differently: the thesis applies to *potential* rather than *actual* perception.[12] This interpretation seems to be in good agreement with the textual data.

Is it true that we are potentially able to perceive that we see whenever we see? Recent psychological experiments suggest that even this weak claim is false. Patients suffering from the so-called "blind sight" syndrome are certainly able to *see*: they are able to point to sources of light, and so on. But they definitely do not *perceive* that they see, not even "potentially" in any interesting sense of this word. They think they are blind, and do not believe anyone who tells them that they are not.[13] The difference between such patients and "normal" people is, in all likelihood, only a matter of degree. We must conclude that Aristotle was *overidealizing* rather than idealizing at the beginning of the *DA* passage.

It goes without saying that Brentano and his followers—e.g., Husserl and Chisholm—made an even greater mistake when claiming (inspired by Aristotle) that we are *actually* aware of all our sensations.[14] They seem to have thought that "introspection" provides us with a full view of our mental lives. The blind sight experiments show that it does not. The "inner eye" is not all-seeing, but rather myopic instead.

Although Aristotle's basic premise is empirically false, this does not make his argument less interesting from a purely logical or conceptual point of view. The reference to an "infinite regress" at (*c*) of the *DA* passage shows that Aristotle thought that there are no limits to (potential) reflective awareness. Just as we may perceive that we see, we may perceive that we perceive that we see, and so on. According to Aristotle, all this perceptual knowledge is provided by sensory organs. Yet we do not have to assume that there infinitely many perceptual organs in order to account for it: for as he pointed out at (*c*)–(*d*), we may simply postulate an instance which "is concerned with itself". Thus, unlimited reflective awareness is not out of reach for agents having only a finite number of sense organs.

As is well-known, the latter idea was adopted by Brentano, became a cornerstone of early phenomenology and has never left the philosophical scene since.[15]

---

[10] *Nicomachean ethics* 1170a28 ff. as translated by W. D. Ross in Barnes (1984, vol. 2, p. 1849).

[11] *De sensu* 448a26 ff. as translated by J.I. Beare in Barnes (1984, vol. 1, p. 711).

[12] We use the terms "potential" and "actual" perception in the sense of Thomason (1973, p. 263).

[13] Weiskrantz (1986).

[14] Brentano (1924, book 2, chapter 2); Chisholm (1981, ch. 7); Berger (1989); Natsoulas (1988), Natsoulas (1989). The latter three articles contain many more references.

[15] See the previous note for references.

Brentano applied Aristotle's idea to sensations instead of senses; but this may not be too great a modification after all, since *aisthēsis* may stand for both concepts. Combining the Aristotelian idea of circular representation with his own idea that each sensation is an object of some sensation, he concluded that all sensations are objects of themselves. All sensations are *en parergoi* "concerned with themselves" whichever other objects they might have; they are, as Ryle has mockingly put it, "self-intimating" or "self-luminous".[16]

Aristotle's idea seems to have made little impact outside phenomenological circles. Even today, it is still frequently being claimed that finite beings cannot represent themselves completely, because this would lead to an infinite regress of representational entities. This misconception is wide-spread; one even encounters it in such unlikely places as *The selfish gene* by the biologist Richard Dawkins.[17] Whereas Dawkins mentions the idea only in passing (as if it were self-evident), others have drawn far-reaching conclusions from it. The philosopher Keith Gunderson, for example, has argued that it may be used to explain our reluctance to accept physicalist mind-body theories: he argues that our finiteness prevents us from seeing ourselves as entities which wholly belong to the physical world. We are, so to say, too small to see ourselves for what we are.[18] The situation would have been different if we had been infinite:

> Robert Nozick suggested that my conclusion would hold only if the mind were finite [...]. A number of us including myself found this a very interesting suggestion, though none of us seemed wholly to understand it.[19]

Mathematician Rudy Rucker (basing himself on the writings of Josiah Royce) has drawn the opposite conclusion from the same fallacious line of thought. According to him, we *can* be totally self-conscious and are therefore infinite:

> Attempts to analyze the phenomenon of consciousness and self-awareness rationally appear to lead to infinite regresses. This seems to indicate that consciousness is essentially infinite.[20]

The *DA* passage may still be used as an antidote against such mysticism.

2. The apparent conflict between the *De anima* and the *De somno*

Although it is easy to get a general idea of what Aristotle was talking about, it is difficult to understand the texts in detail. To begin with, there seems to be a conflict between the *DA* and *DS* passages. It sounds rather contradictory to say, on the one hand, that we perceive by sight that we are seeing, while simultaneously asserting on the other hand that "it is not by sight that one sees that one is seeing ... but by a part common to all the sense-organs".

Previous commentators have reacted in various ways to the apparent conflict. Some simply ignored one of both passages. This was Aquinas' response, for example. Although he wrote a fine commentary on the *DA* passage, he did not mention it in the *Summa theologica* and preferred the *DS* view without giving any justification for his choice.[21] Brentano reacted in a similar way. He accepted the *DS* view at first, but later switched to the *DA* theory without explaining his change of

[16] Ryle (1949, chapter VI, section 2).

[17] Dawkins (1976, p. 63).

[18] Gunderson (1975).

[19] Gunderson (1975, p. 129, note 33).

[20] Rucker (1984, p. 51). He took his inspiration from Royce (1912, Appendix: The one, the many and the infinite (pp. 504–7)).

[21] Aquinas, commentary on the *De anima*, §586 (in Foster & Humphries (1951)); Aquinas, *Summa theologica*, *quæstio* 78, *articulus* 4, *ad* 2; *ibid.*, *quæstio* 87, *articulus* 4, *objectio* 3 and *ad* 3 (Aquinas 1894)

mind.[22] One commentator, Block, has argued that there is a genuine discrepancy: according to him, Aristotle simply changed his views in the course of time.[23] This suggestion has not been accepted by anyone.

The general opinion among the commentators is that there is no real conflict between the passages. But it is seldom indicated how the apparent conflict is to be explained away. One usually gets vague remarks like Sorabji's:

> The *DS*—supplementing, but not, I think, contradicting the *DA*—says that we are aware of our own seeing through the central sense faculty.[24]

This is, of course, totally uninformative. Charlton likewise says that "there is no real conflict" without making himself clear.[25] Kahn has the following verdict:

> On this as on many other points, the most accurate analysis is that given long ago by Rodier (1900, II, p. 266): "ce n'est pas, en effet, en tant que sens spécial et differencié que le sens de la vue nous donne la conscience de la vision; c'est en tant qu'il participe aux caractères communs de toute sensibilité".[26]

This may be true, but does not say much.

Two commentators have made more substantial suggestions. Firstly, Neuhäuser has argued that *tēs prōtēs* (the first part) at (*d*) refers to the organ which the senses have in common.[27] This would imply that the *DA* says exactly the same as the *DS*. We do not think that this interpretation can be justified on the basis of (*c*)–(*d*). As was already recognized by Aquinas,[28] Aristotle was definitely referring to the first member of some (infinite) series of parts of the sensory system. This first element might in fact be identical with the common organ, but it might just as well be some other part of the sensory system, e.g., *to horōn* ("that which sees") of (*e*). As was to be expected, Neuhäuser's interpretation has been rejected by all subsequent commentators. Secondly, Modrak has suggested that:

> Aristotle may . . . have in mind different cases in the two works. In the *DA* he is considering reflective awareness from the perspective of an isolated sense modality; however, in higher animals such as ourselves, more than one sense is active at any given moment. When more than one sense is active, the objective of reflective awareness would be a perception resulting from the convergence of several senses and hence would fall under the common sense as in the *DS*.[29]

This makes sense in itself. However, there is no evidence whatsoever that Aristotle had different cases in mind in the two works. He seems to have been thinking of all cases of reflective awareness in both passages. So we have to find another solution.

3. A solution

We think that the solution is quite simple: all that is needed is some close reading. At (*a*) of the *DA* passage Aristotle presents the following claim:

(A1) If one sees that *p*, then one perceives (by some part of the sensory system) that one sees that *p*.

---

[22] Brentano (1924, book 2, ch. 2, §10, note on pp. 185–186).
[23] Block (1964).
[24] Sorabji (1974, p. 72).
[25] Charlton (1981, p. 109).
[26] Kahn (1966, pp. 56–7).
[27] Neuhäuser (1878, pp. 63–4).
[28] Aquinas, commentary on the *De anima*, §586 (in Foster & Humphries (1951).
[29] Modrak (1987, p. 201).

At (*d*) he reaches the conclusion:

(A2) If one sees that *p*, then one does not hear, smell, feel or taste that one sees that *p*. Rather, one *perceives by sight* that one sees that *p*.

On the other hand, the *DS* claims that:

(A3) If one sees that *p*, then one does not *see* that one sees that *p*; rather, one perceives this by a part which is common to all sense organs.

Now suppose that the following principle holds:

(1) If one perceives by sight that *p*, then one sees that *p*.

If we accept (1), (A1)–(A3) have the following absurd consequence (by propositional logic):

(2) One does not see that *p*.

In other words, seeing is impossible. On the other hand, if (1) is invalid, then there is no problem. The question is therefore: did Aristotle accept (1) or did he not? The answer is clear: he emphatically rejected it in both the *DA* and the *DS* passages.

Let us examine the *DA* first. At (*e*), an imaginary opponent proposes the following principle:

(3) To perceive by sight (*tēi opsei aisthanesthai*) is to see (*horan*).

This principle evidently implies (1). The opponent points out that (A2) and (3) entail that "that which sees is coloured". He regards this as a *reductio ad absurdum* of (A2). Aristotle does not accept it as such because he thinks that "that which sees" is indeed coloured: he says so at (*g*). Nevertheless, he does not accept the opponent's argument, for he regards (1) as unacceptable. According to him, there are clear cases of perceiving by sight which are not cases of seeing. For example, you cannot *see* that it is dark (you do not see anything when it is dark), but *can* perceive this *by sight*. So not all perception by sight involves seeing. (1) and hence (3) are to be rejected. As (*f*) says, seeing is just one form of perceiving by sight: "to perceive by sight is not a single thing". Reflective awareness of seeing is apparently similar to perceiving that it is dark: it is perception by sight which is not seeing (by sight).

There is one objection which may be raised against the above interpretation: subsentence (iii) of sentence (*e*) looks like a *recapitulation* of the main *DA* conclusion. So it seems that we should have formulated it as "if we see that *p*, then we *see* that we see that *p*" after all. However, this reading of (iii) of (*e*) is mistaken. It is not a recapitulation of the *DA* conclusion, but a *consequence* of it *given* the opponent's unacceptable premise (3). Thus, the imaginary opponent is a sly fellow: he ascribes a claim to Aristotle which does not follow from Aristotle's remarks unless we accept his own erroneous assumption.

Now, how is it possible to *perceive by sight* that *p* without *seeing by sight* that *p*? What is the difference between both kinds of perceiving? Here the *DS* passage comes in. Since it is rather sketchy, we expand it to the following more systematic account, for which there is much additional textual evidence.[30]

1. The sensory apparatus consists of five special senses: the organ of sight, the organ of hearing, and so on. Each special sense consist of two basic parts: a common part, i.e., a part which belongs to all other special senses as well (this is *to koinon morion tōn aisthētēriōn hapantōn* which is referred to in the *DS* passage), and a special part (*idion morion*), i.e., a part which belongs to no other special

[30]Cf. Kahn (1966, pp. 50–70).

sense. The common part may be regarded as a sense in its own right, even though it does not have a special part; we call it "the common sense".[31]

Figure 1 illustrates the points we have just made. The large pentagram represents the sensory apparatus as a whole. The small triangular regions **a**, **b**, **c**, **d**, and **e** represent the special parts of the special senses. The pentagonal region **k** in the middle is the common part of the senses. One of the special senses, the one consisting of **a** and **k**, has been shaded; the other special senses are representable by means of similar quadrilaterals. Notice that there are no parts which belong to several special senses but not to the common part.

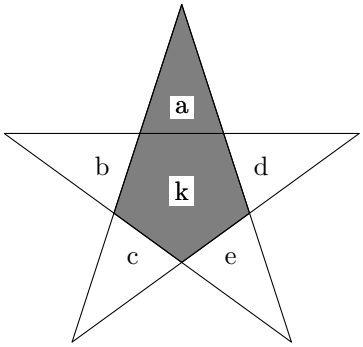

Figure 1. The sensory apparatus.

The corresponding anatomical picture is as follows. The common part of the sense organs is identical with the heart. The special part of a sense organ consists of that organ "minus" the heart, i.e., it is the "sum total" (mereological union, sum or supremum) of the peripheral parts of the organ. The special part of sight, for example, consists of the eyes and perhaps some of the blood vessels with which they are connected to the heart, the special part of hearing is identical with the ears, and so on. Thus, the special sense organs have a considerable spatial extension. They are spread out between the surface of the body and the heart. If any of them were confined to the periphery there would be no *koinon morion tōn aisthētēriōn hapantōn*.

2. The perceptual activities of the five special parts are described by five special perceptual verbs, one for each part. The perceptual activity of the special part of sight (the eyes) is called "seeing", that of the special part of hearing (the ears) "hearing", and so on. Thus, "to see that p" = "to perceive by the special part of sight that p", and similarly for "to hear" and so on. The "part which sees" (*to horōn*) which is referred to in the *DA* passage is identical with the special part of sight.

The claim that seeing is an activity of the special part of sight (i.e., the eyes) may seem surprising. This claim has, indeed, given rise to some controversy in the nineteenth century.[32] Modern scholars no longer doubt, however, that Aristotle regarded seeing as a peripheral process.[33] It is the eyes, or even more specifically, the *eye-jelly* (*hē korē*), which are responsible for all seeing. This does not imply that the proper functioning of the central organ is not a necessary condition for seeing to take place; it only implies that it is not a sufficient condition. Nor does it imply that the eyes do not keep the heart informed of what they see. But even when they do so (which is not always the case, as the theory of dreaming in the *De insomniis* tells us), the common part does not *see* what the eyes see: it can only *perceive* what they *see*.[34]

[31] Our main reason for regarding the common part as a sense in its own right is that this makes it easier to specify the informal reading of the formula t^*Ap below.

[32] See Block (1964).

[33] Cf. Bäumker (1877), Slakey (1961), Sorabji (1974).

[34] It is, strictly speaking, not entirely correct to say that *the eyes* see; nor should one use terms like *to horōn*. We perceive *by means of* our eyes, and whenever we do so, *we* are the perceiving entities. We may assume that Aristotle personified the parts of the sensory apparatus only for the sake of brevity, and we do so for the same reason.

The common part has no special part and there is no special verb to describe its perceptual activity either. This does not mean that it does not have its own distinctive perceptual functions. There would be no reason to assume that the senses overlap if their common part did not make its own contribution. The *DS* contains an example of an *idia aisthēsis* of the common part, a perceptual activity which can only be ascribed to it:[35] it is the only basic part which can perceive that sweet things are different from white ones. The special parts of the sense organs cannot make such "typically intermodal" perceptual judgments.

3. How are perceptual ascriptions to senses related to perceptual ascriptions to their basic parts? We think the only reasonable view is the following one. First, perceptual ascriptions are *expansive*: if a is a part of b and the agent perceives by a that p, then he perceives by b that p.[36] So if an agent perceives by either the common or the special part of sight that p, then he perceives by sight that p; and if his sense of sight perceives that p, then *he* perceives that p. Second, all perception is *grounded* in perception by basic parts: one cannot perceive that p unless at least one of one's senses perceives that p, and a sense cannot perceive that p unless at least one of its basic parts perceives that p. In other words, there is no "extrasensory perception" or "emergence".

Similar remarks apply to the special perceptual verbs. For example, to see = to see by sight = to perceive by the special part of sight. Notice that the sense of sight is the *only* sense with which one can see, since there is no other sense which has *to horōn* as one of its parts. Seeing is a special function not only of *to horōn*, but of sight (*opsis*) as well.

The assumption that perceptual predicates are expansive implies that all perceptual activity of the common part is *koinē aisthēsis*, i.e., perception which is shared by all sense organs. As soon as the common part perceives that p, all sense organs perceive that p. As a consequence, the *DS* passage is not entirely correct. It states that it is impossible to perceive by taste or sight that sweet things are different from white ones. But this is not true: all perceptual activity of the common part, including its *idia aisthēsis* (if we are allowed to use this term in this non-Aristotelian way), is *koinē aisthēsis*, so the organs of taste and sight—and hearing as well!—perceive that sweet things are different from white things whenever the common part perceives this. The *DS* passage would have been expressed more fortunately if it had asserted that one cannot *taste* or *see* that sweet things are different from white ones, or equivalently, that one cannot perceive this by the *special* parts of taste and sight.

4. The above explains how it is possible to perceive by sight that p without seeing (by sight) that p: the common part may be the only basic part of sight which is doing the perceiving. As Figure 2 illustrates, there are no less than *three* ways in which one may perceive by sight: (a) the *koinon morion* **k** may be the only perceptually active part, (b) special part **h** (*to horōn*) may be the only active part, or (c) both parts may be active. There is no seeing unless "the part which sees" is involved; so (a) represents a situation in which one perceives by sight but does not see.

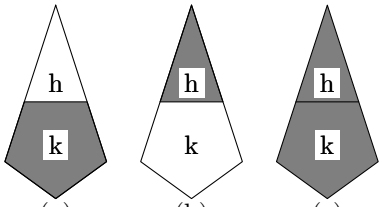

Figure 2. Perceiving by sight.

[35]It should be noted that Aristotle never applied the term in exactly this way.

[36]The term "expansive" comes from Goodman (1951, p. 49).

The *DS* asserts that perception of seeing is perception of type (a): it is perception by sight which is not seeing because it is exclusively provided by the common part. As Aristotle said, it is not the *idion morion* of sight which is responsible for our perceiving that we see; it is not by exercising the *special function* of sight (viz., seeing) that we perceive that we see; "we do not *see by sight* that we see". We could also say that it is not by sight *qua sight* that we perceive that we see, or "that it is not as a special and distinct sense that vision gives us awareness of vision", to recall Rodier's remark which has been quoted above. Nevertheless, we do *perceive by sight* that we see. For we do so by "a part which is common to all the organs of sense", and hence belongs to sight too; because perceptual ascriptions are expansive, its perception of seeing counts as perceiving by sight. It is by sight *qua* containing the common part that we perceive that we see.[37]

It is clear that the above implies that there is no conflict between the *DA* and *DS* accounts. In fact, (A3) entails (A2) because perceiving by the common part of sight is perceiving by sight *simpliciter*. (2) does not follow; seeing is possible after all.

This, then, is our solution to the problem of reconciling the *DA* and *DS* passages. Our interpretation shows that the *DA* and *DS* passages form one continuous story in which the common part is gradually singled out as the *locus* of reflective awareness. The *DS* does not contradict the *DA*: it merely extends the latter account to the intrasensory level. It supplements the *DA* account, which is exactly what Sorabji said: but he did not make clear what he meant, whereas we hope to have done so.

There is one commentator who has come close to the above interpretation. (Rodier may have come close too, but was not sufficiently explicit.) Osborne has made a remark about the *DA* passage which shows that she was aware of the fact that Aristotle distinguishes between seeing and perceiving by sight:

> There were three options canvassed: that we perceive that we see (a) by a sense other than sight (*heterai*); (b) by the same sense, that is *opsei*, but not in the same way, that is not by seeing, *horan*; (c) by the same sense, and in the same way—in fact that we see that we see.[38]

It is clear that we think that option (b) expresses Aristotle's view best. Osborne herself calls the paragraph "inconclusive". She argues neither for nor against option (b), and in fact seems to forget her own suggestion as she goes along.

We want to make two final remarks.

First, the senses are often said to be "parts" of the "sensory soul" (*hē aisthētikē psuchē*). This suggests that the sensory soul is somehow related to the perceptual activity of the sensory apparatus as a whole. If this is correct, the sensory psyche is not confined to the heart (let alone a point without extension): it covers the whole region between the heart and the distal parts of the sensory system. Interestingly enough, Aristotle's contemporary Diocles of Carystus ascribed a similar diffuse location to the psyche.[39]

Secondly, the above theory is not as antiquated as it may seem. Replace "blood vessels" by "nerves" and "heart" by "cerebral association areas" ("tertiary cortex"), and one obtains a theory which is as acceptable to a modern physician as it was to his distinguished colleagues from ancient Athens.[40] Thus, Aristotle's remarks on

[37] Notice that the *Nicomachean ethics* and *De sensu* passages which have been quoted in Section 1 similarly say that we *perceive* that we see, and not that we *see* that we see.

[38] Osborne (1983, p. 402).

[39] Lokhorst (1982).

[40] The most significant difference between Aristotle's theory and modern ones is that that it is no longer assumed that any part which belongs to several special sense organs belongs to

reflective awareness cannot be dismissed by saying that they are based on outdated physiology.

4. Analysis of the arguments: basic principles

So far, we have only stated Aristotle's basic theses. We have not yet analyzed his arguments and determined whether they are correct. Commentators have generally found them "very difficult" to understand.[41] For example, Kosman wrote that the *DA* arguments are "neither straightforward nor clear". Hamlyn accused Aristotle of considering "impossible circumstances" at (*b*) of the *DA* passage, and concluded that "the solution to the problem which Aristotle is attacking is impossible". Hamlyn regarded the argument at (*c*) as "better", but Kosman found it "puzzling" too. As regards (*e*)–(*g*), Hamlyn came to the conclusion that "there seems no way of making the argument coherent".[42]

We shall show that all three arguments become understandable and perfectly unproblematic if the following principles are assumed.

1. Perception is *transparent*:

(Tr) For all basic parts x and y: if one perceives by x that one perceives by y that p, then one perceives by x that p.

Many commentators have realized that the argument at (*b*) of the *DA* passage presupposes some such principle of transparency. Nobody has realized that this principle plays a vital role in the argument at (*c*) too. Nor has anyone realized that it sheds a great deal of light on the *DS* passage: we shall use it to explain why it is in the context of a theory a sleep (of all places) that Aristotle introduced (A3).

Despite its indispensability, there is a problem with (Tr): it looks implausible and *ad hoc*. Why should one assume that one sees that the wind is blowing if one sees that one hears that the wind is blowing? The latter circumstance certainly does not arise very often, and Hamlyn even regarded all circumstances of this type as "impossible". Kosman has tried to defend the principle of transparency by pointing out that Moore and Sartre similarly regarded sensations as "diaphanous".[43] But it is doubtful whether they were really having the same principle in mind. For this reason, it would be better if we were able to derive (Tr) from some more plausible principles. Fortunately, it is not difficult to find such principles.

First, we assume that perception is *veridical* (infallible):

(Ver) For all basic parts x: if one perceives by x that p, then p.

It is a well-known fact that Aristotle accepted this principle; for example, *DA* 428a11 asserts that "sensations are always true". This is one of the respects in which *aisthēsis* differs from *phantasia*, which may be false (*ibid.*).

Secondly, we assume that perception is *closed under simple consequence*: each basic part perceives the consequences of what it perceives.

(RM) If q is a consequence of p (according to the theory of perception), then basic part a perceives that q whenever it perceives that p.

(RM) is a strong rule, and it is certainly inappropriate if one is interested in *actual* perception. But we are not interested in this kind of perception: as we have said before (Section 2), we are interested in *potential*, *idealized* perception. (RM) seems acceptable enough if one wants to capture this notion of perception. For

all sense organs (cf. Point 1 of the present Section). There may be a specifically visual-auditory integrating centre which does not belong to the organ of smell, and so on.

[41] " ... his very difficult analysis of self-awareness ... " (Schiller (1975, p. 294), on *DA* III.2).

[42] Kosman (1975, pp. 500–4); Hamlyn (1968, pp. 121–3).

[43] Kosman (1975, p. 514 ff.).

this reason, modern perceptual logicians (who are usually interested in similarly idealized notions of perceiving) invariably adopt it too.[44]

It is clear that (Ver) and (RM) jointly imply (Tr). This may dispel some of the mystery with which the latter principle has always been surrounded in previous commentaries.

2. The special part of sight is the only special part which can determine whether the predicate "is coloured" applies. "Is coloured" is a "characteristically visual predicate" which expresses the "characteristically visual property" of being coloured.

(C1) The special part of sight is the only special part which can perceive that something is coloured.

In other words, one can see that something is coloured; the eyes may conceivably inform the common part of this fact; but there is no part which does not belong to sight which can perceive this. One cannot hear or smell that, e.g., rainbows are coloured. (C1) should not be strengthened to "the special part of sight is the only *basic* part which can perceive that something is coloured", because we want to leave open the possibility that the common part perceives everything which is perceived at all. (Cf. Section 8 below.)

(C1) is obviously related to Aristotle's claim that colour is the "proper object" (*idion aisthēton*) of sight (*DA* II.6-7). However, Aristotle used the term "object of perception" in a very broad sense. He did not only apply it to properties or qualities of physical objects, but also to those objects themselves. For example, *DA* II.6 mentions colours and flavours as the proper objects of sight and taste, whereas coloured things and bodies in which flavours reside play the same role in *DA* II.7 and II.10.[45] So rainbows may be said to be proper objects of sight as well.[46]

Is there any predicate which is characteristic for the common part in the sense of (C1)? Yes, there are many, and the *DS* passage mentions one of them: the common part is the only basic part which can ascertain whether the predicate "is sweet but not white" applies. Being sweet but not white is a proper object of the common part, and a complex physical object like a sweet black Brazilian cigar in its full gustatory-olfactory-visual splendour presumably is one too.

The proper objects of the common part should not be confused with the so-called "common sensibles" (*koina aisthēta*). This notion is reserved for *aisthēta* which are perceivable by more than one special part (*DA* III.1). No object of perception is both a proper object of the common part and a common sensible.

Nor should the term *aisthēton* ("object of perception" or "percept") be confused with the term *aisthētikon* ("entity which perceives" or "percipient"). The latter term refers to any entity which is perceptually active. If some basic part perceives something, then that part, the sense(s) to which it belongs, and the person to whom it belongs are *aisthētika* (by the expansiveness of perception).

3. All things which are seen have a certain property in common:

---

[44] See, e.g., Hintikka (1969),Thomason (1973), Clark (1976), Bacon (1979).

[45] Cf. Schiller (1975, pp. 287–9).

[46] It is nowadays customary to regard the objects of perception as *facts* (assuming that perception is veridical, as we have done). When I see that the cursor on my screen is blinking, I perceive *the fact that the cursor is blinking*. It is this *fact* which is the object of my perception. (It is a *proper* object of sight, since I cannot hear or feel whether it is the case.) Aristotle never used the term "object of perception" in this way. Cursors and blinking are examples of proper objects of sight, the fact that a particular cursor is blinking is not. Bacon's assertion that Aristotle was a trope theorist *avant la lettre* is accordingly rather suspect. (Bacon (1988); he ascribes this view to Küng (1963) too.) Trope theorists view existent tropes (atomic facts such as the fact that Coriscus is white) as the basic objects of veridical perception; they regard individuals (such as Coriscus) and properties (such as being white) as abstractions. This view is not to be found in Aristotle. He regarded individuals and properties as the basic objects of perception.

(C2) One can only see coloured things.

The latter assumption may look strange. However, it is suggested by subsentence (ii) of sentence (*e*) of the *DA* passage. From a modern point of view, "emits or reflects light" is a better example of a predicate which applies to all things which are seen. (All things which are perceived by the common part have a property in common too: the property of being a sensible object.)

4. Aristotle certainly accepted the following assumption:

(Ekh) The common and special parts of sight exist.

5. On the other hand, we have to assume that he regarded the following two principles as *unacceptable*:

(*C) It is impossible to see that something is coloured.

(*Inf) For any finite number n: it is not the case that there are n basic parts. I.e., the sensory system has an infinite number of basic parts.

(*C) and (*Inf) will be used as the bottom lines of *reductio ad absurdum* arguments. If it can be shown that some principle entails either (*C) or (*Inf), then it should be rejected.[47]

6. Finally, we have to adopt some simple logical principles. This is inevitable: all kinds of rational argumentation presuppose some logic. Aristotle did not specify his logical principles, but we will do so since we want to be as explicit as possible.

It goes without saying that our basic logic should be as weak as possible. Therefore we will use (i) classical propositional logic; (ii) the quantificational principles of monadic free logic (without identity); and (iii) a very weak theory of identity. The choice of (i) needs no defense; we have already used it in the previous Section. (ii) is more questionable. We could also have used classical monadic predicate logic (arguably the most natural modern reformulation of syllogistic logic), but using the more expressive language and weaker quantificational theory of free logic leads to greater formal elegance.[48] In any case, we do not want to consider predicates of degree higher than one (except identity). It would be all too anachronistic to ascribe the modern view of relations to Aristotle. (iii) is justified because Aristotle clearly realized that identity is a rather weak concept in perceptual contexts. In the Coriscus examples of *De sophisticis elenchis* 179a25 ff., he pointed out that the following expressions are not at all logically equivalent:

(1) c is identical with b and a perceives that b is approaching;
(2) there is someone who is identical with c and whom a perceives as approaching (a perceives of c that he is approaching);
(3) a perceives that c is approaching.

(3) might be false even if (1) and (2) are true because a might fail to identify the person who is approaching as c. It might be the case that you see that someone is coming towards you; this person might in fact be identical with Coriscus; but you might fail to recognize him as Coriscus, because he is, for example, still too far away or wearing a mask.[49] Neither (1) nor (2) entails (3) in the systems we will present.

The principles which we have mentioned in this Section enable us to reconstruct Aristotle's arguments as clear and correct logical arguments, as we will now show.

[47] We do not assume that there are six basic parts because we will have to consider arbitrarily large sensory systems below.

[48] For this reason, most modern accounts of the logic of perception are based on free logic; see, e.g., Hintikka (1969), Thomason (1973), Clark (1976), Bacon (1979).

[49] Thomason discusses the Coriscus examples in more detail (Thomason 1973, pp. 276–7).

5. A formal language

The above informal principles are awkward to work with, especially since we will have to deal with statements which contain large numbers of nested perceptual expressions. So let us formalize. This is desirable in any case. Using a symbolic language leads to a much greater clarity of expression than any natural language can provide. It makes it easy for the readers to determine the exact content of our proposals: they can calculate their consequences for themselves and do not have to rely on some kind of *einfühlendes Verstehen* (empathetic "understanding").

In order to keep matters simple, we assume that there is only one agent. The alphabet of our formal language $\mathscr{L}$ consists of:

(1) A denumerable set *VAR* of individual variables;
(2) A denumerable set *CONST* of individual constants;
(3) A denumerable set *PRED* of monadic predicates;
(4) Two special individual constants 'k', 'h' $\in$ *CONST*;
(5) A special monadic predicate 'C' $\in$ *PRED*;
(6) The logical symbols '$\neg$', '$\wedge$', '$\forall$', '$=$', 'E', 'A';
(7) The parentheses '(' and ')'.

The logical signs '$\vee$', '$\leftrightarrow$', '$\exists$', '$\bot$', and '$\top$' are defined as usual. '$\bot$' stands for a preselected contradiction; '$\top$' is an abbreviation of '$\neg\bot$'. We read 'k', 'h', 'C' and 'E' as follows:

k "the common part (*koinon morion*) of the agent's sense organs";
h "the agent's special part of sight" ("that which sees", "*to horōn*");
C "is coloured (*chrōmatos*)";
E "exists" ("is a genuinely existent thing").[50]

We use 'x', 'y' and 'z' as metavariables on *VAR*, 'a', 'b' and 'c' as metavariables on *CONST*, 's', 't' and 'u' as metavariables on *VAR* $\cup$ *CONST*, and 'P' as a metavariable on *PRED*, all of them possibly with subscripts. The set *WFF* of well-formed formulas (wffs) is the smallest set such that:

(1) $Pt \in$ *WFF*, $s = t \in$ *WFF*, $Et \in$ *WFF*;
(2) if $p, q \in$ *WFF*, then $\neg p$, $(p \wedge q)$, $\forall xp$, $tAp \in$ *WFF*.

We use p and q as metavariables on *WFF*. Informal readings of some selected formulas:

tAp "t is a basic part of the agent's sensory system and the agent perceives (*aisthanetai*) by means of t that p" (or more concisely, "t is a basic part and t perceives that p").
kAp "the agent's common part perceives that p".
hAp "the agent's special part of sight perceives that p", "the agent's organ of sight sees that p", "the agent sees that p"—cf. Section 3.
$\exists xxAp$ "the agent perceives that p".

Finally, we define:

(1) $\bigvee_{i=1}^{0} p_i \stackrel{\text{def}}{=} \bot$; $\bigvee_{i=1}^{n+1} p_i \stackrel{\text{def}}{=} p_{n+1} \vee \bigvee_{i=1}^{n} p_i$;
(2) $t_0 \ldots t_1 Ap \stackrel{\text{def}}{=} p$; $t_{n+1} \ldots t_1 Ap \stackrel{\text{def}}{=} t_{n+1} A t_n \ldots t_1 Ap$;
(3) $t^* Ap \stackrel{\text{def}}{=} kAp \vee tAp$;
(4) $sOt \stackrel{\text{def}}{=} \exists x(x = t \wedge sAEx)$;

[50] Hintikka (1969), Thomason (1973) and Bacon (1979) have argued that perceptual contexts require a distinction between two "modes of individuation", perceptual and physical, and that the difference between them should be captured by means of two corresponding existence predicates and two sorts of quantification. (This proposal is criticized in Clark (1976).) We want to keep matters simple and will not make this finer discrimination. Our existence predicate may be read as "exists physically or visually or auditively or ... or tactually, or in any other way".

(5) $s^*Ot \stackrel{\text{def}}{=} \exists x(x = t \wedge s^*AEx)$;

(6) $Bt \stackrel{\text{def}}{=} t = k \vee t = h \vee tA\top$.

We read these expressions as follows.

t^*Ap "t is one of the agent's basic parts, and the agent perceives by means of the smallest sense to which t belongs that p";[51]

k^*Ap "the agent perceives by means of the common part that p";

h^*Ap "the agent perceives by sight (*tēi opsei*) that p";

sOt "the agent's basic part s perceives t";[52]

s^*Ot "the smallest sense to which s belongs perceives t";

Bt "t is one of the agent's basic parts".

We presuppose the usual definitions of free and bound occurrences of variables. (N.B.: the first occurrence of x in xAp is free.) A sentence is a formula without free individual variables. We write $p[t/x]$ for the result obtained from p by replacing all free occurrences of x in p by t (relettering bound variables if necessary to avoid binding a free variable in p).

6. Four systems of perceptual logic

We define a logical system $\mathcal{P}_0$ (basic Aristotelian logic of perception) in the language $\mathscr{L}$ as follows.

Definition: $\mathcal{P}_0$ is the smallest set of wffs which is closed under the following axiom schemata and rules.

(PT) p, if p is a propositional tautology;

($\forall$1) $\forall xp \rightarrow (Et \rightarrow p[t/x])$;

(=1) $t = t$;

(=2) $s = t \rightarrow ((s = u \leftrightarrow t = u) \wedge (u = s \leftrightarrow u = t) \wedge (Ps \leftrightarrow Pt) \wedge (sAp \leftrightarrow tAp))$;

(Ver) $\forall x(xAp \rightarrow p)$;

(C1) $\forall x \forall y(xACy \rightarrow (x = k \vee x = h))$;

(C2) $\forall x(hOx \rightarrow Cx)$;

(Ekh) $Ek \wedge Eh$;

(MP) $p \rightarrow q, p/q$;

($\forall$2) $p \rightarrow (Et \rightarrow q)/p \rightarrow \forall xq$, provided that t does not occur in $p \rightarrow \forall xq$;

(RM) $p \rightarrow q/tAp \rightarrow tAq$.

All these axioms and rules are straightforward translations of their informal counterparts in Section 4 above. (PT), (MP), ($\forall$1) and ($\forall$2) jointly constitute monadic free logic, MFL. (=1) and (=2) embody the "very weak concept of identity" which we have introduced in Section 4, whereas (Ver), (RM), (C1), (C2) and (Ekh) are the specifically perceptual principles we have discussed there.[53]

Notice that E is an intensional, referentially opaque predicate: $s = t$ does not imply $Es \leftrightarrow Et$. The defined predicate O, on the other hand, is extensional or

[51] Because we encounter non-basic parts of the perceptual system (such as the sense of sight) only as the subjects of perceptual ascriptions, this "contextual definition" is sufficient for our purposes. If we had wanted to be completely explicit, we would have used mereological tools and would have defined t^* as the mereological sum of k and t. (On mereology, see Eberle (1970).)

[52] Similar proposals for defining the direct object construction ("perceiving of") in terms of the propositional attitude construction ("perceiving that") have been made by Hintikka (1969, p. 166, formula (15)) and Bacon (1979, p. 294, §63, formula (22)). Thomason's approach is related but different (Thomason 1973, p. 274 ff.).); we could have used it as well. (Aristotle's own writings about perception are not sufficiently explicit to enable us to make a choice.)

[53] Our treatment of quantification is the same as Garson's (Garson 1978, Garson 1984). His account may be seen as a simplified and generalized version of Thomason's Q3, on which the perceptual logics of Thomason (1973) and Bacon (1979) are based.

referentially transparent (as it should be): $s = t \rightarrow (uOs \leftrightarrow uOt)$ is a theorem by (=2).

We may prove the following theorems. $\vdash_0 p$ means that p is a theorem of $\mathcal{P}_0$, i.e., that p follows from the axioms by means of a finite number of applications of the rules.

(Tr) $\vdash_0 \forall x \forall y(xAyAp \rightarrow xAp)$, by (Ver) and (RM).

This is the principle of transparency from Section 4. (Tr) is formally identical to axiom Q from the logic of agency. It there captures the legal maxim *Qui facit per alium facit per se*.[54] We may analogously say: [*sensus*] *qui percipit per alium* [*sensum*] *percipit per se*.

Hamlyn writes that Aristotle "seems to assume that if I perceive by sense Y that I see X, I must therefore perceive X by Y".[55] This weaker principle of transparency (which he incorrectly instantiates to "if I see that I am tasting a strawberry, then I see a strawberry" instead of "if I perceive by sight that I am tasting a strawberry, then I perceive a strawberry by sight") is also derivable:

(Tr′) $\vdash_0 \forall x \forall y(x^*AyAp \rightarrow x^*Ap)$, by (Tr) and MFL.

(Tr′) follows from (Tr) and MFL, but the converse is not true. As we shall see, we really need the stronger (Tr) to make sense of the first two *DA* arguments.

The *DA* and *DS* axioms from Section 3 may be expressed as follows.

(A1) $\forall x(xAp \rightarrow \exists y yAxAp)$;[56]
(A2) $\forall x \forall y(xAyAp \rightarrow (x = k \vee x = y))$;
(A3) $\forall x \forall y(xAyAp \rightarrow x = k)$.

Definition: $\mathcal{P}_1$ is the smallest extension of $\mathcal{P}_0$ which is closed under (A1). Similarly, $\mathcal{P}_2 \stackrel{\text{def}}{=} \mathcal{P}_1 + (\text{A2})$, and $\mathcal{P}_3 \stackrel{\text{def}}{=} \mathcal{P}_1 + (\text{A3})$.

We write $\vdash_i p$ for derivability in $\mathcal{P}_i$ (where $0 \leq i \leq 3$). It is clear that $\vdash_1 p \Rightarrow \vdash_2 p \Rightarrow \vdash_3 p$. The difference between the three theories is that the *locus* of reflective awareness gradually comes into sharper focus, as the following theorems may make clearer:

(T1) $\vdash_2 \forall x(xAp \rightarrow x^*AxAp)$, by (A1), (A2) and MFL.
(T2) $\vdash_3 \forall x(xAp \rightarrow kAxAp)$, by (A1), (A3) and MFL.

A formula like (A1) instantiates to $hAp \rightarrow \exists x xAhAp$. In accordance with the pronunciation rules of the preceding section, this is read as "if one sees that p, then one perceives that one sees that p". (Ekh) is absolutely indispensable to derive this conclusion; it cannot be replaced by a weaker axiom such as $\exists x x = k \wedge \exists x x = h$.

7. Reconstruction of the *De anima* arguments

The first two *DA* arguments are not arguments for (A2), but arguments against (*):

(*) $\neg \exists x x^* AxAp$.

In other words, reflective awareness of the perceptual activity of a basic part is not provided by the smallest sense to which that part belongs: it is provided by some part outside that sense (and hence by some other sense). For example, one cannot perceive by sight that one sees, but only by some part which does not belong to sight. It is important to realize that a *reductio ad absurdum* of (*) is not a proof of (A2). Aristotle apparently thought that we either have to accept (A2) or (*).

---

[54] Chellas (1991, note 6). He thanks Noyes Leech for drawing his attention to this legal maxim.

[55] Hamlyn (1968, pp. 121–2).

[56] It does not make a difference whether one adopts (A1) or $\forall x(xAp \rightarrow \exists y y^* AxAp)$.

7.1. The first argument. The argument at (*b*) goes as follows. Suppose that we add (*) to $\mathcal{P}_1$ and call the resulting system $\mathcal{P}_{1^*}$; $\vdash_{1^*} p$ means that p is derivable in $\mathcal{P}_{1^*}$. We may then prove:

(1) $\vdash_{1^*} hACt \to \exists x xAhACt$, by (Ekh) and (A1);
(2) $\vdash_{1^*} (Es \wedge sAhACt) \to sACt$, by (Ekh) and (Tr);
(3) $\vdash_{1^*} (Es \wedge Et \wedge sACt) \to (s = k \vee s = h)$, by (C1);
(4) $\vdash_{1^*} (Es \wedge Et \wedge sAhACt) \to (kAhACt \vee hAhACt)$, by (2), (3) and (=2);
(5) $\vdash_{1^*} (kAhACt \vee hAhACt) \to \exists x x^* AxACt$, by (Ekh) and ($\forall$1);
(6) $\vdash_{1^*} (Es \wedge Et) \to \neg sAhACt$, by (4)–(5) and (*);
(7) $\vdash_{1^*} Et \to \neg\exists x xAhACt$, by (6) and ($\forall$2);
(8) $\vdash_{1^*} Et \to \neg hACt$, by (1) and (7);
(9) $\vdash_{1^*} \neg\exists x hACx$, by (8) and ($\forall$2).

In other words, one cannot see that something is coloured. This conclusion is identical with the unacceptable (*C) from Section 4, so we reject (*).

7.2. The second argument. Suppose that the agent's sensory apparatus has at most n basic parts. We can express this by means of the following axiom:

($\exists n$) $\exists y_1, \ldots, y_n \forall x (Bx \to \bigvee_{i=1}^{n} x = y_i)$.

We call the resulting system $\mathcal{P}_{1^*n}$ and write $\vdash_{1^*n}$ for derivability in $\mathcal{P}_{1^*n}$. We can now prove that one cannot perceive anything.

(1) $\vdash_{1^*n} (Et_1 \wedge t_1 Ap) \to \exists x_2, \ldots, x_{n+1}(x_2 At_1 Ap \wedge \ldots \wedge x_{n+1} \ldots x_2 At_1 Ap)$, by n applications of (A1).
(2) $\vdash_{1^*n} (Et_1 \wedge \ldots \wedge Et_{n+1} \wedge t_1 Ap \wedge \ldots \wedge t_{n+1} \ldots t_1 Ap) \to \bigvee_{i=2}^{n+1} \bigvee_{j=1}^{i-1} (Et_1 \wedge \ldots \wedge Et_{n+1} \wedge t_1 Ap \wedge \ldots \wedge t_{n+1} \ldots t_1 Ap \wedge t_i = t_j)$, by ($\exists n$). Not all t_i, $1 \leq i \leq n+1$, can be different: there has to be some first $i \leq n+1$ such that there is some $j < i$ such that $t_i = t_j$.
(3) $\vdash_{1^*n} (Et_1 \wedge \ldots \wedge Et_i \wedge t_i \ldots t_{j+1} At_j \ldots t_1 Ap \wedge t_i = t_j) \to t_j At_j \ldots t_1 Ap$, by one application of (=2) and $i - (j+1)$ applications of (Tr). $t_j At_j \ldots t_1 Ap$ implies $t_j^* At_j^* At_{j-1} \ldots t_1 Ap$ by (RM), which reminds us of Aristotle's remark at (*c*) that "some [sense] is concerned with itself".
(4) $\vdash_{1^*n} (Et_1 \wedge \ldots \wedge Et_j) \to \neg t_j At_j \ldots t_1 Ap$, by (*).
(5) $\vdash_{1^*n} \neg(Et_1 \wedge \ldots \wedge Et_{n+1} \wedge t_1 Ap \wedge \ldots \wedge t_{n+1} \ldots t_1 Ap)$, by (2)–(4) and propositional logic.
(6) $\vdash_{1^*n} (Et_1 \wedge t_1 Ap) \to \neg\exists x_2, \ldots, x_{n+1}(x_2 At_1 Ap \wedge \ldots \wedge x_{n+1} \ldots x_2 At_1 Ap)$, by (5) and ($\forall$2).
(7) $\vdash_{1^*n} Et_1 \to \neg t_1 Ap$, by (1) and (6).
(8) $\vdash_{1^*n} \neg\exists x x Ap$, by (7) and ($\forall$2).

In other words, perception is impossible. This conclusion is unacceptable because it implies (*C) from Section 4. So we either have to reject the assumption that the sensory system contains only a finite number of basic parts, or (*). Rejecting the former assumption means that the sensory system is essentially infinite. In contrast to Royce and Rucker, Aristotle did not take this possibility seriously, as we have already seen when rejecting (*Inf) in Section 4 above. So it is (*) which is once again to be rejected.

Assuming that we have to choose between (*) and (A2), the above two arguments establish (A2): $\mathcal{P}_2$ is preferable to $\mathcal{P}_{1^*}$. Reflective awareness of the activity of a basic part is provided by (some part of) the (smallest) sense to which that part belongs.

Notice that Hamlyn's claim that the argument at (c) is "better" than the one at (b) is nonsensical.[57] Both arguments are equally correct.[58]

7.3. The third argument. The argument at (e)–(g) proceeds as follows. An imaginary opponent proposes adding (**) to $\mathcal{P}_2$:

(**) $\forall x(x^*Ap \to xAp)$.

We call the resulting system $\mathcal{P}_{2^*}$ and write $\vdash_{2^*}$ for derivability in $\mathcal{P}_{2^*}$. The opponent points out that we may now prove that "that which sees" (*to horōn*, i.e., the special part of sight) is coloured. The latter claim is correct—but the argument does not depend on (**)! One can prove that*to horōn* is coloured even when one does not accept (**). This is in agreement with Aristotle's own opinions: he rejected (**), but nevertheless asserted that*to horōn* is coloured.

(1) $\vdash_{2^*} hAp \to h^*AhAp$, by (A2) and (Ekh);
(2) $\vdash_{2^*} h^*AhAp \to hAhAp$, by (**) and (Ekh);
(3) $\vdash_{2^*} hAhAp \to hAEh$, by (Ekh) and (RM);
(4) $\vdash_{2^*} hAp \to hAEh$, by (1)–(3), or by (Ekh) and (RM);
(5) $\vdash_{2^*} hAEh \to (Eh \wedge h = h \wedge hAEh)$, by (Ver) or (Ekh), and (=1);
(6) $\vdash_{2^*} (Eh \wedge h = h \wedge hAEh) \to \exists x(x = h \wedge hAEx)$, by ($\forall$1);
(7) $\vdash_{2^*} \exists x(x = h \wedge hAEx) \to \exists x(x = h \wedge Cx)$, by (C2);
(8) $\vdash_{2^*} \exists x(x = h \wedge Cx) \to Ch$, by (=2);
(9) $\vdash_{2^*} hAp \to Ch$, by (4)–(8).

The first three steps are redundant, so we do not need (**).

Aristotle endorses (9) at (g). Thus,*to horōn* shares a property with all things which it sees, namely the property of being coloured. As Aristotle puts it in *DA* II.5 and elsewhere, each special part has the same "shape" or "form" as the things which it perceives.

Aristotle rejects (**) at (f): not all perceiving by sight is seeing. The counterexample he gives is rather unclear, but he seems to mean that you cannot see that something is not coloured (dark) but *can* perceive this by sight. This is provable in our systems: $\forall x(\neg Cx \to \neg hOx)$ is a consequence of (C2), whereas $\forall x(\neg Cx \to \neg h^*Ox)$ is not derivable.[59] It might be the case that $\exists x(\neg Cx \wedge kOx)$. Whatever the exact nature of Aristotle's counterexample, the point which he wants to make is clear: there are cases of perceiving by sight which do not involve seeing, so (**) is unacceptable.

It is a surprising fact that we are able to reconstruct the third *DA* argument. It looks invalid at first sight because it seems to rest on a confusion between the propositional attitude construction and the direct object construction. This has been pointed out by Hamlyn:

This argument is [...] irrelevant in any case, since his concern should be with seeing *that* one sees, and he should show that this involves seeing the thing which sees; this he fails to do.[60]

Hamlyn himself came to the conclusion that "there seems no way of making the argument coherent". Our reconstruction shows that it is, in fact, rather

[57] Hamlyn (1968, p. 122).

[58] The proof we have presented bears some similarity to the argument in Sanford (1975). His proof is not directly applicable to the present case, however, since it is only concerned with extensional logic. Despite its promising title, Stekla (1970) sheds no light at all on Aristotelian infinite regress arguments.

[59] We cannot yet prove that the latter formula is not derivable, but we will be able to prove this by using the techniques which are presented in Section 9 below.

[60] Hamlyn 1968, p. 122.

simple—although it should be kept in mind that it crucially depends on the possibly controversial (Ekh) and (RM) and the particular account of quantification and definition of the direct object construction which we have adopted.

(**) has a curious consequence which is not derivable without it (i.e., which is not derivable in $\mathcal{P}_2$): it implies that one *sees* that "that which sees" is coloured, provided that one sees at all.

(10) $\vdash_{2^*} hAhAp \to hACh$, by (9) and (RM);
(11) $\vdash_{2^*} hAphACh$, by (1), (2) and (10).

8. Reconstruction of the *De somno* passage

Aristotle does not present any argument for (A3) in the *DS* passage, but it is clear what makes it so attractive it to him. To begin with, we note that it implies:

(T3) $\vdash_3 \exists xxAp \to kAp$, by (A1), (A3), (Tr), and MFL.[61]

This agrees with *De sensu* 449a8: "there is one part of the soul with which it perceives everything, although it perceives different things with different parts". The common part is responsible for "intersensory integration". (T3) is an idealization of the same sort as axiom (A1), on which it depends: it only characterizes "potential" or "fully alert" perception, not the kind of inattentive perception which is referred to in the theory of dreaming.

(T3) has the following consequence:

(T4) $\vdash_3 \neg kA\top \to \neg\exists xxA\top$.

In other words, if the common part does not perceive, no basic part (and hence no sense organ whatever) perceives. If it ceases functioning, there is no perception at all. It is the *sine qua non* of perception. As Aristotle says in the *DS* passage, the common part is the *koinē dunamis* ("common power") underlying all perception. Or as he says elsewhere (e.g., in *De insomniis* 461a6), it is the *archē tēs aisthēseōs*, the basis of all perception. It is clear that (T4) is attractive in the context of a theory of sleep: it allows us to identify the common part as the "sleep centre", the part whose inactivity causes a total shutdown of the whole perceptual apparatus. This explains why it is precisely in the *De somno* that Aristotle introduces (A3). No commentator has been able to explain this fact before.

(A3) turns the common part into a perceptual counterpart of "absolute entities" like the *primum mobile* (*Physics* VIII.5, *Metaphysics* XII.7) and the *summum bonum* (*Nicomachean ethics* I.2).[62] Everything which is moved is ultimately moved by the prime mover, and everything which is desired for the sake of something is ultimately desired for the sake of the final good. Similarly, everything which is perceived (including the *aisthētika* themselves) is perceived by the common part.

This point may be made clearer by means of some illustrations. Suppose that a is an existent thing which is seen, i.e., that $Ea \wedge hAEa$ is true.[63] In this case, we have $kAhAEa \wedge kAEa \wedge kAkAEa \wedge \neg hAhAEa \wedge \neg hAkAEa$ by the axioms and rules of $\mathcal{P}_3$, so the situation may be represented as in Figure 3(a), in which a broken arrow from **x** to **a** indicates that **x** perceives **a** and a solid arrow from **x** to **y** indicates that **x** perceives that **y** perceives **a**. A picture of a situation in which something is desired for the sake of the final good would look exactly similar. (Except that it should be drawn by nothing but broken arrows; the relation "is desired for the sake of" is extensional, just like **O**.) The case of movement is slightly different because the prime mover does not move itself, so the circular arrow at the top

[61] As a result, Figure 2(b) no longer represents a possible situation.

[62] The term "absolute entity" comes from Beth (1959, pp. 9–12). Beth gave many examples of absolute entities, but did not notice the psychological application we are giving.

[63] We do not yet have the right to use the notion of truth, but we will acquire this right soon enough (namely in the next Section).

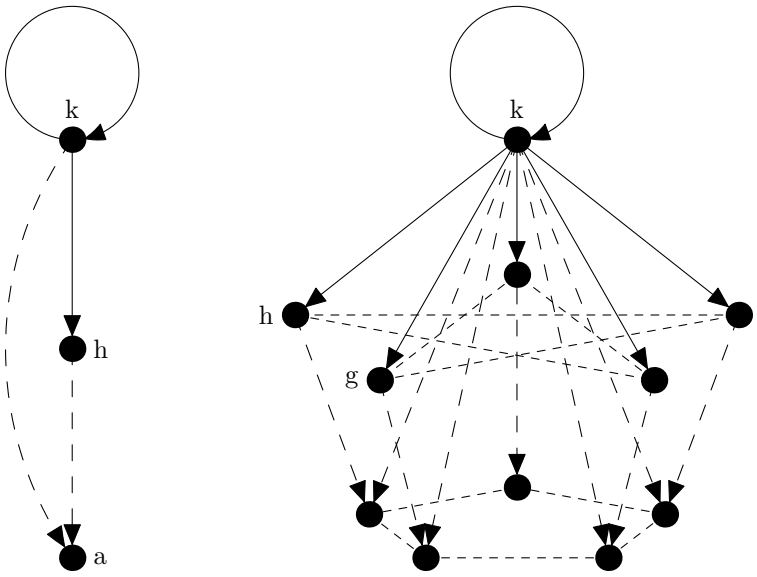

Figure 3. The common part as an absolute entity.

should be omitted; but it is analogous in all other respects.[64] (The divine mind of *Metaphysics* XII.9 represents the opposite case: it only thinks about itself, so its thinking is representable by an isolated circular arrow.)

When there are five special parts which all perceive something (and exist in the same way as *to horōn* and the common part do), the situation becomes as depicted in Figure 3(b). The entity at the top represents the common part, the second layer the special parts, and the bottom layer some entities which are perceived by the latter. **g** is the special part of *geusis*, taste. The common part perceives everything which is perceived by any basic part, including that part itself. The diagram reminds us of Aristotle's description of the common part as a point belonging to several lines.[65] It also reminds us of Florentius Schuyl's description of the psyche as "a spider sitting in the middle of its web".[66]

The correspondence between these various cases is pleasing. Unfortunately, (A3) also has a less welcome consequence: it implies that all senses (but not, of course, their basic parts) have the same *aisthēta*:

(T5) $\vdash_3 \forall x \forall (x^* * Ap \leftrightarrow y^* Ap)$, by (T3).

This is an inevitable consequence of the assumptions that (1) the senses share a common part which perceives everything which they perceive individually and that (2) perceiving by this common part counts as perceiving by the senses themselves.

[64] The number of points lying between a and the absolute entity k might, of course, be greater than one in the case of desire and movement. It might conceivably even be infinite (the relevant relations might, for example, be *dense*). However, Aristotle would protest against the latter suggestion: it is precisely because of his *horror infiniti* that he introduced his various absolute entities.

[65] *DA* 427a9 ff. and 431a20 ff. See the discussion of these passages in Modrak (1981, pp. 417–9). The diagram which Modrak presents on p. 418 is identifiable as part **hkg** of Figure 3(b). Modrak notes (Modrak 1981, note 29, pp. 417–8) that Alexander Aphrodisiensis drew the common part as the centre of a circle at which radii (standing for the faculties) terminate.

[66] "Ut aranea in aranei centro" (Schuyl, "Ad lectorem", in Descartes (1662)).

It is easy to define a "dissective" (instead of "expansive") notion of perception to which (T5) does not apply, e.g., $t@p \stackrel{\text{def}}{=} tAp$, $t^*@p \stackrel{\text{def}}{=} kAp \wedge tAp$.[67] However, it is clear that $\vdash_0 t^*p \rightarrow t@p$. So this is precisely the notion of perception which Aristotle rejects at (f) of the *DA* passage. The *DA* and *DS* passages cannot be reconciled if it is assumed that they are formulated in terms of this concept.

9. Consistency of the interpretation

The above theory is adequate in the sense that it enables us to present formal counterparts of many Aristotelian claims and arguments. However, this is nothing to be proud of: for if our theory were inconsistent, we would not only be able to derive all his conclusions, but all their negations as well. Writers on Aristotle's psychology are seldom worried by such considerations. We are, however, able to prove that our explanatory theory is consistent.

Definition: $\mathcal{P}_i$ $(0 \leq i \leq 3)$ is consistent iff[68] it is not the case that $\vdash_i p$. (I.e., $\mathcal{P}_i$ is consistent iff not every formula is derivable.)

Definition: A type 0 model is a sequence
$\mathfrak{M} = \langle \mathbf{W}, \mathbf{I}, \mathbf{R}, \mathbf{E}, \mathbf{A}, \mathbf{C}, \mathbf{k}, \mathbf{h}, \mathbf{V} \rangle$,
where:

- $\mathbf{W}$ is a non-empty set (of "perceivable worlds" or "perceptual alternatives").
- $\mathbf{I}$ is a non-empty set (of "individuals").
- $\mathbf{R} \subseteq \mathbf{I} \times \mathbf{W} \times \mathbf{W}$. $\mathbf{Riwv}$ means that $\mathbf{v}$ is a perceptual alternative of $\mathbf{i}$ at $\mathbf{w}$.
- $\mathbf{E}$ is a function from $\mathbf{W}$ into the power set of the set of functions from $\mathbf{W}$ into $\mathbf{I}$. $\mathbf{E_w}$ is the set of "things" or "substances" which exist at $\mathbf{w}$.
- $\mathbf{A}$ and $\mathbf{C}$ are functions from $\mathbf{W}$ into the power set of $\mathbf{I}$. $\mathbf{A}$ is the property of being a perceptually active basic part (basic *aisthētikon*). $\mathbf{C}$ is the property of being coloured.
- $\mathbf{k}$ and $\mathbf{h}$ are functions from $\mathbf{W}$ into $\mathbf{I}$. $\mathbf{k}$ is the common part of the senses and $\mathbf{h}$ is the special part of sight.
- The following conditions hold for all $\mathbf{w} \in \mathbf{W}$, $\mathbf{d}$, $\mathbf{e} \in \mathbf{E_w}$:
  [Ver] if $\mathbf{d_w} \in \mathbf{A_w}$ then $\mathbf{Rd_www}$;
  [C1] if $\mathbf{d_w} \in \mathbf{A_w} \setminus \{\mathbf{k_w}, \mathbf{h_w}\}$ then for some $\mathbf{v} \in \mathbf{W}$, $\mathbf{Rd_wwv}$ and $\mathbf{e_v} \notin \mathbf{C_v}$;
  [C2] if $\mathbf{h_w} \in \mathbf{A_w}$ and for all $\mathbf{v} \in \mathbf{W}$, if $\mathbf{Rh_wwv}$ then $\mathbf{d_v} \in \mathbf{E_v}$, then $\mathbf{d_w} \in \mathbf{C_w}$;
  [Ekh] $\mathbf{k} \in \mathbf{E_w}$ and $\mathbf{h} \in \mathbf{E_w}$.
- $\mathbf{V}$ is a function which assigns a function from $\mathbf{W}$*into*$\mathbf{I}$ (i.e., an "individual concept") to each term $t \in VAR \cup CONST$ and a function from $\mathbf{W}$ into the power set of $\mathbf{I}$ (i.e., a "property") to each predicate $P \in PRED$. In addition, $\mathbf{V}$ satisfies the following conditions: $\mathbf{V(k) = k}$, $\mathbf{V(h) = h}$, and $\mathbf{V(C) = C}$.

Definition: Type 1, type 2 and type 3 models are type 0 models which satisfy the additional conditions [A1], [A1] and [A2], [A1] and [A3], respectively, for all $\mathbf{w} \in \mathbf{W}$, $\mathbf{d}$, $\mathbf{e}$, $\mathbf{f} \in \mathbf{E_w}$:

[A1]
- If $\mathbf{d_w} \in \mathbf{A_w}$ then for some $\mathbf{g} \in \mathbf{E_w}$: $\mathbf{g_w} \in \mathbf{A_w}$ and for all $\mathbf{v} \in \mathbf{W}$, if $\mathbf{Rg_wwv}$ then $\mathbf{d_v} \in \mathbf{A_v}$; furthermore,
- if $\mathbf{d_w} \in \mathbf{A_w}$, $\mathbf{Rd_wwv}$, $\mathbf{e_v} \in \mathbf{A_v}$, $\mathbf{Re_vvu}$, and $\mathbf{f_w} \in \mathbf{A_w}$, then $\mathbf{Rf_wwu}$.

[67] The term "dissective" comes from Goodman 1951, p. 48.
[68] "Iff" is short for "if and only if".

[A2] If $\mathbf{e_w} \in \mathbf{A_w} \setminus \{\mathbf{k_w}, \mathbf{d_w}\}$ then there is some $\mathbf{v} \in \mathbf{W}$ such that $\mathbf{Re_wwv}$ and $\mathbf{d_v} \notin \mathbf{A_v}$.

[A3] If $\mathbf{e_w} \in \mathbf{A_w} \setminus \{\mathbf{k_w}\}$ then there is some $\mathbf{v} \in \mathbf{W}$ such that $\mathbf{Re_wwv}$ and $\mathbf{d_v} \notin \mathbf{A_v}$.

Definition: We define $\mathfrak{M} \models_{\mathbf{w}} p$ (p is true at $\mathbf{w}$ in $\mathfrak{M}$) as follows:

- $\mathfrak{M} \models_{\mathbf{w}} Et$ iff $\mathbf{V}(t) \in \mathbf{E_w}$;
- $\mathfrak{M} \models_{\mathbf{w}} Pt$ iff $\mathbf{V_w}(t) \in \mathbf{V_w}(P)$;
- $\mathfrak{M} \models_{\mathbf{w}} \neg p$ iff $\mathfrak{M} \not\models_{\mathbf{w}} p$;
- $\mathfrak{M} \models_{\mathbf{w}} p \wedge q$ iff $\mathfrak{M} \models_{\mathbf{w}} p$ and $\mathfrak{M} \models_{\mathbf{w}} q$;
- $\mathfrak{M} \models_{\mathbf{w}} \forall x p$ iff for all $\mathbf{d} \in \mathbf{E_w}$, $\mathfrak{M}(\mathbf{d}/x) \models_{\mathbf{w}} p$, where $\mathfrak{M}(\mathbf{d}/x)$ is the model which is identical to $\mathfrak{M}$ except that its function $\mathbf{V}$ assigns $\mathbf{d}$ to x;[69]
- $\mathfrak{M} \models_{\mathbf{w}} s = t$ iff $\mathbf{V_w}(s) = \mathbf{V_w}(t)$;
- $\mathfrak{M} \models_{\mathbf{w}} tAp$ iff $\mathbf{V_w}(t) \in \mathbf{A_w}$ and for all $\mathbf{v} \in \mathbf{W}$, if $\mathbf{RV_w}(t)\mathbf{wv}$ then $\mathfrak{M} \models_{\mathbf{v}} p$.

Definition: $\models_i p$ (p is valid in all type i models) iff for all type i models $\mathfrak{M}$, for all $\mathbf{w} \in \mathbf{W}$ (where $\mathbf{W}$ belongs to $\mathfrak{M}$), $\mathfrak{M} \models_{\mathbf{w}} p$. (As always, $0 \leq i \leq 3$.)

Theorem 1: For all i, $0 \leq i \leq 3$: if $\vdash_i p$ then $\models_i p$.

Proof: by calculation.

Theorem 2: $\not\models_0$ (A1). $\not\models_1$ (A2). $\not\models_2$ (A3). $\not\models_3$ (*) and $\not\models_3$ (**). $\mathcal{P}_0$, $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ are therefore consistent.

Proof: It is easy to construct falsifying models in each case. This implies by Theorem 1 that the indicated formulas are not derivable in the respective systems, which are therefore consistent.

Is it possible to axiomatize the notion of validity which has been defined above? Yes, it is. We only have to add the following axiom to all systems $\mathcal{P}_i$, $0 \leq i \leq 3$:

(A4) $(tAp \wedge tAq) \rightarrow tA(p \wedge q)$.

(A4) says that each basic *aisthētikon* integrates its own percepts: there is *intra*sensory integration in addition to the *inter*sensory integration which we have met in (T3) above. This agrees with the texts: cf. *DA* 426b8 ff. and *De sensu* VII. (A4) reflects the fact that there is only one alternativeness relation $\mathbf{R}$ in our models.

Theorem 3: For all i, $0 \leq i \leq 3$: $\models_i p$ iff p is derivable in $\mathcal{P}_i + (\text{A4})$.

Proof: As usual.[70]

The above models are no more than technical tools for checking the consistency of our explanatory calculi. They may not reflect Aristotle's ontological views. It may be possible to devise models which are closer to Aristotle's own views (e.g., models in which both individuals and properties are basic, or models in which tropes are basic, if one wants to regard Aristotle as a trope-theorist), but we will not try to present such models here.[71]

10. Conclusion

The goals we have set ourselves have been reached: we have reconciled the *DA* and *DS* accounts and given a logical analysis of the arguments they contain. All arguments have turned out to make good logical sense, even when they seemed

[69] Thus, we quantify over individual concepts; as noted above, our treatment of quantification is the same as Garson's ((Garson 1978, Garson 1984).

[70] Cf. Gabbay (1976, part I) on *C-2*; Chellas (1980), exercise 3.15 and part III, on *EMC*; Garson (1978) and Garson (1984) on *QS*.

[71] Bacon (1988) contains examples of alternative modellings; on trope-theoretical models, see also Bacon (1989).

suspect at first sight. We have formulated our interpretation in a clear and objective way, with the result that its consistency can be checked. Finally, we have pointed out some applications of his insights. (Recall the refutation of Dawkins, Gunderson and Rucker in Section 2.)

One might complain that the logical reconstruction we have offered is unduly anachronistic, but we think that this complaint would not be justified.

First, the formal systems we have presented are no more than straightforward translations of Aristotle's principles into a symbolic language. It is no more anachronistic to translate them into a formal language than into, say, American English, Cantonese or Volapük. The models from Section 9 are admittedly more anachronistic, but we do not ascribe them to Aristotle; they are no more than technical instruments which we have used to examine his views.

Secondly, modern interpretations of ancient theories are inevitably anachronistic. They are always stated in terms which *we* can understand. For this reason, the most anachronistic reformulations of ancient theories are often the most illuminating and exciting ones. Aristotle's theory that the psyche is the "form" or "shape" of the body had completely disappeared from the philosophical scene until it was realized in the 1970s that it may be regarded as a kind of proto-functionalism.[72] Similarly, the *De memoria* looked like a collection of vague metaphors until the connectionists rediscovered it in the 1980s and saw that it anticipated some of their own most cherished insights.[73] Our commentary had a more modest goal than these sweeping reinterpretations. Nevertheless, it shows that some of Aristotle's insights may be regarded as contributions to the logical study of propositional attitude ascriptions. A proof such as the one in Section 7.2 is directly applicable to some of the multimodal doxastic logics which have recently been proposed in the journal *Artificial Intelligence*.[74] If we had not translated the Stagyrite's insights into the language of symbolic logic—the *lingua franca* of disciplines as wide apart as mathematics, computer science, linguistics and philosophy—we would never have spotted the connection.

Thirdly, the logical approach leads to more insight than any informal one because it forces us to think harder. We cannot afford to be sloppy; we can fool neither the reader nor ourselves. That this has a beneficial effect has repeatedly been demonstrated in the above. No commentator has been able to derive the principle of transparency from more plausible assumptions before; no one has recognized that it is vital to understand the second *DA* argument and the *DS* passage; no one has been able to explain why (A3) makes its first appearance in the *DS*; and no one has ever made sense of the third *DA* argument. Instead, sober commentators like Hamlyn dismissed most of the arguments as muddle-headed nonsense, whereas most of the others tried to convey an impression of understanding by being nebulous.

Finally, Aristotle himself would certainly have been delighted if he had foreseen that some parts of his psychology would one day be scrutinized with logical tools—especially if he had known that they would turn out to be correct in the process.

For the above reasons we think that the logical approach to Aristotle's psychology has much to recommend itself. It is not a perversion of what he said, but only makes his statements more interesting. At the same time, it shows that the tools

[72] On the question whether Aristotle was a functionalist, see, e.g., Modrak (1987) and Shields (1991).

[73] Anderson, Pellionisz & Rosenfeld (1990, pp. 1–13).

[74] Fagin & Halpern (1988).

which have been devised by contemporary logicians may profitably be employed in the area of philosophy.[75]

It is clear that there is still a lot of work which needs be to done along the above lines.[76] The present paper is just a pioneering effort. Despite two millennia of close reading, the study of the Philosopher's writings has not yet neared its completion: it may, in fact, be standing on the brink of an exciting new era.[77]

[75]Some interesting additional remarks which justify our approach to Aristotle's theories are to be found in Code (1976, p. 182).

[76]The *kath' hauta / kata sumbebēkos* distinction seems to be a suitable next target.

[77]The author is indebted to G. Berger, H.-N. Castañeda, M. J. L. Degenaar, P. J. van der Eijk, M. C. Nussbaum, J. Sperna Weiland and J. B. M. van Rijen for useful comments on previous versions of the present paper.

CHAPTER 2

Ontology, Semantics, and Philosophy of Mind in Wittgenstein's *Tractatus*: A Formal Reconstruction

Abstract

This paper presents a formal explication of Wittgenstein's early views on ontology, the syntax and semantics of an ideal logical language, and the propositional attitudes. It is shown that Wittgenstein gave a "language of thought" analysis of propositional attitude ascriptions, and that his ontological views imply that such ascriptions are truth-functions of (and supervenient upon) elementary sentences. Finally, an axiomatization of a quantified doxastic modal logic corresponding to Tractarian semantics is given.

First published in *Erkenntnis* **29** (1988), 35–75. ISSN 0165–0106.

0. Introduction

Historically, Wittgenstein's *Tractatus* is primarily a forerunner of Tarski's and Carnap's later contributions to semantics. However, the latter do not faithfully reflect Wittgenstein's ideas: for example, Wittgenstein's idea that predicates are names of properties is absent from Tarski's work, while Carnap's "state-descriptions" are certainly different from descriptions of states of affairs in the Tractarian sense (cf. Section 3 below). Therefore the question arises: is it possible to develop a semantical system which is both faithful to the *Tractatus* and as precise as Tarski's and Carnap's contributions? This is the question which we shall try to answer in the present paper. The effort will be rewarding: not only will it turn out that it is indeed *possible* to give a formal reconstruction, it will moreover appear that such a reconstruction has various features which are still interesting today. Thus, it not only yields a *truth-functional* analysis of quantification, modalities and propositional attitude ascriptions, it also shows that the *Tractatus* contains a quite modern language of thought theory and even a variant of the currently popular doctrine of psychophysical supervenience.

Our formal reconstruction of the *Tractatus* is not the first one to appear. As early as 1966, Stegmüller—condemning the average interpretation of the *Tractatus* as nothing but "a bunch of very unclear statements, which should first be explicated themselves" (Stegmüller 1966)—gave a formalization of the picture theory; shortly after, Suszko (1968), Wolniewicz (1968) and Mudersbach (1968) began to formalize Tractarian ontology. The formal approach has been taken up by perhaps a dozen philosophers since then. However, none of the previous contributions is wholly successful. First, none of them gives a comprehensive formalization of *both* object ontology *and* situation ontology *and* semantics; as a result, the *interrelations* between such subjects as the independence of states of affairs, the describability of the world by elementary sentences and the principle of truth-functionality have

remained unclear. Secondly, Wittgenstein's remarks on propositional attitude ascriptions have never been discussed in formal terms before,[1] let alone his claim that they are truth-functional or the question whether they are definable in terms of elementary sentences. Finally, all previous reconstructions are rather inelegant. The present reconstruction certainly avoids the first two defects; we hope it avoids the third too.

The plan of the paper is as follows. Because of the primacy of the ontological in the *Tractatus*, we start with this subject in Section 1. Section 2 discusses the syntax and Section 3 the semantics of sentences and pictures; the propositional attitudes are treated separately in Section 4. Section 5 presents the logical system the preceding results lead up to. Finally, the moral will be drawn in Section 6. Comparisons with earlier formalizations will continually be made as we go along.

1. The Ontology of the *Tractatus*

1.1. Objects and States of Affairs. For Wittgenstein, "objects" (*Gegenstände*) are the basic building-blocks of the world. They are the "substance of the world" (TLP[2] 2.021 ff.); all possible worlds have the same substance (TLP 2.022, 2.023, 2.024). The number of objects cannot be determined *a priori*; "it is a matter of physics to find out" (NB p. 127). However, Wittgenstein assumes the existence of at least one object (TLP 2.0211–2.0212, 2.026, 4.2211); on the other hand, he never refers to more than $\aleph_0$ objects (TLP 4.1272; NB p. 127). Denoting the set of objects (*Gegenstände*) by "$\mathbf{G}$", we therefore stipulate:

Definition 1: $\mathbf{G}$ is a set such that $1 \leq \mathit{Card}(\mathbf{G}) \leq \aleph_0$.

It is important to realize that the category of "objects" is a very general one. Relations and properties, if there are such things, are objects too: "Auch Relationen und Eigenschaften etc. sind *Gegenstände*" (NB 16.6.15); "'Objects' also include relations; [...] 'thing' and relation are on the same level" (Lee 1980, p. 120). (This is Wittgenstein's so-called "realism" about relations and properties.) It cannot be settled *a priori* what kinds of objects there are; this can only be found out by empirical investigation, not by logic. Therefore we shall not explicitly distinguish between different kinds of objects and treat them all on a par.

The next step in the Tractarian composition of the world is constituted by "states of affairs" (*Sachverhalte*). States of affairs are concatenations of objects (TLP 2.03; cf. TLP 2.01, 2.0272, 3.21). Wittgenstein seems to have been uncertain as regards the maximum complexity a state of affairs may have. At first, he seems to have accepted only *finite* concatenations of objects: "The *infinitely* complex state of affairs seems to be a monstrosity!" (NB 23.5.15). Later on, however, he seems to have abandoned his repugnance to infinitely complex states of affairs (TLP 4.2211). We adopt the earlier view for the sake of simplicity.

In order to define the set of states of affairs $\mathbf{SA}$, we first introduce the set $\mathbf{G^*}$ of all finite concatenations of members of $\mathbf{G}$.

Definition 2: $\mathbf{G^*}$ is the smallest set such that:

(a) if $\mathbf{g}, \mathbf{g}' \in \mathbf{G}$, then $\mathbf{g} * \mathbf{g}' \in \mathbf{G^*}$;

(b) if $\mathbf{g} \in \mathbf{G}$ and $\mathbf{s} \in \mathbf{G^*}$, then $\mathbf{g} * \mathbf{s} \in \mathbf{G^*}$.

Notice that $\mathit{Card}(\mathbf{G^*}) = \aleph_0$, *even if* $\mathit{Card}(\boldsymbol{G}) = 1$! Therefore Suszko (1968, p. 24) (following Wolniewicz) made an error in claiming that "if there were finitely many objects then there would exist only finitely many configurations of them, i.e., finitely many states of affairs".

---

[1] Exceptions are Lokhorst (1985 *a*), Lokhorst (1985 *b*), the precursors of the present paper.

[2] Here and in the following, "TLP" stands for the *Tractatus* (Wittgenstein 1971). "NB" stands for the second edition of the *Notebooks*, including the appendices (Wittgenstein 1979). We shall occasionally provide our own translations.

G* is, in general, *not* the set of states of affairs, as might be supposed. Certain additional restrictions may exclude some concatenations of objects from being states of affairs. For example, if $\mathbf{f}$ is a property and $\mathbf{g}$ a particular ($\mathbf{f}$, $\mathbf{g} \in \mathbf{G}$), then $\mathbf{f} * \mathbf{g}$ may well be a state of affairs, viz., the situation that $\mathbf{g}$ and $\mathbf{f}$ are concatenated, or the situation that $\mathbf{g}$ has property $\mathbf{f}$. But in this case, $\mathbf{g} * \mathbf{f}$ will presumably not be a situation at all. (It might be one if $\mathbf{g}$ were a second-order property.) The same goes for relations. If $\mathbf{R} \in \mathbf{G}$ is an n-ary relation, then $\mathbf{R} * \mathbf{g}_0 * \ldots * \mathbf{g}_n$ is the state of affairs that $\mathbf{R}$, $\mathbf{g}_0$, ... , $\mathbf{g}_n$ are concatenated, the situation that $\mathbf{R}$ is exemplified by $\mathbf{g}_0$, ... , $\mathbf{g}_n$, or, as Suszko (1968, p. 22) expresses it, the $\mathbf{R}$-configuration of $\mathbf{g}_0$, ... , $\mathbf{g}_n$. In this case, $\mathbf{g}_0 * \mathbf{R} * \mathbf{g}_1$ will presumably not be a situation at all.

However, we cannot give an *a priori* list of conditions an element of **G*** must meet if it is to be counted as a "well-formed" state of affairs: we do not even know, for example, whether there are binary relations or not, for this is an empirical matter. Therefore we simply stipulate that $\mathbf{SA} \subseteq \mathbf{G^*}$.[3]

What is the cardinality of **SA**? In the first place, $\mathbf{SA} \neq \emptyset$. This follows from the requirement in TLP 2.011 that each object occurs in some situation. Since $\mathbf{G} \neq \emptyset$, **SA** cannot be empty either.

However, we can do better than this by taking TLP 4.463 into account: here mention is made of "infinite logical space". As logical space is generated by **SA** (Section 1.3), **SA** must be infinite too. Therefore we stipulate:

Definition 3: $\mathbf{SA} \subseteq \mathbf{G^*}$ is a set such that:

(a) for each $\mathbf{g} \in \mathbf{G}$ there is at least one $\mathbf{s} \in \mathbf{SA}$ such that $\mathbf{s} = \mathbf{g}_0 * \ldots * \mathbf{g}_n$ and $\mathbf{g} = \mathbf{g}_i$ for some i, $0 \leq i \leq n$, and

(b) $Card(\mathbf{SA}) = \aleph_0$.

Clause (b) is important in connection with propositional attitude ascriptions: it implies that these are, in general, not definable in terms of elementary sentences (see note 27 below).

1.2. Situations, Facts and Worlds. The essential clue to understand Tractarian situation ontology has been provided by Suszko (1968): Tractarian situations are the elements of a complete atomic Boolean algebra. We shall adopt this suggestion, but turn Suszko's algebra "upside down" (i.e., consider its dual) as this leads to a more natural conception of possible worlds. The latter now become the "mereological sums" (suprema) of the possible situations they contain as "parts".[4] Given **SA**, an algebra of situations $\mathcal{S}$ is therefore defined as follows:

Definition 4: $\mathcal{S} = \langle \mathbf{S}, \sqcup, \sqcap, -, \mathbf{1}, \mathbf{0} \rangle$ is a complete atomic Boolean algebra such that $\mathbf{SA} \subseteq \mathbf{S}$.

- **S**, the universe of $\mathcal{S}$, is Wittgenstein's "logical space" (*logischer Raum*). The elements of **S** are called "situations" (*Sachlagen*) or "possible situations".
- "$\sqcup$" stands for "supremum" (least upper bound). "Totality" (*Gesamtheit*) is comprehensible if it is read as "supremum" up to TLP 3 and as "set" from then on.
- "$\sqcap$" stands for "infimum" (greatest lower bound), while "$-$" stands for "complement".

[3] At this point, various "forms" may be introduced. In view of TLP 2.0141, the *Form des Gegenstandes* is: $FG(\mathbf{g}) = \{\langle i, \mathbf{g}_0 * \ldots * \mathbf{g}_i \rangle : n \geq i \text{ and } \mathbf{g} = \mathbf{g}_i\}$. A similar definition has been given by Mudersbach (1978). By Definition 3, clause (a), $FG(\mathbf{g}) \neq \emptyset$ (for any $\mathbf{g} \in \mathbf{G}$). This is Mudersbach's Axiom 5. The *Form des Sachverhaltes* is: $FS(\mathbf{g}_0 * \ldots * \mathbf{g}_n) = \{\langle i, FG(\mathbf{g}_i) \rangle : i \leq n\}$. Cf. the definition of the "structure" of **s** by Czermak (1979). The point which is made in the text is that $FG(\mathbf{g})$ and $FS(\mathbf{s})$ may be *proper* subsets of $\mathscr{P}(N \times \mathbf{G^*})$ and $\mathscr{P}(N \times \mathscr{P}(N \times \mathbf{G^*}))$, respectively (where "$\mathscr{P}$" denotes the power-set), but that it cannot be determined in advance *which* subsets they are. Forms are not *a priori*.

[4] Any Boolean algebra may be regarded as a mereology (theory of parts and wholes). A mereological view of the *Tractatus* has also been argued for by Simons (1986).

- $\mathbf{1}$ is the impossible situation, $\mathbf{0}$ the necessary situation. These are the two improper (*uneigentliche*) situations.
- A situation $\mathbf{s}$ "exists" (*besteht*) in a situation $\mathbf{s}'$ iff $\mathbf{s} \sqsubseteq \mathbf{s}'$. Synonyms: "$\mathbf{s}$ is the case (*ist der Fall*) at $\mathbf{s}'$", "$\mathbf{s}$ is contained (*ist enthalten*) in $\mathbf{s}'$".
- A dual atom of $\mathbf{S}$ is called a "possible world" (*mögliche Welt*, NB 19.9.16; cf. TLP 2.022) or "world" (*Welt*, TLP *passim*) for short. The set of dual atoms will be denoted by "$\mathbf{W}$".[5]
- Some (arbitrary) element $\mathbf{w}_0 \in \mathbf{W}$ is "the" world, the "actual world" (*die wirkliche Welt*, TLP 2.022), "our world" (*unsere Welt*, TLP 6.1233), "the world in which we live" (*die Welt worin wir leben*, NB p. 127).[6]
- A situation "exists" or "is the case" *simpliciter* iff it exists (is the case) in $\mathbf{w}_0$.
- A "fact" (*Tatsache*) is a situation which is the case.[7]

Using the above informal paraphrases of our technical terms, some features of Tractarian situation ontology can already be given a precise interpretation. For example, it is an elementary thesis of Boolean algebra that $\mathbf{w}_0 = \bigsqcup\{\mathbf{s} \in \mathbf{S} : \mathbf{s} \sqsubseteq \mathbf{w}_0\}$; translation in informal terms yields "Die Welt ist die Gesamtheit der Tatsachen", i.e., TLP 1.1. More or less the same is stated in TLP 1 and 1.11–1.2, (which shows that the beginning of the *Tractatus* is rather repetitive, as Menger (1980) already pointed out in his plea for a formal analysis of the work).

1.3. Complete Sets of States of Affairs. A *complete* set of situations is a set which contains, for every situation, either this situation or its complement, but not both, and no other elements. We require that $\mathcal{S}$ satisfies the following additional conditions:

Condition 1: For every complete set of states of affairs $\mathbf{K}$: $\bigsqcup \mathbf{K} \neq \mathbf{1}$.

Condition 2: If $\mathbf{w}$, $\mathbf{w}' \in \mathbf{W}$ and $\mathbf{w} \neq \mathbf{w}'$, then there is at least one $\mathbf{s} \in \mathbf{SA}$ such that $\mathbf{s} \sqsubseteq \mathbf{w}$ and $\mathbf{s} \not\sqsubseteq \mathbf{w}'$.

Condition 1 states that all states of affairs are *independent* in the sense of Boolean algebra. This seems to be a good explication of the Tractarian thesis that the states of affairs are independent (*unabhängig*, TLP 2.061): it enables us to prove such passages as TLP 2.062 and 4.27, which have baffled many commentators. Consider 2.062 as an example (we shall discuss 4.27 in a moment). Here it is said that the existence of a state of affairs cannot be inferred from the existence of another state of affairs. This is easily provable: assume one *could* do the latter, i.e., that $\mathbf{s} \sqsubseteq \mathbf{s}'$ for some $\mathbf{s}$, $\mathbf{s}' \in \mathbf{SA}$. Then $\bigsqcup\{-\mathbf{s}, \mathbf{s}'\} = \mathbf{1}$, which contradicts Condition 1; hence the assumption is false, Q.E.D. As we shall see, all Tractarian remarks on the independence of elementary sentences may also be proved using Condition 1.

At this point, Condition 2 cannot yet be very well justified. However, it yields one nearly Tractarian thesis: "Die Gesamtheit der bestehenden Sachverhalte *bestimmt* die Welt"; this is TLP 2.04 upon emendation along the lines indicated by

[5] It is no anachronism to discuss the *Tractatus* in terms of modern "possible worlds": the latter notion was introduced in twentieth-century philosophy by Carnap (1956, p. 91), who took it in his turn directly from the *Tractatus*.

[6] The relation between the terms "world" and "reality" (*die Wirklichkeit*) is hard to understand. Czermak (1979) suggests interpreting "reality" as the partition $\langle\{\mathbf{s} \in \mathbf{SA} : \mathbf{s} \sqsubseteq \mathbf{w}_0\}, \{\mathbf{s} \in \mathbf{SA} : \mathbf{s} \not\sqsubseteq \mathbf{w}_0\}\rangle$ (cf. TLP 2.06, 4.0621); but "reality" may sometimes mean "logical space" or "the world" as well.

[7] However, "fact" is sometimes used in the sense of "concatenation" (see note 11); in this case the term may refer to a non-existent situation. In "Komplex und Tatsache" (ca. 1931; repr. in Wittgenstein (1964)) the requirement that facts exist is explicitly dropped.

Griffin[8]. Moreover, there is indirect reason to accept Condition 2: as we shall see in Section 3.5, it has the consequence that each possible world is completely describable by elementary sentences, which is certainly a prominent thesis of the *Tractatus*.

By Boolean algebra, the conjunction of Conditions 1 and 2 is equivalent to:

Condition 3: For every complete set of states of affairs $\mathbf{K}$: $\bigsqcup\mathbf{K} \in \mathbf{W}$.[9]

It follows from Condition 3 that **SA** is a set of generators of $\mathcal{S}$. Therefore, $\mathcal{S}$ may briefly be characterized as a *complete atomic Boolean algebra independently generated by* **SA**. It follows that $Card(\mathbf{W}) = 2^{Card(\mathbf{SA})} = 2^{\aleph_0}$, while $Card(\mathbf{S}) = 2^{Card(\mathbf{W})}$, as has also been concluded by Suszko (1968, p. 21). The *Tractatus* explicitly mentions the finite analogue of the first property of $\mathcal{S}$: n states of affairs generate 2^n worlds (TLP 4.27). It does not mention the second property.

1.4. Summary. Recapitulating this section, we say that a *Tractarian ontological system* is a quadruple $\Sigma = \langle \mathbf{G}, \mathbf{SA}, \mathcal{S}, \mathbf{w}_0 \rangle$ satisfying Definitions 1–4 and Condition 3.

2. Syntax of Sentences, Thoughts and Pictures

2.1. Syntax of Sentences. The building-blocks of sentences (*Sätze*) are names (*Namen*: TLP 3.202, 3.26, 4.0311, 4.22, 5.55). The category of names is a very general one. For example, no explicit distinction is made between names of individuals (particulars) and predicates: predicates are simply names too, viz., names of properties and relations. As NB 31.5.15 says: "[Two] names are necessary for an assertion that *this* thing possesses *that* property". Therefore we shall not explicitly distinguish between, say, "proper" names (designating individuals) and predicates, but treat them on the same footing (cf. Section 1.1). As there is a bijection from $\boldsymbol{N}$ to some $\mathbf{G}$ (Definition 15b), the cardinality of $\boldsymbol{N}$ satisfies the same restrictions as that of $\mathbf{G}$. Thus (cf. Definition 1):

Definition 5: $\boldsymbol{N}$ is a set such that $1 \leq Card(\boldsymbol{N}) \leq \aleph_0$.

Just as states of affairs are concatenations of objects, so elementary sentences (*Elementarsätze*) are concatenations of names (TLP 4.22; cf. TLP 3.14, 3.21, 4.221). Designating the set of elementary sentences by "$\boldsymbol{EL}$" (for "elementary language"), we accordingly have (cf. Definitions 2, 3):

Definition 6: . $\boldsymbol{N^*}$ is the smallest set such that:

(a) if $a, a' \in \boldsymbol{N}$ then $a * a' \in \boldsymbol{N^*}$;

(b) if $a \in \boldsymbol{N}$ and $p \in \boldsymbol{N^*}$, then $a * p \in \boldsymbol{N^*}$.

Definition 7: $\boldsymbol{EL} \subseteq \boldsymbol{N^*}$ is a set such that:

(a) For each $a \in \boldsymbol{N}$ there is a $p = a_0 * \ldots * a_n \in \boldsymbol{EL}$ such that $a = a_i$ for some i, $0 \leq i \leq n$, and

(b) $Card(\boldsymbol{EL}) = \aleph_0$.

The above definitions do not uniquely specify one set of elementary sentences; rather, they specify the broad conditions any such set must comply with. Because the structure of language reflects the structure of reality, syntax cannot be fully specified *a priori*. As TLP 5.55 says: "Since [...] we are unable to give the

[8]Griffin (1965, ch. 5). Emendation of TLP 2.04 is needed anyway, because it cannot be brought into line with TLP 2.06 and 2.063 otherwise. The emended version may be compared with TLP 2.05 and 4.26–4.28.

[9]Proof: by Condition 2 it cannot be the case that there are two different $\mathbf{w}$, $\mathbf{w}' \in \mathbf{W}$ such that $\bigsqcup\mathbf{K} \sqsubseteq \mathbf{w}$ and $\bigsqcup\mathbf{K} \sqsubseteq \mathbf{w}'$ (where $\mathbf{K}$ is a complete set of states of affairs). Therefore $\bigsqcup\mathbf{K} \in \mathbf{W}$ or $\bigsqcup\mathbf{K} = \mathbf{1}$; hence $\bigsqcup\mathbf{K} \in \mathbf{W}$ by Condition 1. That Condition 3 implies Conditions 1 and 2 is obvious.

number of names with different meanings, we are also unable to give the composition of elementary sentences".

Suppose, for example, that there are both individuals (particulars) and properties. Then we may distinguish between "proper" names (names of objects of the former kind) and predicates (names of objects of the latter kind); but in this case presumably not all concatenations of names will be well-formed elementary sentences, for not each concatenation will correspond to a similarly structured state of affairs. For example, if F is a predicate and a a proper name ($F, a \in \boldsymbol{N}$), then $F * a$ may well be an elementary sentence, but $a * F$ will presumably be as ill-formed as the situation $\mathbf{g} * \mathbf{f}$ from Section 1.1. (This is not to say that predicates cannot be predicated in turn: we may have $G * F$ ($G, F \in \boldsymbol{N}$), where G is a second-order predicate.) Thus, the syntax of elementary sentences parallels the structure of states of affairs. Syntactical form mirrors ontological form.[10] As the latter is not *a priori* determinable, the former is not either.

According to the *Tractatus*, all sentences are built from the elementary sentences by means of the operation of joint negation N (TLP 5.5–5.51, 5.52, 6.001). Much fuss has been made over this operator, especially in connection with the Tractarian account of quantification; a good discussion is Soames (1983). Our solution will be simpler than Soames's in that we shall allow arbitrary *countable* sets of sentences as arguments for joint negation. Unfortunately, this does not agree well with TLP 5.32, where it is asserted that "All truth-functions are results of the successive application to elementary sentences of a finite number of truth-operations" (where "truth-operation" means "connective": see our discussion of TLP 5.54 in Section 4.1 below). However, our solution not only makes it possible to define quantification in terms of joint negation, it also enables us to express the independence of the elementary sentences and the principle of truth-functionality within our language (see Sections 3.4–3.6). So let us ignore TLP 5.32 and run the risk of making the *Tractatus* more interesting than it actually is!

In order to formulate the just-mentioned principles, we additionally need one other operator which is not *expressis verbis* to be found in the *Tractatus*: the unary modal connective $\Box$ (for "it is necessary that").

Thus, given some set $\boldsymbol{EL}$, the language $\boldsymbol{L}$ is defined as follows:

Definition 8: . $\boldsymbol{L}$ is the smallest set such that:

(a) $\boldsymbol{EL} \subseteq \boldsymbol{L}$,
(b) if $P \subseteq \boldsymbol{L}$, then $NP \in \boldsymbol{L}$, provided $1 \leq Card(P) \leq \aleph_0$,
(c) if $p \in \boldsymbol{L}$, then $\Box p \in \boldsymbol{L}$, and
(d) If $p \in \boldsymbol{L}$, then $(x)px \in \boldsymbol{L}$, where px is like p except that at least one occurrence of some name occurring in p has been replaced by x.

Negation, countable conjunction, the connective $\Diamond$ (for "it is possible that") and the universal quantifier (x) may be defined as follows (therefore the clause (d) was in fact superfluous):

- $\neg p = N\{p\}$.
- $\bigwedge P = N\{\neg p : p \in P\}$.
- $\Diamond p = \neg\Box\neg p$.
- $(x)px = \bigwedge\{p[a/x] : a \in \boldsymbol{N}\}$, where $p[a/x]$ is like px except that all free occurrences of x in px have been replaced by a.

[10] Various "syntactical forms" may be defined in the same way as the ontological forms of note 3. The *Form des Bildelementes FB* of an element $e \in \boldsymbol{E}$ [$\boldsymbol{E} \supseteq \boldsymbol{N}$; see Section 2.3] is wholly analogous to *FG*, and the *Form der Abbildung FA* of an element $b \in \boldsymbol{EB}$ [$\boldsymbol{EB} \supseteq \boldsymbol{EL}$; see Section 2.3] is wholly analogous to *FS*. Similarly to $FG(\mathbf{g})$, $FB(e) \neq \emptyset$ by def. 7a; but for the rest, *FB* and *FA* are no more *a priori* than *FG* and *FS* are. The *Form der Darstellung* or *Form des Zusammenhangs* is the same as the *abbildende Beziehung*, which is defined in Section 3.2.

Notice that variables may range over objects, properties, second-order properties, etc.; we cannot settle *a priori* what they range over since language depends on ontology. Therefore, we do not know *a priori* of what order Tractarian logic is; we only know that its order must be smaller than $Card(\boldsymbol{N})$. (It does not have to be 2, as Skyrms (1981) supposes.)

This completes the description of the construction of the ideal logical language $\boldsymbol{L}$ out of $\boldsymbol{EL}$. It will be seen that $\boldsymbol{L}$ does not contain an identity-sign, which is as it should be, for this is explicitly forbidden in TLP 5.53–5.5352 (see Section 3.1).

2.2. Syntax of Thoughts. Thoughts (*Gedanken*) are similar to sentences: "Thinking is a kind of language. [...] A thought is a kind of sentence" (NB 12.9.1916; cf. TLP 4). Analogously to sentences, thoughts are constructed from "psychical constituents that have the same sort of relation to reality as words" (NB p. 130). Denoting the set of thought-elements ("mental names") by $\boldsymbol{TE}$ and the set of elementary thoughts by $\boldsymbol{ET}$, we therefore stipulate that (cf. Definitions 5–7):

Definition 9: $\boldsymbol{TE}$ is a set such that $Card(\boldsymbol{TE}) \leq \aleph_0$.

Definition 10: $\boldsymbol{TE^*}$ is the smallest set such that:

(a) if $e, e' \in \boldsymbol{TE}$ then $e * e' \in \boldsymbol{TE^*}$;

(b) if $e \in \boldsymbol{TE}$ and $t \in \boldsymbol{TE^*}$, then $e * t \in \boldsymbol{TE^*}$.

Definition 11: $\boldsymbol{ET} \subseteq \boldsymbol{TE^*}$ is a set such that:

(a) For each $e \in \boldsymbol{TE}$ there is a $t = e_0 * \ldots * e_n \in \boldsymbol{ET}$ such that $e = e_i$ for some i, $0 \leq i \leq n$, and

(b) $Card(\boldsymbol{ET}) = \aleph_0$.

Given $\boldsymbol{ET}$, the set of thoughts ("language of thought") $\boldsymbol{T}$ is constructed as follows (cf. Definition 8):

Definition 12: $\boldsymbol{T}$ is the smallest set such that:

(a) $\boldsymbol{ET} \subseteq \boldsymbol{T}$,

(b) if $T \subseteq \boldsymbol{T}$, then $NT \in \boldsymbol{T}$, provided $1 \leq Card(\boldsymbol{T}) \leq \aleph_0$,

(c) if $t \in \boldsymbol{T}$, then $\Box t \in \boldsymbol{T}$, and

(d) if $t \in \boldsymbol{T}$, then $(y)ty \in \boldsymbol{T}$, where ty is like t except that at least one occurrence of some thought-element occurring in t has been replaced by y.

We shall return to thoughts in Sections 4.2 ff.

2.3. Syntax of Pictures. As the *Tractatus* leaves us completely in the dark with regard to the structure of pictures in general, we shall consider no other pictures than sentences and thoughts. The sets $\boldsymbol{E}$, $\boldsymbol{EB}$ and $\boldsymbol{B}$ of pictorial elements (*Bildelemente*, TLP 2.1514), elementary pictures (*Elementarbilder*, a term not to be found in the *Tractatus*), and pictures (*Bilder*), respectively, are therefore defined as follows:

Definition 13: $\boldsymbol{E}$, $\boldsymbol{EB}$ and $\boldsymbol{B}$ are sets such that $\boldsymbol{E} = \boldsymbol{N} \cup \boldsymbol{TE}$, $\boldsymbol{EB} = \boldsymbol{EL} \cup \boldsymbol{ET}$, and $\boldsymbol{B} = \boldsymbol{L} \cup \boldsymbol{T}$.

2.4. Summary. Recapitulating this section, we say that a *Tractarian pictorial system* is a 9-tuple $\Pi = \langle \boldsymbol{E}, \boldsymbol{TE}, \boldsymbol{N}, \boldsymbol{EB}, \boldsymbol{ET}, \boldsymbol{EL}, \boldsymbol{B}, \boldsymbol{T}, \boldsymbol{L} \rangle$ satisfying Definitions 5–13 above.

3. Pictorial and Linguistic Representation

3.1. Basic Picture Theory.

Definition 14: A Tractarian interpretation for a pictorial system

$$\Pi = \langle \boldsymbol{E}, \boldsymbol{TE}, \boldsymbol{N}, \boldsymbol{EB}, \boldsymbol{ET}, \boldsymbol{EL}, \boldsymbol{B}, \boldsymbol{T}, \boldsymbol{L} \rangle$$

as described in Section 2.4 is a pair $I = \langle \Sigma, \delta \rangle$ such that:

(a) $\Sigma = \langle \mathbf{G}, \mathbf{SA}, \mathcal{S}, \mathbf{w}_0 \rangle$ is a Tractarian ontological system as described in Section 1.4;
(b) $\delta : \boldsymbol{E} \mapsto \mathbf{G}$ is a function such that $\delta \restriction \boldsymbol{N}$ is a bijection; and
(c) $\boldsymbol{B} \subseteq \mathbf{S}$.

The above provides the basis of the picture theory. $\delta(e)$ is the denotation (*Bedeutung*, "meaning" in NB) of e. When $\mathbf{g} = \delta(e)$, we say that e denotes or stands for (*steht für*, *bedeutet*, *vertritt*) $\mathbf{g}$ (TLP 3.203–3.221, 3.323, 4.0311, 4.0312). In this case, $\mathbf{g}$ is the object corresponding to or correlated with the pictorial element e (*der dem Bildelement entsprechende, zugeordnete Gegenstand*: cf. TLP 2.13, 2.1514, 5.526). δ is *not* a function of the *Sachlage* $\mathbf{s}$ under consideration: Tractarian names are rigid designators (cf. Cocchiarella (1984), Soames (1983)).

In clause (b), "$\delta \restriction \boldsymbol{N}$ is a bijection" means that every object $\mathbf{g} \in \mathbf{G}$ has precisely one name. This follows from:

(i) $\delta \restriction \boldsymbol{N}$ is a surjection, i.e., for every $\mathbf{g} \in \mathbf{G}$ there is at least one $a \in \boldsymbol{N}$ such that $\mathbf{g} = \delta(a)$; otherwise there would be unnamed objects and hence indescribable situations, which contradicts TLP 4.26 (see Section 3.5).
(ii) $\delta \restriction \boldsymbol{N}$ is an injection, i.e., for every $\mathbf{g} \in \mathbf{G}$ there is at most one $a \in \boldsymbol{N}$ such that $\mathbf{g} = \delta(a)$. This is Wittgenstein's famous identity-theory, clearly expressed in TLP 5.53: "Identity of object I express by identity of sign, and not by using a sign for identity. Difference of objects I express by difference of signs." (See TLP 5.53–5.5352.)

Clause (c) expresses the Tractarian thesis that every picture is a situation.[11] Some situations may be regarded from two different points of view: they may be regarded as situations in their own right (in which case they will appear as, e.g., concatenations of objects), or they may be regarded as pictures (in which case they will appear as, e.g., concatenations of pictorial elements). There is no conflict between these two perspectives; the identification of pictures with situations is inconsequential from a semantical point of view.[12]

3.2. Senses. On the basis of δ a function $\sigma : \boldsymbol{B} \mapsto \mathbf{S}$ is defined as follows:

Definition 15: $\sigma : \boldsymbol{B} \mapsto \mathbf{S}$ is a function such that:
(a) If $e = e_0 * \ldots * e_n \in \boldsymbol{EB}$, then $\sigma(e) = \delta(e_0) * \ldots * \delta(e_n)$,
(b) $\sigma(NP) = \bigsqcup\{-\sigma(b) : b \in P\}$ (where $P \subseteq \boldsymbol{B}$), and
(c) $\sigma(\Box b) = \mathbf{0}$ if $\sigma(b) = \mathbf{0}$, $\sigma(\Box b) = \mathbf{1}$ otherwise (where $b \in \boldsymbol{B}$).

σ(b) is the sense (*Sinn*) of b ("Das Bild stellt eine mögliche Sachlage im logischen Raume [...] dar", TLP 2.202; cf. 2.11, 2.221). When $\mathbf{s} = \sigma(b)$, we say that $\mathbf{s}$ is "represented" (*dargestellt*, *abgebildet*) by b, that b "shows" (*zeigt*) $\mathbf{s}$, and that "b says (that) $\mathbf{s}$ is the case" (*b sagt, daß* $\boldsymbol{s}$ *der Fall ist*). As TLP 4.022 says: "The sentence *shows* how things stand *if* it is true. And it *says that* they do so stand". σ itself may be called the "pictorial relationship" (*abbildende, darstellende Beziehung*, TLP 2.1513–4).

Clause (a) is a succinct formulation of the picture theory for elementary pictures (cf. TLP 2.15, 3.1432, 4.0311, 4.21). As TLP 2.1514 says, in the case of elementary

[11] Actually, TLP 2.141 says that every picture is a *fact* (this is repeated in 2.14 for sentences). However, requiring every picture to be an existing situation would seem unduly stringent. In our opinion, 2.14 and 2.141 primarily draw attention to the similarity of structure between states of affairs and elementary pictures: both are concatenations (of objects and pictorial elements, respectively). These passages do not expressly mean to say that every picture exists as part of $\mathbf{w}_0$. We could require the latter, but as nothing seems to be gained by this we take *Tatsache* here as *mögliche Tatsache* (cf. note 7).

[12] In order to clarify matters further we might introduce a function π mapping situations-as-situations onto situations-as-pictures, and a function π' mapping situations-as-pictures into situations-as-situations (cf. Favrholdt (1964)). π is not an injection, for different situations may be the same from the pictorial point of view: "'A' is the same sign as 'A'" (TLP 3.203).

pictures, "the *abbildende Beziehung* [σ] consists of the correlations [*Zuordnungen*] of the picture's elements with objects". Supposing that $e_0 = R$ is an n-ary predicate and $\mathbf{g}_0 = \mathbf{R}$ an n-ary relation, we see that if $\mathbf{g}_i = \delta(e_i)$ for i, $0 \leq i \leq n$, then the R-configuration of $e_1, \ldots, e_n$ represents the $\mathbf{R}$-configuration of $\mathbf{g}_1, \ldots, \mathbf{g}_n$. Or to put it differently, let $R' = \{\langle e_i, \ldots, e_{i+n}\rangle \in \boldsymbol{N}^n : R * e_i * \ldots * e_{i+n} \in \boldsymbol{EB}\}$ and $\mathbf{R}' = \{\langle \mathbf{g}_i, \ldots, \mathbf{g}_{i+1}\rangle \in \mathbf{G}^n : \mathbf{R} * \mathbf{g}_i * \ldots * \mathbf{g}_{i+1} \in \mathbf{SA}\}$: then the "fact" that $e_i, \ldots, e_{i+n}$ stand in the relation R' says that $\mathbf{g}_i, \ldots, \mathbf{g}_{i+1}$ stand in the relation $\mathbf{R}'$—which is exactly what the notorious TLP 3.1432 affirms. In conjunction with Definition 14b, Definition 15a implies that there is a 1-1 correspondence between elementary sentences and states of affairs (i.e., $\sigma \restriction \boldsymbol{EL} \mapsto \mathbf{SA}$ is a bijection).

Clause (b) is equivalent to the following two claims taken together:

(i) $\sigma(\neg b) = -\sigma(b)$, and
(ii) $\sigma(\bigwedge P) = \bigsqcup\{\sigma(b) : b \in P\}$.

Both formulae are in a general way justified by TLP 5.2341: "The sense of a truth-function of p is a function of the sense of p". More specifically, (i) is justified by TLP 5.2341 ("Negation reverses the sense of the sentence") and by TLP 4.0621 ("The sentences p and $\neg p$ have opposite sense"); it also explains why p and $\neg\neg p$ have the same sense (TLP 4.0621). Justification for (ii) is harder to find; the *Tractatus* is silent on the semantics of conjunctions. However, NB 5.6.15 asserts that "$p \wedge \neg p$ is that thing [...] that p and $\neg p$ have in common". According to our formalization, the sense of $\bigwedge P$ is, indeed, the greatest common part (infimum) of the senses of the elements of P, which seems an acceptable way to give this assertion its due. It will be noted that if $p \notin \boldsymbol{EL}$, then $\sigma(p)$ is not a *new* element of $\mathbf{S}$ (if $\sigma(p) = \mathbf{s}$, $\mathbf{s}$ already belonged to $\mathbf{S}$), which is precisely what is asserted in TLP 3.42.

Because the senses of contradictions and tautologies are improper situations (for $\sigma(p \wedge \neg p) = \sigma(p) \sqcup -\sigma(p) = \mathbf{1}$, while $\sigma(p \vee \neg p) = \sigma(p) \sqcap -\sigma(p) = \mathbf{0}$), they are improper pictures themselves (cf. TLP 4.462) and they may even be called "senseless" (*sinnlos*, TLP 4.461). But as TLP 4.4611 emphasizes, contradictions and tautologies are not "nonsensical" (*unsinnig*): this is a term reserved for pseudosentences, i.e., for sentence-like entities which do not belong to $\boldsymbol{L}$ at all. For example, the metalinguistic assertions of "the ladder-language" in which the *Tractatus* discusses object language are *unsinnig*—which implies that the whole *Tractatus* is *unsinnig* (a conclusion which is, indeed, drawn in TLP 6.54).

Clause (c) introduces an $S5$-like semantical analysis of modal sentences. We adopt the analysis by von Wright (1982), which singles out $S5$ as the correct formalization of the notion of modality in the *Tractatus*. Von Wright's analysis is not without its critics; for example, Perzanowski (1985) regards several other modal logics as more suitable for this role. However, a strong point in favour of $S5$ is that its semantics make clear (as the semantics of other modal logics do not) why modal sentences do not violate the principle that the world is completely describable by elementary sentences (Section 3.5). As a consequence these sentences do not violate the principle of truth-functionality either (Section 3.6).[13] In view of the important role of both principles in the *Tractatus*, $S5$ seems to be the modal logic which agrees best with the *Tractatus*.

3.3. Truth. A function $TV : \boldsymbol{B} \times \mathbf{S} \mapsto \{T, F\}$ assigning a truth-value T (true) or F (false) to b at $\mathbf{s}$ is defined as follows:

Definition 16: $TV(b, \mathbf{s}) = T$ if $\sigma(b) \sqsubseteq \mathbf{s}$; $TV(b, \mathbf{s}) = F$ otherwise.

[13] For this reason $S5$ is sometimes, rather confusingly, called an "extensional" modal logic, e.g., by Perzanowski (1985). We shall see that truth-functionality and extensionality must be sharply distinguished (Section 3.6).

A picture b is said to be true (*simpliciter*) iff it is true at $\mathbf{w}_0$. Thus, a picture is true iff its sense exists, or in other words, it is true iff it says that $\mathbf{s}$ is the case and $\mathbf{s}$ is indeed the case (cf. the definitions of "exists", "is the case" and "says that" in Section 1.2 and Section 3.2). A picture is said to be valid (in the interpretation under consideration) iff $TV(b, \mathbf{s}) = T$ for all $\mathbf{s} \in \mathbf{S}$ (in this interpretation).

Definition 16 is an extension of TLP 4.25 to pictures and situations in general: "If the elementary sentence is true the [corresponding] state of affairs exists; if the elementary sentence is false the [corresponding] state of affairs does not exist" (cf. TLP 4.21). It is evident that there is a 1-1 correspondence between senses and partitions $\langle\{\mathbf{s} \in \mathbf{S} : TV(b, \mathbf{s}) = T\}, \{\mathbf{s} \in \mathbf{S} : TV(b, \mathbf{s}) \neq T\}\rangle$ of logical space $\mathbf{S}$, which is in eminent agreement with TLP 2.11, 2.201, 4.1 and also explains TLP 4.024: "To understand a sentence means to know what is the case if it is true". (For knowing $\sigma(p)$ amounts to knowing the corresponding partition.) Specifying the (actual) truth-value of a sentence serves to narrow down the range (*Spielraum*) the "logical place" (*logischer Ort*) of $\mathbf{w}_0$ may occupy (TLP 4.463, 5.5262). This function is not fulfilled by tautologies and contradictions. As these determine the partitions $\langle\mathbf{S}, \emptyset\rangle$ and $\langle\emptyset, \mathbf{S}\rangle$, respectively, the former leave the whole subset $\mathbf{W}$ of logical space $\mathbf{S}$ open to the world, while the latter leave no point for it at all (TLP 4.46–4.4611).

Some observations (provable by elementary Boolean algebra):

(a) $TV(e_0 * \ldots * e_n, \mathbf{s}) = T$ iff $\mathbf{g}_0 * \ldots * \mathbf{g}_n \sqsubseteq \mathbf{s}$, where $\mathbf{g}_i = \delta(e_i)$ for all i, $0 \leq i \leq n$ (see TLP 4.21). It is evident that this truth-condition is hardly "related to the Tarski-type truth-definition for atomic sentences" (the latter claim has been made by Hintikka & Hintikka (1983, p. 158)).

(b) $TV(\neg b, \mathbf{w}) = T$ iff $TV(b, \mathbf{w}) = F$. This holds for worlds, but *not* for all situations: we have $TV(b, \mathbf{s}) = T$ for all b iff $\mathbf{s} = \mathbf{1}$, while we may have $TV(b, \mathbf{s}) = TV(\neg b, \mathbf{s}) = F$ if $\mathbf{s} \notin \mathbf{W}$.

(c) $TV(\bigwedge P, \mathbf{s}) = T$ iff $TV(b, \mathbf{s}) = T$ for all $b \in P$, and similarly for universally quantified sentences (or thoughts).

(d) b is valid iff $\sigma(b) = \mathbf{0}$.

(e) $\Box b$ is valid if b is valid; otherwise $\neg\Box b$ is valid.

3.4. The Independence of Elementary Sentences. The independence of states of affairs is reflected in the independence of elementary sentences. Let us call a set of sentences "independent" (*unabhängig*) if the situations described by these sentences are independent, and let a "state-description" be a complete set of elementary sentences, i.e., a set which contains for every elementary sentence either this sentence or its negation, but not both, and no other elements (Carnap 1956, p. 9). By definition of σ the independence of $\boldsymbol{EL}$ may be given the following expression, which is immediately provable by Condition 1 on $\mathcal{S}$ (Section 1.3):

Theorem 1: For every state-description $\boldsymbol{SD}$, $\Diamond \bigwedge \boldsymbol{SD}$ is valid.

Thus, all members of any state-description are compossible, just as all members of any complete set of states of affairs are compossible.

With Theorem 1, all Tractarian assertions on the independence of elementary sentences may be proved. For example, TLP 4.211 says that elementary sentences do not contradict each other (cf. TLP 6.3751). Indeed, suppose that $\sigma(p) = -\sigma(q)$ for some $p, q \in \boldsymbol{EL}$. Then $\sigma(p) \sqcup \sigma(q) = \mathbf{1}$, which contradicts Theorem 1. Similarly, TLP 5.134 says that elementary sentences cannot be deduced from each other (cf. TLP 2.062, already discussed in Section 1.3). Indeed, suppose p follows from q ($p, q \in \boldsymbol{EL}$). Then $\sigma(p) \sqsubseteq \sigma(q)$, for as TLP 5.122 states, "If p follows from q, the sense of 'p' is contained in the sense of 'q'" (cf. the definition of "is contained in" in Section 1.2). It follows that $\sigma(p) \sqcup -\sigma(q) = \mathbf{1}$, which contradicts Theorem 1.

3.5. The Complete Describability of the World by Elementary Sentences. TLP 4.26 asserts:

> If all true elementary sentences are given, the result is a complete description of the world. The world is completely described by giving all elementary sentences, and adding which of them are true and which false.

In our formal reconstruction, this follows immediately from Condition 2 on $\mathcal{S}$ (Section 1.3) and Definitions 15 and 16 of σ and TV:

Theorem 2: If $\mathbf{w} \neq \mathbf{w}'$ then there is at least one $p \in \boldsymbol{EL}$ such that $TV(p, \mathbf{w}) = T$ while $TV(p, \mathbf{w}') = F$.

Because two worlds cannot contain precisely the same states of affairs, they cannot agree on all elementary sentences (cf. the discussion of TLP 2.04 in Section 1.3); if two worlds differ, there is at least one elementary sentence describing the difference.

Theorems 1 and 2 are in conjunction equivalent to:

Theorem 3: $\sigma(\bigwedge \boldsymbol{SD}) \in \mathbf{W}$ for every state-description $\boldsymbol{SD}$,

which may also be directly derived from Condition 3 on $\mathcal{S}$ (Section 1.3). There is a 1-1 correspondence between worlds, the state-descriptions describing them, and the states of affairs existing in them: each state-description describes precisely one world, and each world is completely described by one state-description. (Therefore state-descriptions are in fact *world*-descriptions.) Non-elementary sentences such as quantified sentences are superfluous as far as the description of the world is concerned. This also applies to modal sentences: because their truth-value is the same in all possible worlds, they do not contribute to the description of any one world in particular.

Theorem 3 has an interesting consequence: for any interpretation and any $p \in \boldsymbol{EL}$, $\bigwedge \boldsymbol{SD} \supset p$ is valid iff $p \in \boldsymbol{SD}$.[14] Notice that this is a definition of validity for elementary sentences *which does not depend on the interpretation*. However, the set of all valid sentences $p \in \boldsymbol{L}$ is recursively definable in terms of the set of all valid elementary sentences, for we have:

Theorem 4: For any interpretation and any state-description $\boldsymbol{SD}$:

(a) For any $p \in \boldsymbol{EL}$: $\bigwedge \boldsymbol{SD} \supset p$ is valid iff $p \in \boldsymbol{SD}$.
(b) For any $p \in \boldsymbol{L}$: $\bigwedge \boldsymbol{SD} \supset \neg p$ is valid iff $\bigwedge \boldsymbol{SD} \supset$ p is not.
(c) For any $P \subseteq \boldsymbol{L}$: $\bigwedge \boldsymbol{SD} \supset \bigwedge P$ is valid iff $\bigwedge \boldsymbol{SD} \supset p$ is valid for all $p \in P$.
(d) For any $p \in \boldsymbol{L}$: p is valid iff $\bigwedge \boldsymbol{SD} \supset p$ is valid for all $\boldsymbol{SD}$.
(e) $\Box p$ is valid if p is valid; otherwise $\neg\Box p$ is valid.[15]

Therefore the above provides a recursive definition of validity for all sentences $p \in \boldsymbol{L}$ which is *independent from interpretations*. It follows that:

Theorem 5: Exactly the same sentences $p \in \boldsymbol{L}$ are valid in all interpretations.

3.6. The Principle of Truth-functionality. The Tractarian principle of truth-functionality is weaker than might be expected. According to present-day definitions, a sentence cannot be truth-functional (i.e., a sentence cannot contain only truth-functional connectives) unless its truth-value is some function of the

[14] Proof. "⇒": suppose $\sigma(\bigwedge \boldsymbol{SD} \supset p) = \mathbf{0}$. Then $\sigma(p) \sqsubseteq \sigma(\bigwedge \boldsymbol{SD})$. As $\sigma(\bigwedge \boldsymbol{SD}) \in \mathbf{W}$ by Theorem 3, therefore $-\sigma(p) \not\sqsubseteq \sigma(\bigwedge \boldsymbol{SD})$; accordingly $\neg p \notin \boldsymbol{SD}$, whence $p \in \boldsymbol{SD}$, Q.E.D. "⇐": if $p \in \boldsymbol{SD}$, then obviously $\sigma(p) \sqsubseteq \bigsqcup\{\sigma(p) : p \in \boldsymbol{SD}\}$, whence $\sigma(\bigwedge \boldsymbol{SD} \supset p) = \mathbf{0}$, Q.E.D.

[15] Proof of (a) is easy. (b)–(e) are most easily proved by realizing that $\bigwedge \boldsymbol{SD} \supset p$ is valid iff p is true at $\sigma(\bigwedge \boldsymbol{SD})$ (for p is true at $\sigma(\bigwedge \boldsymbol{SD})$ iff $\sigma(p) \sqsubseteq \sigma(\bigwedge \boldsymbol{SD})$ iff $\sigma(\bigwedge \boldsymbol{SD} \supset p) = \mathbf{0}$), and then using observations (b)–(e) on truth from Section 3.3.

truth-values of the *subsentences* it contains and of the way it is built up from these; the truth-values of other sentences do not count (see, e.g., Humberstone (1986)).[16]

The Tractarian formulation of truth-functionality seems more liberal: "The sentence is a truth-function of the elementary sentences. (The elementary sentence is a truth-function of itself.) The elementary sentences are the truth-arguments of sentences" (TLP 5-5.01). It does not seem to be required here that only the truth-values of the elementary *subsentences* of a sentence matter as to its truth-value; the latter truth-value may as well be a function of the truth-values of *all* elementary sentences.

We shall take the Tractarian principle of truth-functionality to mean the latter. Thus, this principle asserts that the truth-values of the elementary sentences jointly determine the truth-values of all sentences; given a state-description, any sentence may assume only *one* truth-value. Or to put it formally: the principle asserts that for all $p \in \boldsymbol{L}$, the relation $\{\langle \boldsymbol{SD}, TV(p, \mathbf{s})\rangle : \boldsymbol{SD}$ is a state-description such that $TV(\bigwedge \boldsymbol{SD}, \mathbf{s}) = T\}$ is a *function*; that is, there are no situations $\mathbf{s}$, $\mathbf{s}'$ verifying the same state descriptions while simultaneously $TV(p, \mathbf{s}) \neq TV(p, \mathbf{s}')$ for some $p \in \boldsymbol{L}$. Intuitively, the principle of truth-functionality is presupposed by the principle of the complete describability of the world by elementary sentences: if the former principle did not hold, two worlds could verify the same elementary sentences and yet differ as regards the truth-value of some other sentence—and they would hence not be *completely* described by elementary sentences. This intuition is borne out by our explication: Theorem 2 (the completely describability thesis) implies the principle of truth-functionality (Theorem 6a). Summarizing the above and adding some refinements, we define:

Definition 17: Let $\mathscr{R}(\boldsymbol{EL}, p) = \{\langle \boldsymbol{SD}, TV(p, \mathbf{s})\rangle : \boldsymbol{SD}$ is a state-description such that $TV(\bigwedge \boldsymbol{SD}, \mathbf{s}) = T\}$.

(a) p is a *truth-function* of $\boldsymbol{EL}$ (in an interpretation I) iff $\mathscr{R}(\boldsymbol{EL}, p)$ is a function (in I).

(b) The *principle of truth-functionality* holds for $\boldsymbol{L}$ iff, for all $p \in \boldsymbol{L}$ and all I, $\mathscr{R}(\boldsymbol{EL}, p)$ is a function.

(c) p is a *determinate* truth-function of $\boldsymbol{EL}$ iff $\mathscr{R}(\boldsymbol{EL}, p)$ is the *same* function in all I.

(d) p is an *indeterminate* truth-function of $\boldsymbol{EL}$ iff $\mathscr{R}(\boldsymbol{EL}, p)$ is a function which *varies* with I.

For the moment, we do not need (d); we shall encounter indeterminate truth-functions in Section 4.6.

Theorem 6: (a) The principle of truth-functionality holds for $\boldsymbol{L}$.

(b) All sentences $p \in \boldsymbol{L}$ are determinate truth-functions of $\boldsymbol{EL}$ (under Tractarian interpretations of $\boldsymbol{L}$).

(c) $\{\langle TV \restriction (\boldsymbol{EL} \times \{\mathbf{w}\}), TV \restriction (\boldsymbol{L} \times \{\mathbf{w}\})\rangle : \mathbf{w} \in \mathbf{W}\}$ is a function.

(d) $\Box(\bigwedge \boldsymbol{SD} \supset p) \vee \Box(\bigwedge \boldsymbol{SD} \supset \neg p)$ is valid for all $\boldsymbol{SD}$ and all $p \in \boldsymbol{L}$.

(e) For all $\boldsymbol{SD}$ and all $p \in \boldsymbol{L}$: either $\bigwedge \boldsymbol{SD}$ and p, or $\bigwedge \boldsymbol{SD}$ and $\neg p$, are not compossible.

(f) For any p, $\boldsymbol{EL} \cup \{p\}$ is not independent.[17]

---

[16] The set of subsentences of a sentence is defined as $Sub(p) = \{p\}$ if $p \in \boldsymbol{EL}$, $Sub(NP) = \{NP\} \cup \bigcup\{Sub(p) : p \in P\}$ and $Sub(\Box p) = \{\Box p\} \cup Sub(p)$.

[17] Proof of (a): by Theorem 2, if $TV(\bigwedge \boldsymbol{SD}, \mathbf{s}) = TV(\bigwedge \boldsymbol{SD}, \mathbf{s}') = T$, then $\mathbf{s} = \mathbf{s}'$ (and $\mathbf{s}, \mathbf{s}' \in \mathbf{W}$, or $\mathbf{s} = \mathbf{s}' = \mathbf{1}$); hence $TV(p, \mathbf{s}) = TV(p, \mathbf{s}')$ for any p, Q.E.D. Proof of (b): it must be shown that $\mathscr{R}(\boldsymbol{EL}, p)$ is the same function in all interpretations I, that is, that $TV(\bigwedge \boldsymbol{SD}, \mathbf{s})$ determines $TV(p, \mathbf{s})$ regardless of I. Proof: by Theorem 4, $TV(p, \mathbf{s}) = T$ just in case $\bigwedge \boldsymbol{SD} \supset p$ is valid; as validity does not depend on I by Theorem 5, $TV(\bigwedge \boldsymbol{SD}, \mathbf{s})$ determines $TV(p, \mathbf{s})$ regardless of I, Q.E.D. (c)–(f) are obvious consequences of (a).

Sentences beginning with a modal operator do not form an exception to the principle of truth-functionality: their truth-values are *constant* functions of the truth-values of the elementary sentences. That is to say, $\mathscr{R}(\boldsymbol{EL}, \Box p)$ is a *constant* function: for all $\mathbf{s} \in \mathbf{S}$, $TV(\Box p, \mathbf{s})$ is the same. This also holds for $\mathscr{R}(\boldsymbol{EL}, \Diamond p)$.

As all logicians know (but most commentators of the *Tractatus* do not), the principle of truth-functionality must be distinguished from the principle of *extensionality*, which says that $(p \equiv p') \supset (q \equiv q')$ is valid in all interpretations, where q' is like q except that some occurrence of subsentence p of q has been replaced by p'. Because of the presence of modal sentences our Tractarian language is, though truth-functional, definitely *not* extensional (see Humberstone (1986) for extensionality without truth-functionality).

4. Propositional Attitude Ascriptions

4.1. Syntax of Propositional Attitude Ascriptions. Wittgenstein begins his discussion of the propositional attitude ascriptions (from now on "thought-ascriptions" for short) by stating that "In the general sentential form a sentence occurs within a sentence only as a basis of truth-operations" (TLP 5.54). This is generally seen as an affirmation of the thesis of extensionality (Section 3.6), for instance by Black (1964).

However, this interpretation seems dubious. In TLP 5.542 Wittgenstein goes on to declare that thought-ascriptions are (appearances notwithstanding: TLP 5.541) no exception to the principle of TLP 5.54, because "A thinks that φ" has the same form as "'φ' says φ".[18] However, not only is it clear that "'φ' says φ" is *not extensional at all* (so this analysis would be pointless if TLP 5.54 really expressed the principle of extensionality): what is more, any analysis according to which "φ" occurs extensionally in "A thinks that φ" would simply be ludicrous, for thought-ascriptions are plainly not extensional.

Must we then conclude that "A thinks that φ" does not contain an occurrence of a subordinate sentence "φ" at all (as has, for example, been done by Black (1964) and Fogelin (1976))? This would not sound very convincing either. Obviously "φ" occurs in *some* sense in the latter sentence (albeit not in an extensional one), just as it occurs in some sense in "'φ' says φ". Both statements would be impenetrable, structureless wholes otherwise, which they not only do not *seem* to be, but also conflicts with the Tractarian view that no sentence is an unstructured whole: "The sentence is articulated" (TLP 3.251). Therefore we propose an alternative, more lenient interpretation of TLP 5.54.

In our view TLP 5.54 says nothing more than that sentences always occur *as sentences* within other sentences, and never as anything else. So sentences never occur in sentences as *names* of sentences, as *names* of facts, as *collections* of sentential *constituents* (names), as *affixes* (TLP 5.02), as *facts*, or whatever: subsentences of sentences have the same status as sentences standing on their own. That is to

[18] We shall use the Greek character "φ" as an *abbreviation for a sentence of ordinary language* and rewrite all quotations from Wittgenstein accordingly. This is necessary in order to prevent confusion with the formal development, where we use "p" as the *name* of the sentence $p \in \boldsymbol{L}$, as is usual in formal logic. Thus, it is correct to say that the name of a sentence $p \in \boldsymbol{L}$ is "p", but it would be absurd to say that the name of φ is "φ". Instead, we must say that the name of "φ" is "'φ'". For example, it is absurd to say that "the sun is shining" ("φ") is the name of (the situation) the sun is shining (φ), but it is correct to say that "'the sun is shining'" ("'φ'") is the name of "the sun is shining". To put it crudely, "p" corresponds to "'φ'", p corresponds to "φ", and $\sigma(p)$ corresponds to φ. Confusion between the names of sentences, sentences, and the senses of sentences would be fatal in the contexts considered here.

The same distinctions apply to the ordinary-language name of the subject, "A", and the formal counterpart of this name, A (although confusion is less serious here). Therefore we use the Greek character "A" in the former case and the Roman character "A" in the latter case.

say, semantically they are *descriptions of situations*, and syntactically they are possible *arguments of connectives* (bases of truth-operations); this is already clear in the case of, e.g., negations, but TLP 5.54 emphatically repeats this principle for all sentences.

Now what should have been demonstrated after TLP 5.541 is that thought-ascriptions are only an apparent violation of the principle just mentioned. However, such a demonstration can only be found in the *Notebooks*. There it is clearly stated that in "A thinks that φ", "φ" plays the same syntactical role as it does in "not φ" and "it is necessary that φ": in "A thinks that φ", "'φ' cannot be replaced by a proper name" (NB p. 95; cf. NB p. 106), nor will it do "to mention only its [i.e., "φ" 's] constituents, or its constituents and form but not in the proper order" (NB p. 94). Instead, "the sentence itself must occur in the statement to the effect that it is thought" (*Ibid.*). The underlying reason is that in "A thinks that φ", "φ" plays the same *semantical* role as in "not φ" and "it is necessary that φ": it is a *description of a situation*, just as in these other cases. And a situation cannot be described by a name or a *Klasse von Namen* (TLP 3.144, 3.142).[19]

As we have said, such a demonstration of the compatibility of thought-ascriptions with TLP 5.54 is, however, not to be found in the *Tractatus*. There it is only remarked that "A thinks that φ" is comparable to "'φ' says φ" (TLP 5.542). This comparison seems most unfortunate: in the latter sentence "φ" *does* occur as a name ("'φ'" is a name of "φ"), so this sentence violates the principle of TLP 5.54 (as we have interpreted it) and can hardly serve to explain why "A thinks that φ" does *not* violate it. Moreover, "'φ' says φ" is a metalinguistic statement and as such *unsinnig* (Section 3.2): but it is utterly implausible to suppose that "A thinks that φ" is *unsinnig*. The latter locution seems to be a perfectly proper part of everyday language. But then what is the function of the comparison in TLP 5.542? In our view it only serves to clarify the semantical analysis of thought-ascriptions (as we shall see in the next section). Thus, TLP 5.542 ff. do not bear on TLP 5.54-5.541 at all, stylistic appearances to the contrary notwithstanding. The *Tractatus* contains a gap between 5.541 and 5.542 which must be filled up with remarks from the *Notebooks*.

Summing up our discussion thus far: "A thinks that φ" is a sentence in which another sentence, "φ", occurs as the "basis" (the argument) of a "truth-operator" (connective). The "truth-operator" in question is "A thinks that ... ", which is comparable to "not ... ", and "it is necessary that ... ". Parallelling this analysis on the syntactic side we introduce the unary connective D_A, which is syntactically analogous to $\neg$ and $\Box$, and which may be read as "A thinks that ... " (*A denkt, daß ...* ; however, any propositional attitude may be substituted here):

Definition 18: $\boldsymbol{L}_D$ is the smallest set such that $\boldsymbol{L} \subseteq \boldsymbol{L}_D$ and if $p \in \boldsymbol{L}_D$, then $D_A p \in \boldsymbol{L}_D$; $\boldsymbol{T}_D$ is the smallest set such that $\boldsymbol{T} \subseteq \boldsymbol{T}_D$ and if $p \in \boldsymbol{L}_D$, then $D_A p \in \boldsymbol{T}_D$; $\boldsymbol{B}_D = \boldsymbol{L}_D \cup \boldsymbol{T}_D$.[20]

[19] This is Wittgenstein's objection to Russell's "theory of judgment", in which the propositional attitude ascription "A judges that a and b are similar", is analysed as "$J\{A, a, b, \text{similarity}, xFy\}$", where "$xFy$" stands for the form "something and something have some relation" (Russell 1984, p. 117). Here we *do* have a "class of names" (with one illegitimate name at that: according to Wittgenstein only objects can be named, forms cannot; cf. NB p. 105). As a result, the contact with situations is lost. Any class containing the appropriate constituents may be taken as the argument of a judgment-ascription, without any regard for situations. For example, $\{\text{this table}, \text{the book}, \text{penholders}, xFy\}$ would qualify (at least according to Wittgenstein); hence "A judges that this table penholders the book" would be a well-formed judgment-ascription on Russell's account (NB p. 103). Wittgenstein regards this as absurd.

[20] The operator "D_A" stands for the "most general" propositional attitude. If we start with several propositional attitudes D_A^n, $n \leq \omega$, then $D_A p$ may be defined as $D_A p = \bigvee\{D_A^i p : 0 \leq i \leq n\}$ (cf. note 30).

4.2. Informal Semantics of Propositional Attitude Ascriptions. Something about the semantics of thought-ascriptions may already be gleaned from the above: in "A thinks that φ", "φ" is a *description of a situation*. Like any sentence, it is a description of the situation which is its *sense*; as the *Notes on Logic* state, "here a *sense*, not a meaning [*Bedeutung*] is concerned" (NB p. 106). This immediately explains why "φ" cannot be "a piece of nonsense" here (as any adequate theory of thought-ascriptions must do: TLP 5.5422), for situations cannot be described by "pieces of nonsense".[21]

So the semantical role of the subordinate sentence of a thought-ascription is already clear. What about the subject of the ascription (referred to by "A") and his or her relation to the sense of "φ", which is "obviously not a relation in the ordinary sense" (NB p. 95)? It is here that TLP 5.542–5.5421 come in.

> It is clear [...] that "A believes that φ", "A thinks φ", "A says φ" are of the form "'φ' says φ": and this does not involve a correlation of a fact with an object, but rather the correlation of facts by means of the correlation of their objects. This shows too that there is no such thing as the soul—the subject, etc.,—as it is conceived in the superficial psychology of the present day. For a composite soul would no longer be a soul.

In order to understand this passage of "almost impenetrable obscurity" (Urmson 1956, p. 133), we first have to understand the statement "'φ' says φ". This is not too difficult. Wittgenstein generally uses the name of a sentence (i.e., the sentence within quotation-marks) to refer to the sentence (cf. TLP 5.12, 5.123, 5.1241, 5.1311, 5.152, 5.44, 5.512, 5.513), and the sentence itself to refer to the situation described by the sentence (e.g., in TLP 5.43). Therefore "'φ' says φ" is a specification of the situation which is described by "φ", that is, it is a specification of the sense of "φ". But this specification is rather uninformative, for in order to describe the sense of "φ", "'φ' says φ" uses "φ" itself. Therefore "'φ' says φ" means nothing more than that the sense of "φ" is the sense of "φ": the statement is a correct but not very informative specimen of sense-specification. (An example: let "φ" = "The sun is shining". Then we have: "'The sun is shining' says the sun is shining". Of course it does; but this does not tell us much about the situation the sun is shining.) The formal rendering of "'φ' says φ" will be clear, given our definition of "p says $\mathbf{s}$ is the case" as $\sigma(p) = \mathbf{s}$ in Section 3.2: this is simply $\sigma(p) = \sigma(p)$, which is, again, correct but rather uninformative.

Now according to TLP 5.542 "A thinks that φ" is analogous to "'φ' says φ". This might be taken to mean that the former sentence is an instance of sense-specification as well. However, this would make thought-ascriptions *unsinnig*, as sense-specifications are metalinguistic assertions not belonging to the language itself (this also holds for their formal counterparts: $\sigma(p) = \mathbf{s} \notin \boldsymbol{L}$). But as we have already said, it is implausible to suppose that thought-ascriptions are nonsensical.

The solution to this problem is to assume that TLP 5.542 gives a *semantical* analysis of thought-ascriptions. Not the thought-ascription itself but its *sense* or its *truth-condition* is in some way analogous to "'φ' says φ". That is to say, "A thinks that φ" describes a *situation* which is in some way similar to the latter sentence; it is true iff that situation is indeed the case. Formally: not $D_A p$, but $\sigma(D_A p)$ is analogous to $\sigma(p) = \sigma(p)$ ("p says that $\sigma(p)$ is the case"). $D_A p$ itself is analogous

[21] Syntactically, Definition 18 already prevents "nonsensical" thought-ascriptions from being well-formed sentences. Since $\boldsymbol{L}_D$ is the *smallest* set satisfying Definitions 8 and 18, it may be proved that $D_A p \in \boldsymbol{L}_D$ implies that $p \in \boldsymbol{L}_D$. Thus, it is *unmöglich einen Unsinn zu urteilen* (TLP 5.5422), for if the subordinate sentence p of $D_A p$ were *unsinnig* while $D_A p$ itself were not, then we would have a nonsensical $p \in \boldsymbol{L}_D$, which is impossible (since the nonsensicality of p implies that $p \notin \boldsymbol{L}_D$: Section 3.2).

to $\Box(p \equiv p)$, which *says that* $\sigma(p) = \sigma(p)$, that is, which describes a situation involving sense-specification. In the same way as $\sigma(\Box(p \equiv p))$ does, the situation $\sigma(D_A p)$ must somehow involve sense-specification.

If this is correct, "A", in "A believes that φ", must refer to at least one picture representing a situation (otherwise there would be nothing to specify the sense of). This picture cannot be an object: objects are incapable of representing because they are simple. Only situations have the "logical complexity" which is required to represent (complex) situations (cf. TLP 2.02, 2.021, 3.142, 3.144, 4.032–4.041). So "A" refers to at least one situation, viz. the representing picture; in TLP 5.542 this situation is called a "fact", to which another "fact", namely the fact (or situation) represented by the former fact, is correlated. This makes it clear that "A" cannot be a name, for names always refer to objects. "A" is a "pseudo-name".

This makes one wonder about the references of pseudo-names. What do they refer to, in order that they may refer to one or more pictures? The answer is provided by Russell's writings of the same period. According to Russell symbols like "A" ("incomplete symbols", as he calls them) do not refer to simple objects, but to certain "logical fictions", namely classes, or series, or series of classes (Russell 1918, p. 253). "Persons are fictions" as well (Russell 1922, p. xix); the person referred to by "A" is similarly a "series of events" or a "class of facts" (Russell 1927, p. 403, p. 405). "The names that we commonly use, like 'Socrates', are really abbreviations for descriptions, not only that, but what they describe are not particulars but complicated systems of classes or series" (Russell 1918, pp. 200–1). Note that incomplete symbols are *not* rigid designators (in contrast to genuine names): the classes of facts they refer to form time-dependent *series*. Therefore their references vary with time (or per situation).

Wittgenstein supplements Russell's view of persons, then, by stating that some of the situations ("events", "facts") constituting a person in a given situation may have a pictorial character. Ascribing the thought that φ to a person amounts to asserting that among the pictures in question there is at least one which represents or models the situation φ.

Summing up, the above leads to the following analysis of thought-ascriptions: "A thinks that φ" is true (in a situation) iff the class of facts referred to by "A" (in that situation) contains at least one picture that says that φ. The latter picture may be, of course, be called a "thought" in the sense of Section 2.2. Thus, "A thinks that φ" is true iff the subject referred to by "A" has a thought that says that φ. It is of course the latter part of the truth-condition (the "business part", as Anscombe (1959, p. 88) called it) that is similar to "'φ' says that φ"; this is where the sense-specification comes in. Or to put it differently, "A thinks that φ" says that A has some thought that says that φ: it is the *sense* of "A thinks that φ" that is similar to "'φ' says that φ". A thought-ascription partially describes a person by means of specifying the sense of one of his thoughts, where the latter is done by employing a subordinate sentence having the same sense as that thought. (Because the sense of the thought is specified by using *another* picture, the latter sense-specification is not as uninformative as that in "'φ' says φ", where the *same* sentence is used to indicate the sense.) "A thinks that φ" is not meta-linguistic itself; but our analysis clearly vindicates the assertion by Clark (1976, p. 81) that "In ascribing thoughts and perceptions we are, very nearly, saying meta-linguistic things".

In view of the above the Tractarian account of thought ascription rather surprisingly turns out to be practically literally identical to that of modern "language

of thought" theorists.[22] This is no shortcoming of our analysis: it only says something about the influentiality of the *Tractatus*![23]

Before proceeding with the formal semantics of thought-ascriptions, it should be remarked that incomplete symbols should not really be admitted in a logically perfect language (this was repeatedly emphasized by Russell). We have made an exception to this by allowing "A" in D_A, but we shall see that allowing this operator does not increase the capacity of the language to describe the world anyway—which is all the more justification to exclude incomplete symbols from an ideal logical language!

4.3. Formal Semantics of Propositional Attitude Ascriptions. The above insights may directly be incorporated into our formal semantics.

Definition 19: A Tractarian interpretation for a doxastic[24] pictorial system $\Pi = \langle \boldsymbol{E}, \boldsymbol{TE}, \boldsymbol{N}, \boldsymbol{EB}, \boldsymbol{ET}, \boldsymbol{EL}, \boldsymbol{B}_D, \boldsymbol{T}_D, \boldsymbol{L}_D \rangle$ as described in Section 2.4 and Definition 18, is a triple $I = \langle \Sigma, \delta, \psi_A \rangle$ such that Σ and δ are as in Definition 14 and $\psi_A : \mathbf{S} \mapsto \mathscr{P}(\mathbf{S})$ is a function such that $\{\sigma(t) : t \in \boldsymbol{T} \cap \psi_A(\mathbf{s})$ and $\Delta(t) = n\} = \bigcap\{\{\sigma(t) : t \in \boldsymbol{T} \cap \psi_A(\mathbf{w})$ and $\Delta(t) = n\} : \mathbf{w} \sqsupseteq \mathbf{s}\}$.

Here $\mathscr{P}(\mathbf{S})$ is the power-set of $\mathbf{S}$. The set of situations $\psi_A(\mathbf{s})$ is the "pseudo-denotation" of the "pseudo-name" A at $\mathbf{s}$.[25] As remarked above, A does not refer rigidly: ψ_A is not a constant function. $\Delta(b)$ is the doxastic degree of b, a notion which is similar to the usual notion of modal degree: $\Delta(b) = 0$ if $b \in \boldsymbol{EB}$, $\Delta(NP) = \max\{\Delta(b) : b \in P\}$, $\Delta(\Box b) = \Delta(b)$, and $\Delta(D_A p) = \Delta(p) + 1$. Notice that $\Delta(b) = 0$ iff $b \in \boldsymbol{B}$. The conditions on Δ and ψ_A will be motivated in a moment.

Definition 15 of σ is extended as follows:

Definition 20: $\sigma : \boldsymbol{B}_D \mapsto \mathbf{S}$ is a function such that the conditions of Definition 15 hold and moreover $\sigma(D_A p) = \sqcap\{\mathbf{s} \in \mathbf{S}$: there is a $t \in \boldsymbol{T} \cap \psi_A(\mathbf{s})$ such that $\sigma(t) = \sigma(p)$ and $\Delta(t) \leq \Delta(p)\}$.

As Definition 16 is kept unchanged, Definition 20 ensures that $TV(D_A p, \mathbf{s}) = T$ iff there is a $t \in \boldsymbol{T}$ such that $t \in \psi_A(\mathbf{s})$ and $\sigma(t) = \sigma(p)$ and $\Delta(t) \leq \Delta(p)$. This is precisely the truth-clause we arrived at in our informal analysis in Section 4.2, except that we have extended our informal account with the condition that $\Delta(t) \leq \Delta(p)$. We have done so because we want the definition of the sense of $D_A p$ to be an *explanatory analysis* of $D_A p$ at the same time. For this to be the case, the definition must be a reductive one, that is, one in which the sense of $D_A p$ does not ultimately rest on the senses of other thought-ascriptions. (Theorem 8 below shows that our definition is a reductive one.) Otherwise, it could, e.g., be the case that $TV(D_A p, \mathbf{s}) = T$ iff there is a $t \in \boldsymbol{T} \cap \psi_A(\mathbf{s})$ such that $\sigma(t) = \sigma(p)$

[22] See, e.g., Field (1978) and Harman (1973). A detailed comparison of Harman's and Wittgenstein's language of thought theories has been carried out by Berghel (1978).

[23] The influence of TLP 5.542 may even be traced back in Wittgenstein's own strikingly similar remarks about "expecting" in the *Philosophische Bemerkungen* (ca. 1929).

> Ist es nicht so, daß meine Theorie ganz darin ausgedrückt ist, daß der Sachverhalt, der die Erwartung von p befriedigt, durch den Satz p dargestellt wird? (Wittgenstein 1964, remark no. 25).
>
> Die Erwartung, der Gedanke, der Wunsch, etc., daß p eintreffen wird, nenne ich erst dann so, wenn diese Vorgänge die Multiplizität haben, die sich in p ausdrückt, erst dann also, wenn sie *artikuliert* sind. (*Ibid.*, remark no. 32).

[24] From Greek "*dokein*" ("to think"), "*doxastikos*" ("pertaining to mere opinion, as opposed to knowledge").

[25] We might, but shall not, impose the following additional conditions on ψ_A. First, $\bigsqcup \psi_A(\mathbf{s}) \sqsubseteq \mathbf{s}$. Secondly, for every $\mathbf{s}' \in \psi_A(\mathbf{s})$ there is a $p \in \boldsymbol{L}_D$ such that $\mathbf{s}' = \sigma(p)$. Thirdly, $Card(\psi_A(\mathbf{s})) \leq \aleph_0$. In this case, A would be locally definable: for each $\mathbf{s} \in \mathbf{S}$, there would always be a true conjunctive sentence $p_{\mathbf{s}}$ such that $\sigma(p_{\mathbf{s}}) = \psi_A(\mathbf{s})$. A similar procedure may be hinted at in TLP 3.24.

and $\Delta(t) > \Delta(p)$. In this case it could be possible that the sense of $D_A p$ rests on that of $D_A D_A p$, with the sense of the latter in turn resting on that of $D_A D_A D_A p$, etc.: this would be one of the most mystifying analyses of $D_A p$ ever put forward!

The special condition on ψ_A in Definition 19 only serves to bring the doxastic interpretations into line with the non-doxastic semantics of Section 3. Using this condition, we may generalize the remarks on the senses and truth-values of sentences $p \in \boldsymbol{L}$ in Section III (up to Theorem 4) to *all* sentences $p \in \boldsymbol{L}_D$ and prove such theorems as $\sigma(D_A p) = \sqcap\{\mathbf{w} \in \mathbf{W} : TV(D_A p, \mathbf{w}) = T\}$ and $TV(D_A p, \mathbf{s}) = T$ iff $TV(D_A p, \mathbf{w}) = T$ for all $\mathbf{w} \sqsupseteq \mathbf{s}$. This would be impossible otherwise.

In order to gain a clear insight in the logical properties of thought ascriptions, we shall give a discussion exactly parallelling Section 3.4–3.6 above.

4.4. The Interdependence of Propositional Attitude Ascriptions. On this subject we may be very brief: as, e.g., $D_A p \equiv D_A \neg\neg p$ and $D_A(p \wedge q) \equiv D_A(q \wedge p)$ are valid, thought-ascriptions are clearly interdependent (i.e., not independent). In this respect they are similar to modal sentences. Therefore, there are no doxastic elementary sentences (as is already clear from the fact that $D_A p$ is not a concatenation of names).

4.5. The Supervenience of Propositional Attitude Ascriptions on Elementary Sentences. Although there are no doxastic elementary sentences, this does not affect the capacity of elementary sentences to provide complete descriptions of all worlds. Condition 2 and Theorem 2 still hold; therefore thought-ascriptions are as redundant as modal sentences as far as the describability of worlds is concerned. They partially describe persons by means of specifying the senses of some of their thoughts, but persons are, just as the worlds they form part of, already completely described by elementary sentences.

In modern parlance, this is expressed by calling thought-descriptions *supervenient* on elementary sentences. As Haugeland (1982, p. 97) defines it:

> Two worlds in $\mathbf{W}$ are *discernible with* language $\boldsymbol{L}$ just in case there is a sentence of $\boldsymbol{L}$ which is true at one, and not at the other. [. . .] $\mathbf{K}$ *weakly supervenes on* $\boldsymbol{L}$ (relative to $\mathbf{W}$) just in case any two worlds in $\mathbf{W}$ discernible with $\mathbf{K}$ are discernible with $\boldsymbol{L}$.

Accordingly, $\boldsymbol{L}$, $\boldsymbol{L}_D$, $\boldsymbol{T}$, $\boldsymbol{T}_D$, $\boldsymbol{B}$ and $\boldsymbol{B}_D$ all weakly supervene on $\boldsymbol{EL}$.

However, not all of Section 3.5 applies to thought-ascriptions: Theorem 4 now no longer provides a definition of validity for *all* sentences. Sentences p such that $\Delta(p) \geq 1$ (i.e., sentences not belonging to $\boldsymbol{L}$) are not covered by it. Indeed, it is readily seen that the validity of thought-ascriptions cannot be recursively defined for all interpretations I because it may vary with I. For example, we may have $\psi_A(\mathbf{s}) = \emptyset$ (for all $\mathbf{s}$) in I and $\psi_A(\mathbf{s}) = \mathbf{S}$ (for all $\mathbf{s}$) in J, with the result that $\neg D_A p$ is valid in I but invalid in J and that $D_A p$ is valid in J but invalid in I (any p).

4.6. The (Indeterminate) Truth-functionality of Propositional Attitude Ascriptions. Because Theorem 2 still holds, clauses (a) and (c)–(f) of Theorem 6 hold under substitution of $\boldsymbol{L}_D$ for $\boldsymbol{L}$. Thus, the principle of truth-functionality holds: the truth-value of a thought-ascription is a function of the truth-values of the elementary sentences. However, clause (b) of Theorem 6 does not hold for $\boldsymbol{L}_D$: as validity may now vary from interpretation to interpretation, we have by Definition 17d:

Theorem 7: Thought-ascriptions are indeterminate truth-functions of $\boldsymbol{EL}$ (under Tractarian interpretations of $\boldsymbol{L}_D$).

Thus, $\boldsymbol{EL} \cup \{D_A p\}$ is not independent, although the specific form of the dependence varies from interpretation to interpretation; in any interpretation either

$\neg\Diamond(\bigwedge \boldsymbol{SD} \wedge D_A p)$ or $\neg\Diamond(\bigwedge \boldsymbol{SD} \wedge \neg D_A p)$ is valid, although it depends on the interpretation which one of both is valid. Of course, any p such that $\Delta(p) \geq 1$ is an indeterminate truth-function of $\boldsymbol{EL}$ as well.

Because $D_A p$ is an indeterminate truth-function of $\boldsymbol{EL}$, it is not sufficient to know the truth-values of the elementary sentences in order to know whether $D_A p$ is true. The case here is similar to the case of the description of the world by elementary sentences. Each state-description uniquely describes one world. But in order to know exactly *which* world it describes, one has to know some semantical facts: in particular, one has to know $\delta \restriction \boldsymbol{N}$—and this is all one has to know, for $\sigma(\bigwedge \boldsymbol{SD})$ is fully determined by $\delta \restriction \boldsymbol{N}$. The same is true for $\mathscr{R}(\boldsymbol{EL}, D_A p)$, the function telling how the truth-value of $D_A p$ depends on the truth-values of the elementary sentences (Definition 17): $\mathscr{R}(\boldsymbol{EL}, D_A p)$ is *some* function, but in order to know *which* function it is, one has to know some semantical facts. In particular, one has to know $\delta \restriction (\boldsymbol{N} \cup \boldsymbol{TE}_A)$ and ψ_A (knowing $\delta \restriction \boldsymbol{N}$ does not suffice), where $\boldsymbol{TE}_A$ is the set of thought-elements occurring in the sentences of "A's language of thought" $\boldsymbol{T}_A = \boldsymbol{T} \cap \bigcup\{\psi_A(\mathbf{s}) : \mathbf{s} \in \mathbf{S}\}$. $\delta \restriction (\boldsymbol{N} \cup \boldsymbol{TE}_A)$ and ψ_A are *all* one has to know in order to determine $\mathscr{R}(\boldsymbol{EL}, D_A p)$, for:

Theorem 8: $\sigma \restriction (\boldsymbol{L}_D \cup \boldsymbol{T}_A)$, and hence $\mathscr{R}(\boldsymbol{EL}, p)$, are fully determined by $\delta \restriction (\boldsymbol{N} \cup \boldsymbol{TE}_A)$ and ψ_A.[26]

Will it ever be possible to know the meanings of all elements of $\boldsymbol{N} \cup \boldsymbol{TE}_A$, and of "$A$"? Assuming that it is unlikely that we may ever know the meanings of more than a finite number of pictorial elements, this depends on the cardinality of $\boldsymbol{N} \cup \boldsymbol{TE}_A$; only the eventual finiteness of $\boldsymbol{N} \cup \boldsymbol{TE}_A$ (and hence of $\mathbf{G}$) would guarantee a humanly possible determinability of $\sigma(D_A p)$ on the basis of denotations.[27]

Because thought-ascriptions are *indeterminate* truth-functions of the elementary sentences, it may be objected that the truth-functional account we have offered is not really a very *illuminating* one. We do not deny this; but the extreme generality of the analysis may well be unavoidable. The assertion that the truth-values of thought-ascriptions cannot vary unless some other features (e.g., physical features ultimately describable by elementary sentences) of the world do seems hazardous enough as it is. When doing logic (philosophy), we can hardly venture beyond this; any other, more specific systematic relationships there may be between elementary sentences and thought-ascriptions may well be of an empirical (or at least partly empirical) nature and should therefore be settled by empirical science. As a consequence, the extreme generality of the truth-functional account (which parallels

[26] Proof: Let $\Delta(b) = n$ (where $b \in \boldsymbol{L}_D \cup \boldsymbol{T}_A$). By Definition 20 and the definition of $\Delta(b)$, $\sigma(b)$ only depends on ψ_A and $\sigma \restriction \{b \in \boldsymbol{L}_D \cup \boldsymbol{T}_A : \Delta(b) = n-1\}$. Repeating this argument as many times as necessary shows that $\sigma(b)$ depends only on ψ_A and $\sigma \restriction \{b \in \boldsymbol{L} \cup \boldsymbol{T}_A : \Delta(b) = 0\}$. Since the latter function only depends on $\delta \restriction (\boldsymbol{N} \cup \boldsymbol{TE}_A)$, $\sigma(b)$ is determined by $\delta \restriction (\boldsymbol{N} \cup \boldsymbol{TE}_A)$ and ψ_A. As $\mathscr{R}(\boldsymbol{EL}, p)$ is known as soon as $\sigma(\bigwedge \boldsymbol{SD})$ and $\sigma(p)$ are given for all $\boldsymbol{SD}$ and p, and $\sigma(\bigwedge \boldsymbol{SD})$ is determined by $\delta \restriction \boldsymbol{N}$, $\mathscr{R}(\boldsymbol{EL}, p)$ is fully determined by $\delta \restriction (\boldsymbol{N} \cup \boldsymbol{TE}_A)$ and ψ_A, Q.E.D.

[27] Even if $\boldsymbol{N} \cup \boldsymbol{TE}_A$ were finite, this would not guarantee the definability (reducibility) of thought-ascriptions in terms of elementary sentences. The latter would obtain in the two following two cases; however, both cases are ruled out by the *Tractatus*. First, if $\mathbf{SA}$ were finite, then $\mathcal{S}$ would be finitely generated by $\mathbf{SA}$. In this case there would be a $q \in \boldsymbol{L}$ such that $\Delta(q) = 0$ and $\sigma(q) = \sigma(D_A p)$, and $D_A p$ would accordingly be reducible to elementary sentences. However, the *Tractatus* assumes $\mathbf{SA}$ to be infinite (Section 1.1). Secondly, if we allowed uncountable disjunctions, with $\sigma(\bigvee P) = \prod\{\sigma(p) : p \in P\}$, then $\sigma(D_A p) = \prod\{\mathbf{w} \in \mathbf{W} : \mathbf{w} \sqsupseteq \sigma(D_A p)\} = \prod\{\sigma(\bigwedge \boldsymbol{SD}) : \sigma(\bigwedge \boldsymbol{SD}) \sqsupseteq \sigma(D_A p)\} = \sigma(\bigvee\{\bigwedge \boldsymbol{SD} : \bigwedge \boldsymbol{SD} \supset D_A p\})$, and we would, again, have explicit definability (of a totally uninformative sort). However, such disjunctions are even less Tractarian than countable conjunctions. (Notice that they would make definite descriptions definable:

$$q(\imath x)px = \bigvee\{q[a/(\imath x)px] \wedge p[a/x] \wedge \bigwedge\{\neg p[a'/x] : a' \in \boldsymbol{N} \setminus \{a\}\} : a \in \boldsymbol{N}\},$$

where $q(\imath x)px$ is introduced in the same way as $(x)px$ in clause (d) of Definition 8.)

the extreme generality of modern psychophysical supervenience theories) is not a defect, but a point in favour of the Tractarian theory.

4.7. Tractatus 5.542 Formalized. In order to demonstrate the adequacy of our formalization, let us show in detail how it ties in with TLP 5.542 and related passages. First, let p be an elementary sentence, $p = a_0 * \ldots * a_n$, $n \in \mathbb{N}$, with $\sigma(p) = \mathbf{g}_0 * \ldots * \mathbf{g}_n$. Then we have $TV(D_A p, \mathbf{s}) = T$ iff there is a $t \in \boldsymbol{T} \cap \psi_A(\mathbf{s})$ such that $\Delta(t) \leq \Delta(p)$ and $\sigma(t) = \mathbf{g}_0 * \ldots * \mathbf{g}_n$. One t that would qualify is an elementary thought $t = e_0 * \ldots * e_n$ such that $\delta(e_i) = \delta(a_i) = \mathbf{g}_i$ for all i, $0 \leq i \leq n$. It is clear that in this case we have a *darstellende Beziehung* between two situations, namely t and $\sigma(p)$, by means of *Zuordnungen* of their elements, for $\sigma(t) = \sigma(p)$ because $\delta(e_i) = \delta(a_i)$ for all i, $0 \leq i \leq n$ (cf. TLP 2.1514). Or to speak very crudely, we have here a correlation of two "facts" (in the sense of "concatenations of elements") by means of a correlation of their "objects" (in the sense of "elements": pictorial elements in the one case, objects *sensu stricto* in the other), which is precisely what TLP 5.542 says. A's thought that p is true if $\mathbf{g}_0 * \ldots * \mathbf{g}_n$ is a fact, and it is false otherwise.[28]

Now let us define A's "soul" (*Seele*) at $\mathbf{s}$ as $\boldsymbol{T}_A(s) = \boldsymbol{T} \cap \psi_A(\mathbf{s})$. Thus A's soul (at $\mathbf{s}$) consists of A's thoughts (at $\mathbf{s}$); it is the currently entertained subset ("theory") of A's language of thought $\boldsymbol{T}_A$. And let us define "logical multiplicity" (*logische Mannigfaltigkeit*, TLP 4.04–4.0412, 5.475) as follows: $\mathit{Mult}(e_0 * \ldots * e_n) = n + 1$, $\mathit{Mult}(NP) = \mathit{Mult}(P) = \max\{\mathit{Mult}(b) : b \in P\}$, $\mathit{Mult}(\Box b) = \mathit{Mult}(D_A b) = \mathit{Mult}(b)$. It will be clear that the logical multiplicity of any non-empty set of pictures is at least 2, even if there existed only one pictorial element (and object). Therefore as soon as A thinks anything at all at $\mathbf{s}$ (i.e., $D_A p$ is true at $\mathbf{s}$ for some p) $\mathit{Mult}(\boldsymbol{T}_A(\mathbf{s})) > 1$: "It is just as impossible that [the subject] should be a simple as that 'φ' should be" (NB p. 119; cf. TLP 5.5421). This makes it clear that Wittgenstein's contention that the soul is complex should definitely not be taken to mean that $\mathit{Card}(\boldsymbol{T}_A(\mathbf{s})) > 1$, as Hintikka (1958, p. 90) considered admissible. Nor should it be taken to mean that thoughts are not "combined" with each other in the soul, taking "combination" in the sense of "conjunction" (i.e., if $T \subseteq \boldsymbol{T}_A(\mathbf{s})$, then $\bigwedge T \in \boldsymbol{T}_A(\mathbf{s})$): it is possible that the soul is "unified" in the sense that it is closed under conjunction, and it is also possible that it is not. In the terminology of psychologists of the time, Wittgenstein asserts that the soul is not *einfach* (simple), while he does not commit himself on the question as to whether it is *einheitlich* (unified).[29]

Because D_A stands for all propositional attitudes, the remarks on perception in TLP 5.5423 can also be easily understood. If $\sigma(a_0 * \ldots * a_n) = \mathbf{g}_0 * \ldots * \mathbf{g}_n$, then to perceive that $a_0 * \ldots * a_n$ is not just to have some isolated psychical elements referring to $\mathbf{g}_0, \ldots, \mathbf{g}_n$ separately: instead, it is to have a thought saying that $\mathbf{g}_0, \ldots, \mathbf{g}_n$ "are related to one another in such and such a way" (TLP 5.5423). For example, it is to have a thought saying that $\mathbf{g}_0, \ldots, \mathbf{g}_n$ are concatenated, in this sequence, or (if $\mathbf{g}_0 = \mathbf{R}$) it is to have a thought representing the $\mathbf{R}$-configuration of $\mathbf{g}_1, \ldots, \mathbf{g}_n$ (to recall Suszko's terminology from Section 1.1). This explains why seeing that $a_0 * \ldots * a_n$ is different from, say, seeing that $a_n * \ldots * a_0$: "for we really see two different facts" in the two cases (TLP 5.5423).[30]

---

[28] Copi has reached essentially the same insights in his fine informal article on TLP 5.542 (Copi 1958).

[29] The soul we are talking about here is the "human soul, with which psychology deals" (TLP 5.641). This empirical soul must be carefully distinguished from the "metaphysical subject", which is *not* a part of the world (5.641), and *simple*, not complex (5.64).

[30] Two *prima facie* different semantical analyses may be given of the propositional attitudes $D_A^n p$, $n \leq \omega$, other than D_A (see note 20). First, we may distinguish between various subsets ("dialects") $\boldsymbol{T}_A^n$, $n \leq \omega$, of the language of thought $\boldsymbol{T}_A = \bigcup\{\boldsymbol{T}_A^i : 0 \leq i \leq n\}$: in this

The above may suffice as a demonstration of the adequacy of our formalization. It will be seen that few if any mysteries remain. The only unsolved problem is a historical one: which psychologists did Wittgenstein accuse of superficiality in TLP 5.5421? A search of the literature reveals that all major psychologists of the period regarded the soul as *einheitlich* but definitely not *einfach*![31]

5. Tractarian Doxastic Modal Logic

Wittgenstein himself was hardly interested in axiomatization (TLP 5.132), so we shall not go too deeply into this subject either. However, axiomatizing gives us a clear picture of what the preceding results lead up to; therefore we here present the doxastic modal logic *DML* (for a given Tractarian language $\boldsymbol{L}_D$) corresponding to the Tractarian semantics (for $\boldsymbol{L}_D$) proposed above.

5.1. Axiomatization of *DML*.

Axiom 1: Every axiom of finitary propositional logic is an axiom.
Axiom 2: $\bigwedge P \supset p$, where $p \in P$.
Axiom 3: $\Box p \supset p$.
Axiom 4: $\Diamond \bigwedge \boldsymbol{SD}$, for every state-description $\boldsymbol{SD}$.
Axiom 5: $\Box(\bigwedge \boldsymbol{SD} \supset p) \vee \Box(\bigwedge \boldsymbol{SD} \supset \neg p)$, for every state-description $\boldsymbol{SD}$.
Axiom 6: $\Box(p \equiv q) \supset (D_A p \equiv D_A q)$.
Rule 1: If $\vdash p$ and $\vdash p \supset q$, then $\vdash q$.
Rule 2: If $\vdash p \supset q$, for all $q \in Q$, then $\vdash p \supset \bigwedge Q$.
Rule 3: If $\vdash p \supset q$, then $\vdash p \supset \Box q$, provided p is fully modalized (i.e., provided every elementary sentence in p occurs within the scope of a modal operator).[32]

Here $\vdash p$ means that p is derivable in *DML*, i.e., that there exists a countable sequence $p_0, \ldots, p_j, \ldots, p_k$ such that $p_k = p$ and for each $j \leq k$, p_j is either an axiom or is inferred from earlier formulas p_i , $i < j$, by a rule of inference.

Theorem 9: $\vdash p$ iff p is valid in all Tractarian interpretations of $\boldsymbol{L}_D$.[33]

---

case, perceptions, judgments, memories, etc., are special kinds of thoughts. To perceive (judge, remember) that p is to have a perception (judgment, memory) that says that p, etc. Secondly, we may distinguish between various "compartments" or "faculties" $\boldsymbol{T}^n_A(\mathbf{s})$, $n \leq \omega$, of the soul $\boldsymbol{T}_A(\mathbf{s}) = \bigcup\{\boldsymbol{T}^i_A(\mathbf{s}) : 0 \leq i \leq n\}$, say the faculties of perception, judgment, memory, etc. In this case, to perceive that p is to have a thought saying that p in one's faculty of perception, etc. As mental faculties may be defined in terms of mental dialects and *vice versa* (for $\boldsymbol{T}^n_A(\mathbf{s}) = \boldsymbol{T}^n_A \cap \boldsymbol{T}_A(\mathbf{s})$ and $\boldsymbol{T}^n_A = \bigcup\{\boldsymbol{T}^n_A(\mathbf{s}) : \mathbf{s} \in \mathbf{S}\}$), both approaches are formally the same. Notice that $D_A p \equiv \bigvee\{D^i_A p : i \leq n\}$ is valid (cf. note 20).

[31] See, for example, the following quotations, from books which explain the concepts of *Einfachheit* and *Einheitlichkeit* at greater length than we have done:

> Unsere Untersuchung hat ergeben, daß, wo immer eine Seelentätigkeit besteht, eine gewisse Mannigfaltigkeit und Verwickelung vorhanden ist. Selbst in dem einfachsten Seelenzustande ist ein doppelter Gegenstand immanent gegenwärtig. [...] Aber der Mangel an Einfachheit war nicht ein Mangel an Einheit. (Brentano 1973, p. 221).
>
> *Einheit* [ist] der treffendere Ausdrück für die Natur der Seele [...] als *Einfachheit*. (Fechner 1860, p. 415).
>
> Woher schöpft man die Überzeugung, daß die Seele ein *einfaches* Wesen sei? [...] Wir [treffen] in dem Bewußtsein [...] eine Mannigfaltigkeit an, die auf eine Vielheit seiner Grundlage hinweist. [...] Nicht als einfaches Sein, sondern als geordnete Einheit vieler Elemente ist die Seele was Leibniz sie nannte: *ein Spiegel der Welt*. (Wundt 1874, pp. 862–3).

[32] See Hughes & Cresswell (1972, p. 127), where it is also shown that Axioms 1 and 3 and Rules 1 and 3 jointly constitute an axiomatization of $S5$.

[33] Proof. "⇒" (soundness): by calculation. "⇐" (completeness): suppose that not $\vdash p$. Then construct a "canonical model" as follows. $\mathbf{S}$ is the power-set of the set of maximally consistent

Theorem 9 holds *regardless of the order of DML* (which contrasts with the general situation for higher-order logic). The reason is clear: Tractarian interpretations correspond to Henkin's *general* models of higher-order logic (which enable completeness proofs), rather than to the so-called "natural" models (which do not). (Cf. Skyrms (1981, pp. 203–5).)

Because of the presence of $\bigwedge$, *DML* is undecidable. Without $\bigwedge$, *DML* would conceivably be decidable. (Whether it would actually be decidable depends on $Card(\boldsymbol{N})$ and thus on $Card(\mathbf{G})$; cf. Soames (1983, p. 588).)

5.2. Some Observations on *DML*. *DML* is an extension of the familiar logical systems $S5$ and $L\omega_1 0$ (classical propositional logic with countable conjunctions). The distinctive non-doxastic axioms of *DML* are Axioms 4 and 5. Axiom 4 is the linguistic counterpart of the thesis of the independence of states of affairs (Section 3.4). This axiom implies that $\vdash \Diamond \bigwedge P$ for any finite $P \subseteq \boldsymbol{SD}$, which formula culminates Suszko's discussion of the independence of states of affairs (Suszko 1968, Axiom 8.16). Axiom 5 is the linguistic expression of the thesis of truth-functionality (Sections 3.6 and 4.6). Notice that the expressibility of these theses within $\boldsymbol{L}_D$ crucially depends on the presence of $\Box$ and $\bigwedge$.

There is one little problem involving Axioms 4 and 5: because of their presence one might hesitate to regard *DML* as a logic at all. According to some definitions, e.g., one given by Perzanowski (1985), a logic should be closed under substitution. But in Axioms 4 and 5 $\bigwedge \boldsymbol{SD}$ may *not* be replaced by any arbitrary sentence $q \in \boldsymbol{L}_D$. Indeed, closure under substitution would have rather unpleasant consequences here. In this case, Axiom 4 would imply $\Diamond(p \wedge \neg p)$ and as we have $\vdash \neg\Diamond(p \wedge \neg p)$ thus bring about the inconsistency of *DML*. On the other hand, Axiom 5 would in this case imply $\Box((p \vee \neg p) \supset q) \vee \Box((p \vee \neg p) \supset \neg q)$, whence $\Box q \vee \Box\neg q$, whence $q \equiv \Diamond q \equiv \Box q$, and thus entail a collapse of *DML* to propositional or "Fregean" logic (cf. Suszko (1968, pp. 11–12)). Now if one insisted on this point, we could remove Axioms 4 and 5 and reintroduce their necessitations as special extra-logical postulates; or we could introduce a set $\mathbf{Z} = \{z_i : 0 \leq i \leq \omega\}$ of special sentential variables playing the role of conjunctions of state-descriptions, and replace Axioms 4 and 5 by $\Diamond \bigwedge z_i$ and $\Box(z_i \supset p) \vee \Box(z_i \supset \neg p)$, respectively. However, the difference seems to be merely a terminological one, for which reason we shall simply call *DML* a logic.

Some interesting formulae of *DML* (easily provable by Theorem 9) are the following:

(a) $\Box \bigwedge P \equiv \bigwedge_{p \in P} \Box p$;
(b) $\Box(x)px \equiv (x)\Box px$ (the Barcan formula and its converse);
(c) $\Diamond(\bigwedge \boldsymbol{SD} \wedge p) \equiv \neg\Diamond(\bigwedge \boldsymbol{SD} \wedge \neg p) \equiv \Box(\bigwedge \boldsymbol{SD} \supset p)$.

It will be noticed that *DML* is rather weak as a doxastic logic, much weaker, in fact, than contemporary doxastic logics based on possible worlds semantics (e.g., Lenzen (1980)). For example, we cannot prove any formula of the form $D_A p$. One of the few positive facts that may be noted is that Axiom 6 clearly reveals the interdependence of thought-ascriptions (Section 4.4). Despite its weakness, one nevertheless may find *DML* too strong: doesn't Axiom 6 imply the "logical

---

sets of *DML*. $\sqcup$, $\sqcap$ and $-$ are set-theoretical intersection (*sic*), union and complementation, respectively; $\mathbf{1} = \emptyset$, $\mathbf{0}$ = the set of all maximally consistent sets. $\mathbf{O} = \boldsymbol{N}$, δ is identity, $\sigma(p) = \{P \subseteq \boldsymbol{L}_D : P \text{ is maximally consistent and } p \in P\}$, and $\psi_A(\mathbf{s}) = \{p : D_A p \in \bigcap \mathbf{s}\}$, where $\mathbf{s}$ is a set of maximally consistent sets. Pictures may be identified with sentences, and these may be identified with arbitrary elements of $\mathbf{S}$ in turn. The model defined in this way is a genuine Tractarian interpretation. Axioms 4 and 5 jointly guarantee that Condition 3 on $\mathcal{S}$ is satisfied; Axiom 6 guarantees that ψ_A and $\sigma(D_A p)$ satisfy Definitions 19 and 20. Since p may be shown to be invalid in this interpretation, the theorem is proved. Cf. Keisler (1971, Ch. 4), on $L\omega_1 0$, and Chellas (1980) on $S5$ and E (the latter is similar to the doxastic fragment of *DML*).

omniscience" of A? For an argument that it does not really do so, we refer to Stalnaker (1976)

5.3. Some Correspondence Results. *DML* turns into a stronger and more interesting doxastic logic if some additional restrictions are imposed on the interpretations. This is apparent from the following correspondence theorems:

Theorem 10: Axiom $(D_Ap \wedge D_Aq) \supset D_A(p \wedge q)$ corresponds to the condition that $\boldsymbol{T}_A(\mathbf{s})$ is closed under finite conjunction.[34]

Theorem 11: Axiom $\bigwedge_{p \in P} D_Ap \supset D_A \bigwedge P$ ("systemic nature of thought") corresponds to the condition that $\boldsymbol{T}_A(\mathbf{s})$ is closed under *arbitrary* conjunction.[35]

Notice that this axiom in turn implies the doxastic Barcan formula $(x)D_Apx \supset D_A(x)px$. Therefore we have the remarkable result that the *Einheitlichkeit* of the soul entails the derivability of the doxastic Barcan formula (in complete axiomatizations)!

Theorem 12: Axiom $D_Ap \supset D_AD_Ap$ ("self-reflexivity of thought") corresponds to the condition that if $t \in \boldsymbol{T}_A(\mathbf{s})$, then there is a $t' \in \boldsymbol{T}_A(\mathbf{s})$ such that $\sigma(t') = \bigcap\{\mathbf{s} \in \mathbf{S} : t \in \boldsymbol{T}_A(\mathbf{s})\}$.

To put it more transparently: the axiom "if A thinks that φ then A thinks that A thinks that φ" corresponds to the condition that no thought belongs to the soul unless the soul contains a thought which *says that* this thought belongs to the soul.

However, the *Tractatus* does not contain the informal counterparts of any of these additional semantical postulates or corresponding axioms. The reason is clear: we are here once more dealing with issues which are to be settled by empirical investigation (in this case: psychology), not by logic.[36]

6. Conclusion

This ends our tortuous path through the Tractarian labyrinth. We certainly have not discussed all topics we might have treated: for example, Wittgenstein's views on functions and the theory of types may presumably also be handled by formal means. However, with the above the foundations of formal Tractarian semantics have been laid; in particular, we have achieved our goal of giving simple truth-functional analyses of quantification, the modalities and the propositional attitudes, which is something previous commentators generally considered impossible. This may suffice for a first start.

How does the *Tractatus* look in the light of our formal analysis? From a general point of view, we have obtained a better idea of the general nature of the work: it anticipates Tarski's and Carnap's later work, but it does so in a rather apodictic way. Deriving the consequences of the statements and clarifying their interrelations is a task almost exclusively left to formal analysis. When carrying out the latter, several weaks points emerge. For example, we encountered various inconsistencies; moreover, the work contains various lacunae which must be filled in by our own imagination—just recall the silence of the *Tractatus* on the syntax and semantics of non-elementary pictures and the conspicuously absent answer to TLP 5.54–5.541.

[34] That is, if $t \in \boldsymbol{T}_A(\mathbf{s})$ and $t' \in \boldsymbol{T}_A(\mathbf{s})$, then $(t \wedge t') \in \boldsymbol{T}_A(\mathbf{s})$. (Here "$\wedge$" is a "mental connective" in the sense of Harman (1973).)

[35] The appellation "systemic", which means approximately the same as our *einheitlich*, is due to Routley & Routley (1975). According to the Routleys all thought is systemic. In the next note we shall see that it is not.

[36] For example, the question as to whether the soul is unified can only be settled by empirical research. (Recent research suggests it is not always unified: "split-brain" patients display manifestly non-systemic thoughts, perceptions and memories, as has been noticed by Barwise (1981).)

Nevertheless, however crudely it may sometimes have been formulated, the work contains much that is still of interest today. Thus, the quantified logic we have extracted from it is a complete higher-order logic, more comprehensive than standard first-order logic because it treats predicates as names; the doxastic logic we arrived at is perfectly acceptable to contemporary doxastic logicians of the "possible worlds" persuasion; the Tractarian semantics of propositional attitude ascriptions strikingly anticipates modern "language of thought" theories; and the Tractarian thesis of the complete describability of the world by elementary sentences is a variant of the doctrine of the supervenience of the mental on the non-mental which is currently coming into vogue.

From a philosophical point of view, the Tractarian philosophy of mind is the most interesting subject we have discussed: here the *Tractatus* manages to combine two (unrelated) theories—the "language of thought" and "supervenience" theories—which certainly do not yet look antiquated today. This is not to say that there are no differences with these modern theories. First, Wittgenstein does not wrestle with the much-debated problem of current "language of thought" theories as to whether thoughts are iconic (picture-like) or discursive (sentence-like), because according to him sentences are iconic too. Secondly, modern psychophysical supervenience theories postulate the supervenience of propositional attitude ascriptions on physical descriptions of the world without imposing further restrictions on these descriptions. The Tractarian theory is more specific in asserting that the physical descriptions in question are, in the final analysis, elementary sentences. Thirdly, according to Wittgenstein meanings are conferred to pictorial elements by the "metaphysical subject", whereas modern theories seek to provide *causal* accounts of the attribution of meanings. To Wittgenstein the latter route is blocked, for he regards believing in causality as a "superstition" (TLP 5.1361, 6.32 ff.).[37] The main weakness of Wittgenstein's account seems to lie in the latter feature, although this may still appeal to some philosophers of a metaphysical bent. However, apart from this feature (which does not belong to logic anyway) the Tractarian theory seems no less attractive and viable than its related modern successors.

All in all, we think our effort has made it clear that the *Tractatus* may fruitfully be discussed in formal terms; the formal approach uncovers various viewpoints which are still interesting in their own right and thereby justifies a greater appreciation of the work than would otherwise be warranted. One may not always agree with the specific form our interpretation has taken: but even in this case a formal account has the advantage over an informal one that it may at least be precisely understood *what* it is one does not agree with. So even if our analysis is not *unantastbar und definitiv*, it may at least facilitate further understanding. *Mögen andere kommen und es besser machen!*

Appendix (1992)

A letter from G. Kreisel (Baden, Switzerland, October 21, 1990; quoted with permission) puts some things we have said in the above in a slightly different light. Kreisel writes as follows about a conversation which he had with Wittgenstein "probably back in 1942":

> (a) It seemed to me too obvious even to mention that Tractatus was concerned with a Boolean algebra; specifically, the algebra generated from the *simples* as elements. Also Wittgenstein would have been horrified at such (for him) pretentious language: one spoke of propositional combinations.

[37] See Kenny (1984) for an exposition of the Tractarian view, and Field (1978) as a protagonist of the modern approach.

(b) His (only) remark to me was in reply to an observation. I said that what I found in Tractatus was compelling only if one assumed that there were *finitely* many simples. Otherwise things became contrived here and there.

(c) As so often, Wittgenstein seemed (to me) quite unduly pleased with me (and at the time I had no idea, why). He said something to the effect: 'Of course, I thought of the primitive case, and if things are clear there, the rest will look after itself. If a foundational scheme doesn't work out as simply as it looks, it's no good at all.'

(d) Today I think I know what he liked about (b). It was a straightforward comment without any agonizing. In normal circumstances this would not be much to write home about, but in 'exact philosophy' a modicum of a sense of proportion was a rarity; just think of Carnap's style (or Tarski's in the 30's, not after the 50's). Besides, when it comes to agonizing, few can match Wittgenstein's particular talent for this activity.

In other words, Wittgenstein seems to have been thinking of a situational Boolean algebra which is generated by a finite number of *Sachverhalte*. In such an algebra, there is only a finite number of *Sachlagen* and only a finite number of worlds.

It is hard, if not impossible, to reconcile this view with TLP 4.463, in which it is said that logical space is "infinite".

It is, however, not difficult to modify our reconstruction in the appropriate way: clause (b) of Definitions 3, 7, and 11 should be changed in such a way that it becomes true that $Card(\mathbf{SA}) = Card(\boldsymbol{EL}) = Card(\boldsymbol{ET}) < \aleph_0$. As a result, the first point which is made in footnote 27 becomes relevant, and the whole construction would indeed become "less contrived here and there".

The above has no effect on what is said in Definitions 2, 6 and 10. We shall have to allow for the possibility that some "objects" have certain relations to themselves. Therefore we have to admit proto-*Sachverhalte* (in the sense of Definition 2) like $\mathbf{b} * \mathbf{a} * \mathbf{a}$, $\mathbf{b} * \mathbf{a} * \mathbf{a} * \mathbf{a}$, and so on, and perhaps even $\mathbf{a} * \mathbf{a}$ in case $\mathbf{a}$ is a property which applies to itself.

CHAPTER 3

Truth-Functionality and Supervenience in the *Tractatus*

Abstract

The purpose of this paper is twofold. First, I want to point out that Wittgenstein's *Tractatus* contains a clear and remarkably modern example of a theory of supervenience. And secondly, I want to argue that this theory of supervenience may be interpreted as a weak form of a principle of truth-functionality—which may exactly be the form of this principle which Wittgenstein himself had in mind.

First published in P. Weingartner & G. Schurz, eds., *Philosophy and Natural Science: Borderline Questions. Reports of the 13th International Wittgenstein Symposium*, Hölder-Pichler-Tempsky, Vienna, 1989, pp. 276–278. ISBN 3–209–00862–0.

The purpose of this paper is twofold. First, I want to point out that Wittgenstein's *Tractatus* contains a clear and remarkably modern example of a theory of supervenience. And secondly, I want to argue that this theory of supervenience may be interpreted as a weak form of a principle of truth-functionality—which may exactly be the form of this principle which Wittgenstein himself had in mind.

1. Supervenience

"Supervenience" means something like "dependence". When something depends on another thing, we also say that that thing supervenes on the other one. Modern supervenience theorists see a lot of relations of supervenience between various domains. For example, it is often said that the moral supervenes on the non-moral. When two organisms are alike in all non-moral respects, they cannot possibly differ in some moral respect; when someone's biography is exactly the same as that of the present president of Austria in all biological, psychological and other non-moral respects, he or she cannot fail to have the same moral status as that president has (whichever that may be). No moral difference without some other difference. It may be the same with the mental: any exact physical duplicate of me must necessarily have precisely the same mind as I have. No mental difference without some physical difference.

The interesting thing about supervenience is that it is a much weaker notion than reducibility. Moral facts may depend on non-moral ones, but no one may be able to spell out the dependence of the supervenient superstructure on the basis in detail; no one may be able to give a "reduction" in terms of non-moral facts. Similarly, the mental may fully depend on the physical, but no one may ever be able to describe the mechanism (or logic?) of the dependence in detail. This is an advantage of the notion, for physicalists often want to defend only a vague, general form of dependence, without wanting to posit any lawlike relationships.

Haugeland has given a definition of supervenience which nicely fits our purposes:

> Two worlds in a class of possible worlds are *discernible with* a given language just in case there is a sentence of this language which is true at one, and not at the other. [...] A language *weakly supervenes on* another language (relative to a class of possible worlds) just in case any two worlds in the class of possible worlds which are discernible with the former language are also discernible with the latter. ((Haugeland 1982), p. 97.)

2. Supervenience in the *Tractatus*

On seeing the above definition, any reader of the *Tractatus* will immediately be reminded of section 4.26 of this work:

> If all true elementary sentences are given, the result is a complete description of the world. The world is completely described by giving all elementary sentences, and adding which of them are true and which false.

In other words, when you take any "possible world" (as Wittgenstein calls it: *Notebooks* 19.9.1916) different from "the real world", there is always at least one elementary sentence which is true in only one of both worlds. Wittgenstein does not state whether he would want to apply this principle to *all* possible worlds, but we may safely assume that he did. So let us extend *TLP* 4.26 to the thesis that worlds may always be discerned by elementary sentences. No difference between worlds without some elementarily describable difference.

Applying Haugeland's definition, it will be clear that we then have a principle of supervenience here. Everything which may be discerned with the whole language may already be discerned by means of the elementary sentences alone, and therefore the whole language supervenes on its subset of elementary sentences. We may extend the definition and likewise say that all facts supervene on elementary facts (*Sachverhalte*, described by elementary sentences), that all properties supervene on elementary properties (attributed by elementary sentences), etc.

3. Truth-functionality as supervenience

Wittgenstein states his thesis of truth-functionality in *Tractatus* 5–5.01:

> The sentence is a truth-function of the elementary sentences. (The elementary sentence is a truth-function of itself.) The elementary sentences are the truth-arguments of sentences.

The principle of truth-functionality as stated here is weaker and more liberal than the definitions which we have become accustomed to nowadays. According to present-day definitions, a sentence cannot be truth-functional unless its truth-value is some function of the truth-values of the *subsentences* it contains and of the way it is built up from these. The truth-values of other sentences do not matter. However, the *Tractatus* does not say that only the truth-values of the elementary *subsentences* of a sentence matter as to its truth-value. The latter truth-value may as well be a function of the truth-values of *all* elementary sentences.

Let us, for the moment, interpret the Tractarian principle of truth-functionality in the latter way. Thus, this principle asserts that the truth-values of all elementary sentences (not necessarily only the ones contained in the sentence as subsentences) jointly determine the truth-values of all sentences. Given what Carnap called a state-description—a set which, for each elementary sentence, contains either this sentence or its negation, and no other elements—any sentence may assume only *one* truth-value.

When we accept this weak formulation, the principle of truth-functionality is easily seen to follow from the supervenience principle we have just mentioned. For if the thesis of truth-functionality did not hold, two worlds could verify the same elementary sentences and yet differ as regards the truth-value of some other sentence. These worlds would hence not be *completely* described by elementary sentences and violate the principle of supervenience.

The converse implication does, of course, not hold: we may conceive of a language which is purely truth-functional but unable to describe any one world completely. Such a language would, however, not be in accordance with the *Tractatus*.

4. The Tractarian principle of truth-functionality

Now did Wittgenstein really have such a weak principle of truth-functionality in mind, or did he accept the principle in its stronger, full-blown modern version? I think it is hard to find evidence for the latter view. Wittgenstein never explicitly banishes modal and doxastic constructions from the ideal language he had in mind. They seem to be perfectly in order, provided they are truth-functional; for otherwise the principle of supervenience of language on elementary language would be violated. And why should they be prohibited, after all? This would not only lead to a drastic impoverishment of language (which is nowhere explicitly advocated in the *Tractatus*), there is also no justification for it on syntactic grounds. Modalizing a sentence is no more mysterious than negating it. And indeed, most of Wittgenstein's remarks on syntactic operators seem to apply equally well to all operators, including modal and other ones which we do not longer call truth-functional nowadays.

Thus, I believe Wittgenstein's principle of truth-functionality, which would be unduly restrictive otherways, may best be regarded as being only such a rather weak claim. This also explains why Wittgenstein himself is so silent about the truth-functionality of, e.g., intentional ascriptions. They are simply no exception to the rule. They differ from what we still call truth-functional compounds nowadays in that their truth-values are not fully determined by the subsentences they contain and the way they are built up from these alone; the additional factors that play a role (according to the *Tractatus*) are spelled out in (Lokhorst 1988 *a*). But they are truth-functions of the elementary sentences nonetheless, so there is no need to pay special attention to them. It is the same with modal sentences: their truth-values are the same in all worlds and in all interpretations which are in accord with the demands the *Tractatus* puts on such interpretations, as I explain in (Lokhorst 1988 *a*), and so they are trivially truth-functional. (Unlike the case of intentional ascriptions, their truth-values *are*, however, determined by their structure alone).

5. In defense of the Tractarian view of truth-functionality

Far from being a drawback, the broad, general nature of the Tractarian principle of truth-functionality is in fact a point in its favour. It is an enviable position to be able to claim that there are some sentences which are basic from an epistemological or scientific point of view, and which jointly determine all truths and falsehoods, without being obliged to say how they manage to do just that. The long history of failed attempts at giving explicit truth-functional definitions of modal and intentional language may indeed suggest that it is well-nigh impossible to do the latter, while the thesis yet remains attractive in its own right. Thus, Wittgenstein may have been wise in going no further.

Modern theories of supervenience, which also argue for dependence without committing themselves to reducibility, stem from the same motivation and share the same appeal. These theories show that it is possible to be precise and yet not

too specific when giving physicalistic accounts of morality and mind. The *Tractatus* shows that the same may be done when one is formulating sweeping statements on the nature of the semantical relations between various kinds of sentences.

CHAPTER 4

Wittgenstein on the Structure of the Soul: A New Interpretation of *Tractatus* 5.5421

Abstract

Tractatus 5.542–5.5421 should be read as follows: anything which represents is complex; the soul is simple; so 'the superficial psychologists of the present day' are mistaken when claiming that the soul represents anything. In contrast to the 'empirical self', with which psychology is concerned, the 'metaphysical' or 'transcendental' soul, subject, or self is a purely fictitious entity (or rather, non-entity) which does not have any positive function.

First published in *Philosophical Investigations*, **14** (4): 324–341, October 1991. ISSN 0190–0536.

0. Introduction

What is the most obscure remark in Wittgenstein's *Tractatus*?[1] We have an *embarras de choix*, but TLP 5.5421 may well lay claim to this epithet. In this passage, Wittgenstein tells us which implications his analysis of propositional attitude ascriptions in TLP 5.542 has for our views on the 'soul' ('mind') or 'subject':

5.542 It is clear, however, that 'A believes that p', 'A thinks p', 'A says p' are of the form "'p" says p': and this does not involve a correlation of a fact with an object, but rather the correlation of facts by means of the correlation of their objects.

5.5421 This shows too that the soul[2]—the subject, etc.—as it is conceived in the superficial psychology of the present day is a monstrosity. For a composite soul would no longer be a soul.

At least four conflicting interpretations have been given of TLP 5.5421:

1. Some commentators claim that it shows that Wittgenstein regarded the mind as a *complex* entity: it is a Humean 'bundle or collection' of thoughts.[3] Perszyk calls this interpretation the 'standard reading' of the passage.[4]
2. According to others, Wittgenstein precisely *rejected* this view: according to them, he wanted to say that the mind is *simple*.[5]

[1] L.J.J. Wittgenstein, *Logisch-philosophische Abhandlung: Kritische Edition*, ed. by B.F. McGuinness and J. Schulte, Frankfurt am Main, Suhrkamp, 1989. Henceforth referred to as 'TLP'.

[2] Wittgenstein uses the word 'Seele'. As S.A. Kripke notes in his *Wittgenstein on Rules and Private Language*, Cambridge, Mass., Harvard University Press, 1982, pp. 49 and 127, 'mind' is often a better translation of this word than 'soul' because it has less religious and philosophical connotations. Since most readers of the English version of the Tractatus will be accustomed to the translation 'soul', we will use the latter term.

[3] For example, J. Hintikka, 'On Wittgenstein's "Solipsism"', *Mind* 67 (1958), 88–91, and M. Black, *A Companion to Wittgenstein's 'Tractatus'*, Ithaca, N.Y., Cornell U.P., 1964, p. 301.

[4] K.J. Perszyk, 'Tractatus 5.54–5.5422', *Philosophia* 17 (1987), 111–126, quotation from p. 117.

[5] G.E.M. Anscombe, *An Introduction to Wittgenstein's Tractatus*, 4th ed., London, Hutchinson, 1971, p. 88. J. Hintikka and B. Wolniewicz, personal communications, 1989.

3. Some commentators want to have it *both* ways. For example, Hacker enigmatically writes: 'The claim should be interpreted thus: the soul conceived of as a unitary simple subject does not exist. But conceived of as a manifold, it is the legitimate subject-matter of psychology'.[6]
4. Sluga, on the other hand, thinks that Wittgenstein wanted to show that the soul is *neither* complex *nor* simple. He credits him with 'the discovery that the notion of the soul or subject is altogether incoherent and that, consequently, there cannot be any such thing'.[7]

Kripke keeps an open mind: 'the obscure passage *Tractatus* 5.5421 ... does not appear to be directed primarily at Hume's theory'.[8] Some commentators see no solution at all: Urmson speaks of 'almost impenetrable obscurity'.[9] Others see no *problem* at all: 'All this is commonplace and evident', Rosenberg remarks after quoting the relevant passages.[10]

The great variety of existing interpretations suggests that the key to TLP 5.5421 has hitherto not been found. After having thought about the passage for many years, we believe that we have finally stumbled across a solution. We think that our interpretation is the first one which really does justice to the passage, in the sense of reconstructing it as a clear and cogent argument. Moreover, it has all other virtues any good interpretation of the passage should have: it is in accord with the textual evidence, it does not make implausible assumptions about Wittgenstein's knowledge of 'the psychology of the present day', it does not treat the passage as just an isolated remark, and it suggests what the historical and systematical sources of his view may have been. Finally, our interpretation shows that the passage is more important than previous commentators have thought: it turns out to be closely connected with the remarks on the perspectival character of the 'metaphysical subject' in TLP 5.631 ff.

All this does not imply, however, that our interpretation of TLP 5.5421 is a *justification* of it. We think that it is an unfortunate remark, which is false (or misleading at best) and should never have made its way into the *Tractatus*.

1. Thoughts

As TLP 5.5421 is presented as a corollary of TLP 5.542 we have to understand the latter passage first. Fortunately, this is not too difficult.

Wittgenstein begins by saying that 'A thinks p' is of the same form as '"p" says p'. What is the form of '"p" says p'? Well, it is an ascription of a fact to a sentence: it says that the sentence 'p' is a picture of the fact that p. If 'A thinks p' is to be similar to '"p" says p', it should also say that some sentence is a picture of the fact that p. It does so if we analyze it as saying that one of A's *thoughts* is a picture of the fact that p. We know that Wittgenstein considered thoughts as similar to sentences; they consist of 'psychical constituents that have the same sort of relation to reality as words'.[11] So 'A thinks p' is in its analyzed form indeed analogous to '"p" says p': both involve the ascription of a fact to a sentence-like picture.

---

[6] P.M.S. Hacker, *Insight and Illusion*, 2nd ed., Oxford, Clarendon Press, 1986, p. 62.

[7] H. Sluga, 'Subjectivity in the Tractatus', *Synthese* 56 (1983), 123–139, quotation from pp. 129–130.

[8] Kripke, op. cit., p. 131.

[9] J.O. Urmson, *Philosophical Analysis*, Oxford, Blackwell, 1956, p. 133.

[10] J. Rosenberg, 'Intentionality and Self in the Tractatus', *Noûs* 2 (1969), 341–358, quotation from p. 342.

[11] Letter to Russell dated 19.8.19, reprinted in L.J.J. Wittgenstein, *Notebooks 1914–1916*, edited by G.H. von Wright and G.E.M. Anscombe, 2nd ed., Oxford, Blackwell, 1979, p. 131.

After remarking that 'A thinks p' is of the same form as '"p" says p', Wittgenstein says that both sentences involve a 'correlation of facts by means of the correlation of their objects'. This remark may be understood if we consider how a picture represents a fact. It is capable of doing so because it is a fact itself: it is the fact that certain pictorial elements are arranged in a certain way. Each pictorial element denotes a corresponding object. The fact that the pictorial elements are arranged in such-and-such a way says that another fact is the case, namely that the objects which correspond to the elements are arranged in the same way. So when a thought represents a fact, we have 'a correlation of facts' (namely the thought on the one hand, and the fact which is pictured by the thought on the other hand) 'by means of a correlation of their objects' (namely the 'psychical constituents' of the thought on the one hand and the objects of the pictured fact on the other hand), which is precisely what TLP 5.542 asserts.

It is important to notice that pictures (including thoughts) are always complex: the fact that certain elements are arranged in a certain way necessarily involves at least two elements. We may call the number of components of a fact its 'logical complexity' (TLP 4.04). A picture and the fact which it represents have the same 'logical complexity', for there are no 'multi-purpose' pictorial elements denoting several objects or 'empty' pictorial elements which do not denote an object. This explains why a thought-ascription 'does not involve a correlation of a fact with an object': an object is simple and does not have the 'logical complexity' which is required to represent a fact.

The above is hardly controversial. Dozens of commentators have offered similar analyses in the past three decades. The only peculiar feature of our interpretation is that we want to read TLP 5.542 as a proposal for a *semantical* analysis of thought-ascriptions. One should not say that 'A thinks p' is *equivalent* to 'one of A's thoughts says p'. For if we did so, thought-ascriptions would become nonsensical. In the ideal language Wittgenstein had in mind, it cannot be expressed that a certain picture represents a certain fact; any attempt to do so would result in metalinguistic nonsense. '"p" says p' is a good example of such nonsense: it tries to *say* what can only be *shown*. It is a correct expression in the ladder-language in which the *Tractatus* is written, but it cannot belong to the object-language. However, there seems to be no good reason to exclude thought-ascriptions from language. Therefore, 'A thinks p' should not be regarded as being *equivalent* to any sentence of the form '"p" says p', 'thought T says p' or suchlike. Rather, it *says that* one of A's thoughts says that p; it *is true* iff one of A's thoughts says that p. In this way thought-ascriptions remain ordinary sentences of language. An additional attractive feature of a semantical analysis is that it may easily be transformed into a *truth-functional* analysis.[12]

2. The Empirical Self

How does the above account apply to TLP 5.5421? Let us begin by considering the first interpretation on the list presented above: the view that Wittgenstein considered the soul or subject as being identical with a Humean bundle or collection of thoughts.

We may call such a bundle of thoughts the 'empirical self'. This is the self which may be studied by psychology (TLP 5.641), and its identification with a bundle of thoughts explains why it may indeed be studied empirically. There even are *two*

[12] G.J.C. Lokhorst, 'Ontology, Semantics and Philosophy of Mind in Wittgenstein's Tractatus: a Formal Reconstruction', *Erkenntnis* 29 (1988), 35–75, and 'Truth-functionality and Supervenience in the Tractatus', in P. Weingartner and G. Schurz, eds., *Reports of the 13th International Wittgenstein-Symposium*, Vienna, Hölder-Pichler-Tempsky, 1989, 276–278.

ways in which psychology may study the empirical self. First, thoughts are facts in the world. They may therefore be investigated in the same way as other facts in the world are studied. And secondly, thoughts may be studied by means of what they represent, that is, by way of their contents. The latter approach would presumably be the method of introspective (or at any rate 'phenomenalistic') psychology.

There is no way to bridge the gap between the results of both methodologies. Psychology can never discover whether a given thought represents a given fact, for a thought can only *show* what it represents, and any sentence that says that this thought represents a certain fact can only be metalinguistic nonsense which does not belong to language. Psychophysical laws in the sense of sense-ascribing laws are as impossible to formulate as semantical laws are.

3. The Simplicity of the Soul (Subject)

According to the adherents of the first standpoint on our list, the soul or subject which is mentioned in TLP 5.5421 is identical with the empirical self which we have defined above. However, there are at least two reasons why this interpretation cannot be correct.

In the first place, Wittgenstein emphatically says that 'a composite soul would no longer be a soul'. The empirical self, however, is complex. It is even *doubly* complex: it is complex because it is a bundle of several thoughts, and it is complex because each thought is complex in itself. Therefore the soul cannot be identical with the empirical self.

There is only one way to counter this objection. One might maintain that Wittgenstein does not say that he himself regards the soul as complex; he merely ascribes this view to the superficial psychologists of his time. Thus, TLP 5.5421 should be read as 'A composite soul would no longer be a soul *according to the superficial psychologists of the present day*'. However, this suggestion does too much violence to the text. The text certainly gives the impression that Wittgenstein proclaims the just-mentioned view as his own standpoint.

In the second place, this interpretation charges Wittgenstein with having a caricatural view of the psychology of his time. There were almost no psychologists in his time who considered the soul, mind or subject as simple. The view that it is a bundle of thoughts was the most popular view in psychology around the turn of the century. We cannot give a full survey of the literature here, but only refer to Weininger's *Geschlecht und Character*, a book which Wittgenstein read and admired highly. Weininger heaped abuse on the psychologists of his time precisely because they regarded the soul as being nothing but a bundle of thoughts. He argued that this view is fine as far as women are concerned, but patently false when one takes male Caucasian geniuses into account.[13] It is quite possible that Wittgenstein got his conception of the psychology of his time from Weininger; there are more traces of his influence.[14]

In sum, we seem to be obliged to accept the second interpretation on our list: the subject or soul is simple. Although it is simple, 'the I is no object' (*Notebooks* 7.8.16): for it 'does not exist' (TLP 5.631) and 'does not belong to the world (TLP 5.632). We will return to a discussion of its nature below.

The above interpretation is not only in accord with what the last sentence of TLP 5.5421 literally says, it is also in accord with the Schopenhauerian remarks in TLP 5.64 about the shrinking of the subject into a point without extension.

---

[13] O. Weininger, *Geschlecht und Charakter: Eine prinzipielle Untersuchung*, Vienna and Leipzig, Braumüller, 1926, Part II, Chaps. 6–9.

[14] See, e.g., Sluga, op. cit. The remark in TLP 5.1362 that "'A knows that p is the case" is senseless when p is a tautology' may also be an echo from Weininger. He wrote that a tautology does not express knowledge and cannot be the object of an act of thought (ibid., part II, ch. 7).

4. How Wittgenstein Changed his Mind

Wittgenstein radically changed his mind on the issue of the complexity of the subject between the time he dictated his notes to Moore (1914) and the time he wrote the final version of the *Tractatus* (1918). The last sentence of the notes dictated to Moore says: 'The relation of "I believe p" to "p" can be compared to the relation of "'p' says (besagt) p" to p: it is just as impossible that *I* should be a simple as that "p" should be'.[15]

Here Wittgenstein still adhered to the view that the subject is a complex entity similar to the empirical self we mentioned above. However, it is obvious that he says just the opposite in TLP 5.5421. He probably changed his mind as a result of reading Schopenhauer, who wrote in *Die Welt als Wille und Vorstellung* that the self is 'an indivisible point' which is the 'centre of all existence'. (The influence of Schopenhauer is very conspicuous in the *Notebooks*.) Confusing TLP 5.5421 with the just-mentioned remark in the notes dictated to Moore is one of the main causes of the 'standard' misinterpretation of TLP 5.5421.

5. The Soul (Subject) Does Not Represent

We have reached the conclusion that the soul or subject is simple. However, as we have already seen in our list of previous commentaries, Sluga claimed that the Tractarian soul *cannot* be simple. To quote him in full:

> He [i.e., Wittgenstein] raises a crucial difficulty for all those who argue that only a simple substance can have mental attributes. If among those attributes is the ability to have representations and if representations of complexes are, by nature, themselves complex, we must ask how a *simple* substance is capable of having *complex* representations.
>
> But that suggestion, combined with the claim that a composite soul is not a soul any longer, seems to lead to the discovery that the notion of the subject is altogether incoherent and that, consequently, there cannot be any such thing.[16]

How do we reply to Sluga's analysis?

Sluga certainly draws attention to an important point. As he makes clear, Wittgenstein's account of representation is, for example, obviously incompatible with Leibniz's theory of simple monads which are 'mirrors of the world'. According to the picture theory there can be no simple monads which represent complex facts. The Many cannot be represented by the One.

However, a similar critique does *not* apply to Wittgenstein's own view of the soul as a simple entity. For he *never says that the soul represents*. Indeed, it is just the other way around: we want to suggest that TLP 5.5421 may most naturally be read as a *straightforward rejection of the view that the soul represents*.

When we adopt this suggestion, the line of argument in TLP 5.542–5.5421 suddenly becomes crystal-clear. Wittgenstein begins by noting that anything which represents a fact must be complex. From this he draws two conclusions. First, that Russell's and Moore's conception of the soul as an (indivisible) object is false. And secondly, that the conception of the psychologists of his time his wrong. For these regard the soul as something which represents (or as a collection of representations). But that implies that it is complex, and that is absurd, for 'a composite soul would no longer be soul'. Therefore the psychologists are wrong in saying that the soul represents.

---

[15] *Notebooks*, op. cit., p. 119.

[16] Sluga, op. cit., pp. 129–130.

This, then, is our new analysis of TLP 5.5421: it is a *reductio ad absurdum* of the view that the soul or subject is a representational entity. The soul is no 'mirror of nature'; rather, it is a windowless Leibnizian monad which differs from a true Leibnizian monad in that it does not represent anything. The soul may perhaps be related in some way to thoughts which represent facts (although it is not clear how we could conceive of such a relation), but it does not represent facts itself.

Besides reconstructing TLP 5.5421 as a clear and cogent argument, our interpretation has at least two other advantages over the 'standard' interpretation.

First, it does not imply that Wittgenstein had a superficial knowledge of the psychology of his times. It is undoubtedly correct to suppose that most psychologists thought that the mind has a representational character. Indeed, most psychologists still think so today.

Secondly, our interpretation fits in very nicely with TLP 5.631. This passage may even be regarded as the clincher for our analysis, for here Wittgenstein unequivocally reaffirms the conclusion we have just ascribed to him: 'There is no such thing as the thinking, representing subject'. Even if there were a subject, it could not think or represent, for then it would be complex. In the same passage, Wittgenstein goes on to assert that 'in an important way there is no subject' at all. This statement also supports our thesis that the Tractarian subject does not represent: for it would be hard to imagine how something that does not exist could represent anything.

Thus, even if our suggestion may seem strange at first, there really is no way to avoid it!

6. The Idleness of the Soul (Subject)

The soul or subject as it is portrayed by Wittgenstein is a pretty useless entity. It does not represent itself and cannot represent any other fact either. Nor does it play any role in the attribution of propositional attitudes: 'A thinks that p' is true iff some of the facts (thoughts) constituting the person A represent the fact that p. The soul does not come in anywhere. Moreover, the Humean argument in 5.631 ff. is designed to show that the soul cannot be represented by any other instance either. Thus, it seems to play no role in representation whatsoever.

Some commentators deny this: they argue that the soul is the *necessary precondition* for representation. According to them, it is the instance which confers meanings on symbols. Pictures, thoughts and sentences would be senseless, lifeless facts in the world if the metaphysical subject did not provide pictorial elements with denotations.[17]

However, this suggestion is not tenable. In the first place, it is utterly incomprehensible. It is inconceivable how a metaphysical, non-worldly instance could ever imbue facts in the world with worldly meanings. It should at least partly 'reach out' to the facts ('stick its fingers into the world') in order to do this. Moreover, it cannot be *simple* if it is to do this. It should have some internal structure corresponding to the structures of the facts which it is relating to each other, for it would not be able to distinguish between different facts otherwise.

In the second place, not the slightest evidence for the suggestion can be found in the *Tractatus*. Wittgenstein discusses only *interpreted* symbols. The interpretation is always considered as given beforehand. 'The pictorial relation which turns a fact into a picture belongs to the picture itself.' (TLP 2.1513) It is not necessary to invoke a soul as a *deus ex machina* which has to give meanings to symbols, for

[17] See, for example, A. Kenny, 'Wittgenstein's Early Philosophy of Mind', in I. Block, ed., *Perspectives on the Philosophy of Wittgenstein*, Oxford, Blackwell, 1981, 140–147.

symbols are meaningful from the very start. So 'let's not imagine the meaning as an occult connection the mind makes between a word and a thing'![18]

Similar objections apply to the suggestion that the soul is the instance which *has* thoughts. It is inconceivable how an extramundane entity can *have* facts in the world. Furthermore, there is no reason why some extramundane entity *should* have them. 'Thoughts think themselves', in the sense that everything that goes on in thinking is completely accounted for by thoughts and what they represent. A soul could only be an inert homunculus, a bystander seeing nothing and doing nothing.

So the hypothesis of the soul or subject seems to be completely superfluous. It does not have any useful property whatsoever beside its simplicity—if that may be called a useful property. With his assumption that the soul is simple, Wittgenstein puts it completely out of action. As William James said in his criticism of the doctrine of Transcendental Egoism: 'The Ego is simply *nothing*: as ineffectual and windy an abortion as Philosophy can show'.[19] We might as well discard it—which is what Wittgenstein himself ultimately does in TLP 5.64 ff.

7. Why is the Soul (Subject) Said to be Simple?

Why did Wittgenstein have such a strong conviction that the soul or subject is, if anything, simple? One superficial answer would be that he was misled by the surface-grammar of language. He thought that the pronoun 'I' is a name denoting a simple object, the I. It is an indexical name, but no less genuinely name-like for that. A Humean argument shows that the denotation of this name is not in the world (TLP 5.631 ff.), and therefore it has to be outside of it, or at least on the boundaries of it. And so he arrives at the strange doctrine of the soul as a chimerical, imaginary object. It has disappeared as an object from the world, but its ghost remains as the "gaseous" or "aethereal" denotation of the indexical 'I'.[20]

We think that this answer is too easy. Wittgenstein would not have forgotten his own lessons on the misleadingness of language so quickly. He will certainly have been sensitive to the possibility that 'I' may be no more a genuine name than 'A' in 'A thinks p' and that both names might have to be replaced by long descriptions of complex facts in a completely perspicuous language. After all, he showed such sensitivity in the notes dictated to Moore which we have quoted above (§4). He must have had deeper reasons for his opinion.

We suggest that the answer may be found in the visual metaphor of TLP 5.633 ff. Nothing in the visual field suggests that it is seen by an eye. But its contents are such that we may introduce a fictitious 'geometrical eye' from whose standpoint it appears to be seen.[21] Such an eye does not really exist; it is a fiction similar to the 'point of sight' (also called 'central point' or simply 'eye') which is used in the theory of perspective.

There is a simple relationship between the geometrical eye and the boundaries of its visual field: the latter determine the former, but not conversely. The eye may be seen as the vertex of an imaginary three-dimensional cone containing the things which are seen in the visual field; the surface of this cone is the boundary of the visual field. The boundary of the visual field (including the geometrical eye) does not belong to the field itself; it is unseen and unseeable. It will be clear that each visual cone determines exactly one 'point of sight'. However, the converse does not hold, for a cone starting from a vertex may fan out in various directions.

---

[18] L.J.J. Wittgenstein, *Blue Book*, pp. 73–74, in *The Blue and Brown Books*, Oxford, Blackwell, 1958.

[19] W. James, *The Principles of Psychology*, Vol. 1, New York, Henry Holt, 1890, p. 365. In addition to 'Ego', James also uses the terms 'Transcendental Subject' and 'Self'.

[20] The latter expressions come from the *Blue Book*, op. cit., p. 47.

[21] The expression 'geometrical eye' comes from the *Blue Book*, op. cit., pp. 63–64.

The I (the subject) is a fiction similar to the geometrical eye. It is the imaginary standpoint (the 'centre of the world', *Notebooks* 5.8.16) from which reality is experienced. 'Experience as a whole is the field to which the philosophical I stands as does the geometrical eye to the visual field.'[22] As in the case of the geometrical eye, the 'location' of the I is wholly determined by the 'limits' or 'boundaries' of reality (although the converse does not hold): the I is the imaginary entity which precisely experiences *this* reality. This explains why TLP 5.64 says that 'the I of solipsism shrinks to a point without extension and the reality with which it is correlated remains'.

Just as the surface of the visual cone (including the geometrical eye) does not belong to the visual field which it encompasses, so the limits of reality (including the metaphysical I) do not belong to reality itself. This explains why TLP 5.632 says that 'the subject does not belong to the world; rather, it is a limit of the world'. It would have been more accurate to say that the I does not belong to the world because it is an *element* of the limit of the world (namely, its 'vertex'), but this seems only a minor lapse.[23]

Now we want to suggest that the crucial point which led Wittgenstein to his conception of the simplicity of the soul was his view that such virtual points of view are *necessarily simple (point-like)*. Thus, he seems not to have thought of such expressions as 'from the point of view (perspective) of the present Government ... ', which involve *collective* viewpoints ascribed to *collective* entities. He seems to have thought that all perspectives are perspectives *from one point*. It was this view which made him say in TLP 5.5421 that the I is point-like. (It is merely a virtual thing, but simple, if anything.)

8. Further Evidence for This Interpretation

The picture sketched above is clear and intuitively attractive. However, one might well ask for more arguments and textual evidence. We think that these are not difficult to give.

First, there can be no doubt at all that Wittgenstein regarded the soul or subject as not really existing (virtual); we have already quoted the evidence in §3. This also explains why it does not do anything (§6): it has a purely hypothetical nature.

Secondly, the suggestion that Wittgenstein saw the metaphysical subject as the (fictitious) owner of experience becomes very plausible if we regard some of his later writings as reactions to his earlier view. In the *Philosophical Remarks* in particular, he argued that the perspectival structure of the experienced visual field does *not* provide one with reasons for postulating a subject as its owner. 'The visual space does not have an owner ... The representation of visual space is the representation of an object and contains no suggestion of a subject'.[24] Or as Moore tells us: 'He said that "Just as no eye is involved in seeing, so no Ego is involved in thinking or having toothache", and he quoted, with apparent approval, Lichtenberg's saying "Instead of 'I think' we ought to say 'It thinks'" ("it" being used, as he said, as

[22] M.U. Coyne, 'Eye, "I", and Mine: The Self of Wittgenstein's Tractatus', *Southern Journal of Philosophy* 20, 1982, 313–323, quotation from p. 317.

[23] Coyne (op. cit.) has no difficuly with TLP 5.632 because she regards the eye as the limit of the visual field and the I as the limit of reality. We say that the eye and I are only elements of these limits (namely, their vertices). We prefer our own interpretation because we do not see how one can make sense of Coyne's talk about the 'shapes' of the visual field and reality if these shapes are assumed to be bounded by points. How could a point delimit a shape?

[24] L.J.J. Wittgenstein, *Philosophical Remarks*, ed. by R. Rhees, Oxford, Blackwell, 1964, §71.

"Es" is used in "Es blitzet").'[25] Here he came close to the 'psychology (philosophy) without a psyche' of Hume and Lichtenberg, in which the I is just a 'grammatical fiction'. He was apparently no longer impressed by Weininger's vehement critique of this view.

Thirdly, the suggestion that Wittgenstein regarded the hypothetical centre of experience as simple because he did not think of perspectives from collective viewpoints, hardly needs textual evidence. The very word 'viewpoint' suggests that this is the most natural way of viewing perspectives. Moreover, similar virtual points occur regularly in the fields of descriptive geometry, projective geometry, geometrical optics, etc., which Wittgenstein must have studied as an engineer, and he may well have had them in mind.

In view of the above three points, our interpretation seems to be backed quite well by textual and general considerations.

In sum, we have come to the conclusion that TLP 5.5421 does not arise from a misunderstanding of language. Rather, it is the result of viewing one's view of reality *more geometrico*. The pronoun 'I' refers to a geometrical fiction, the I, which is the virtual centre of all experience. Such a centre can only be point-like, never complex, and that is why it is said in TLP 5.5241 that a 'composite soul would no longer be a soul'.

Thus, TLP 5.5421 expresses an opinion which is interesting in itself and is closely connected with the remarks on the metaphysical subject in the passages on solipsism. It is illuminated by them and illuminates them in turn. One might even say that these remarks can hardly be understood without taking TLP 5.5421 into account. TLP 5.5421 plays a more prominent role than previous commentators have usually thought.

9. Why TLP 5.5421 Should be Rejected

Nevertheless, it would have been better if TLP 5.5421 had never made its way into the *Tractatus*.

In the first place, it mars the unity of the work. The soul or subject is totally different from the other entities that populate the treatise. Nowhere else do we encounter virtual entities virtually representing real facts. A consequence of this is that the doctrines of the *Tractatus* do not apply to it. The picture theory is a good example: it does not explain how virtual representation by point-like virtual entities might work.

In the second place, the idea of a metaphysical soul or subject distinct from the empirical self is unfortunate in itself. As long as its purely fictitious character is stressed, the idea is innocent enough; but then why introduce it? In geometry and physics, virtual entities may often be very convenient. A good example in mechanics is offered by Hertz's 'invisible masses' (which Wittgenstein referred to in the *Notebooks*, 6.12.14). These pseudo-objects (*Scheingegenstände*) enabled him to give a unified account of mechanics. But in philosophy of mind nothing much seems to be gained by introducing the pseudo-object of the metaphysical self.[26] Moreover, postulating such an entity is a dangerous thing to do. Even if its purely

[25] G.E. Moore, 'Wittgenstein's Lectures in 1930–33', in his *Philosophical Papers*, London, Allen and Unwin, 1959, 252–324, quotation from p. 309. Wittgenstein may have got the Lichtenberg quotation from Weininger, op. cit., part II, ch. 7.

[26] A similar critique applies to Dewan's proposal to regard the mind as a virtual governor of the brain of the same type as the virtual governors which are defined over grids of electrical generators. The idea is useful in electrical engineering, but does not have any explanatory value in the philosophy of mind. See E.M. Dewan, 'Consciousness as an Emergent Causal Agent in the Context of Control System Theory', in G.G. Globus, G. Maxwell and I. Savodnik, eds., *Consciousness and the Brain: A Scientific and Philosophical Inquiry*, New York, Plenum Press, 1976, 181–198.

hypothetical character is stressed, it is all too easily imaginable that someone might remark that *his* mind does not seem imaginary to *him*. And thus it is all too easy to fall prey to illegitimate reifications of the same sort as Schopenhauer's reification of Kant's transcendental unity of apperception. As long as the idea has no obvious value, it is best to avoid it.

In the third place, the introduction of a soul or subject distinct from the empirical self is completely unnecessary within the framework of the *Tractatus*. Everything which Wittgenstein says about 'my language', 'my world', and the 'truth of solipsism' may just as well, nay, better be said without bringing a separate subject into play. We will devote the remaining part of this paper to a demonstration of how this may be done.

10. The Empirical Self as the Centre of Existence

Let us suppose that the pronoun 'I' does not refer to an imaginary point, but to a collection of facts; it is an indexical quasi-name of the collection of facts which constitute me. (It is not a genuine name because collections of facts cannot be named.) Thus, we return to Wittgenstein's remark in the notes dictated to Moore that 'it is just as impossible that *I* should be a simple as that 'p' should be'. Which consequences does this have?

First, we may easily make sense of the remarks about 'my language' and 'my world' in TLP 5.6, as Hintikka pointed out a long time ago.[27] 'My language' is just another expression for my empirical self, the collection of thoughts I have or may have. (Remember that thoughts are similar to sentences.) Or alternatively, 'my language' may be regarded as the collection of all sentences which *express* all the thoughts I may have. 'My world' is the collection of all (possible) facts which are represented by the thoughts I may have. At any given time, I may entertain any given subcollection of my collection of possible thoughts, which explains why 'no portion of our experience is *a priori*' (TLP 5.634). But I cannot transgress the boundaries of what I *may* think (I cannot have a thought which I cannot have), and therefore 'the limits of my language are the limits of my world' (TLP 5.6).[28] As *my* language is *the* language for me, 'the only language which I understand' (TLP 5.62), *my* world is *the* world for me, the only world which I experience (TLP 5.641).

Secondly, we may easily transpose the remarks we made above about the 'perspectival' character of the I to this new setting. The only difference is that the 'viewpoint' from which the world is experienced is no longer simple, but complex. It is a *plane* or *region* rather than a point.

It is helpful to consider the case of the visual field again. It may be possible to define a point-like 'geometrical eye' as the point from which the visual field is seen, although this should not be granted too quickly—how should one, for example, account for stereopsis, seeing in depth, if one limits oneself to one point-like geometrical eye? But even if it were possible to define such an abstract eye, it is more realistic to consider the *two* eyes taken together, which are both complex in themselves, as the standpoint (region) from which the visual field is seen. The visual field is seen from the *collective* perspective of a pair of eyes. Wittgenstein seems not to have thought of this possibility, which is, however, completely clear in itself.

---

[27] Compare Hintikka, op. cit. (1958).

[28] 'Ich kann mir nichts ausser meinem Denken denken; denn dadurch, dass ich es denke, wird es ja mein Denken, und fällt unter die unvermeidliche Gesetze desselben.' ('I can't think anything which goes beyond my thinking; for the very fact that I am thinking it turns it into my thinking, and makes it fall under the inevitable laws of thinking.') Wittgenstein? No, J.G. Fichte, *Die Bestimmung des Menschen*, Berlin, Voss, 1800, p. 157.

According to us, the case of the 'I', 'soul' or 'self' is analogous. These entities, too, may perhaps be defined as simple virtual points of perspective, but it is more realistic to regard them as collections of various real elements (thoughts), each of them complex in itself—that is, to identify them with the empirical self. We may grant that 'the I enters into philosophy because "the world is my world"', as Wittgenstein writes in TLP 5.641. However, that does not tell us anything about the nature of the I. In particular, it does not tell us that it is simple. The I may be as complex as the eyes jointly are.

Thirdly, it is easy to do justice to the remarks about the non-encounterability of the self. The empirical self is a collection of facts, all of them belonging to the world. It is, in principle, possible that this self may encounter itself in the world, in the sense that it could contain a picture of every fact of which it is composed. It even does not have to be *infinite* to be capable of this feat, as Hacker wrongly supposes.[29] For we could envisage *circles of representation* in the sense that a picture of a picture of a fact (or a picture to the n-th degree of a fact) might be identical with this fact itself. In this way, there could be a picture of every picture, while the total collection of pictures would yet remain finite.

However, even if the empirical self contained a picture of every picture of which it consists, this would not guarantee that the empirical self is able to *recognize* the facts of which it is composed as being elements of itself. For as we saw above in our remarks about the impossibility of formulating psychophysical laws, it is impossible to represent, say or think what facts *represent*. Even if the self saw itself completely (for example, if it were identical with the brain and observed itself in action with the celebrated 'autocerebroscope'), it would not be able to recognize itself *as* itself, because it could never see, say or think that the facts which it observes represent exactly the same facts which it is representing itself. Therefore the empirical self is in a sense 'invisible' or 'unknowable' to itself, even if it saw and knew itself completely.

Fourthly, our account has the (minor) advantage over the view of the self as the purely imaginary centre of all experience that it does not exclude the possibility that two minds might have exactly the same contents. Two different collections of pictures might represent exactly the same facts, whereas minds which are completely determined by their experiences can only be different if they have at least one different experience.

11. Living with the Empirical Self

In short, all the things which Wittgenstein says about the metaphysical self may just as well be said about the empirical self. It seems wise to use Occam's razor and excise the metaphysical self from the *Tractatus*.

This does not imply that everything which Wittgenstein says about the empirical self is unobjectionable. For example, the idea that thoughts are of a linguistic nature seems to be an unfortunate relapse to the mediaeval doctrine of a 'language of thought' or its short revival in German Idealism (Humboldt, Schelling, Hegel, Herder). Hegel[30] boldly asserted that 'we think in nouns', and it is not unfair to suggest that Wittgenstein came rather close to this view, which was repudiated by almost all major philosophers who came before and after him.[31]

[29] Hacker, op. cit., pp. 77–78.

[30] Quoted without reference by J. Hadamard, *The Psychology of Invention in the Mathematical Field*, 2nd ed., Princeton, N.J., Princeton U.P., 1949, p. 68.

[31] I am ignoring the philosophers who have proclaimed themselves to be the philosophers of modern cognitive science (Fodor, Pylyshyn and the like). First, cognitive scientists themselves are usually bewildered by the claims these philosophers make on their behalf, and secondly, this kind of philosophy is rapidly dying out anyway (cf. the last footnote below).

Wittgenstein seems to have been led to it because he only had a representational theory of sentences. (His picture-theory of sentences might better be called a sentence-theory of pictures.) He derived this theory from Hertz's theory of 'dynamical models'; the only modification which he made was to give a linguistic (semantical) twist to it.[32] With hindsight, this linguistic twist seems regrettable. Hertz's notion of models (including his account of dynamical *mental* models) is still applicable to modern psychological models of mental representation such as the 'Boltzmann machine', while a theory of mental sentences is hopelessly out of date here.[33] But whatever the defects of the theory of the empirical self, these may be amended, for this notion seems to have a good deal of substance. With the metaphysical self it is different, and therefore we had better forget about TLP 5.5421 and the passages related to it. The simple metaphysical self is dead, but the complex empirical self is alive and well: *l'âme est morte, vive l'âme*.

[32]A good account of Hertz's influence on the picture-theory may be found in J. Griffin, *Wittgenstein's Logical Atomism*, Oxford, Oxford U.P., 1964.

[33]The 'Boltzmann machine' is a device composed of simple elements analogous to neurons whose collective behaviour is described by the laws of statistical mechanics. It is able to make 'dynamical internal models' of the statistical structure of its environment which exactly conform to the definition H. Hertz gave of such models in his *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*, Leipzig, Barth, 1894. See, e.g., D.H. Ackley, G.E. Hinton and T.J. Sejnowski, 'A Learning Algorithm for Boltzmann machines', *Cognitive Science* 9, 1985, 147–169, repr. in J.A. Anderson and E. Rosenfeld, eds., *Neurocomputing: Foundations of Research*, Cambridge, Mass., MIT Press, 1988, 638–649. This anthology contains many more examples of non-sentential psychological models of a 'Hertzian' kind.

CHAPTER 5

Multiply Modal Extensions of da Costa's $\mathscr{C}_n$, $1 \leq n \leq \omega$, Logical Relativism, and the Imaginary

Abstract

How should *our* logic express what *other* logics deem necessary? How should we give a rational account of forms of rationality which are different from ours? The present paper answers these questions. It shows how to enrich logical systems with operators which describe what is necessary, rational and imaginary according to other systems. Although only da Costa's paraconsistent calculi are treated in detail, the construction is generally applicable. As a result the thesis of logical relativism—people from different cultures may live in different cognizable worlds—may henceforth be discussed in terms of modal logic and possible world semantics.

First published in *The Journal of Non-Classical Logic* **5**, nr. 2 (1989), 7–22. ISSN 0102–3411.

Or, s'il y a plusieurs mondes, comme ... presque toute la philosophie a pensé, que sçavons nous si les principes et les règles de cettuy touchent pareillement les autres? Ils ont à l'avanture autre visage et autre police.

M. E. de Montaigne, *Apologie de Raimond Sebond*
(de Montaigne 1580)

1. Introduction

When one does not restrict one's attention to just one logic, but bears in mind that there is a plurality of logics around (intuitionistic logic, multi-valued logics, paraconsistent logics, etc.), it seems obvious that notions such as necessity, possibility and rationality are not *absolute*, but *relative* to the particular logical system under consideration. Yet, the logic-relative nature of these notions is not generally recognized, and logical systems which take it into account do not seem to have been constructed up to now. In this paper, we will try to fill this lacuna. We have selected da Costa's well-known series of paraconsistent logics $\mathscr{C}_n$, $0 \leq n \leq \omega$, to make a first study of logic-relativized notions of necessity, possibility and rationality, and to indicate some philosophical areas (Vasil'ev's "imaginary logic", the logic of belief, and Lévy-Bruhl's "logical relativism") which are illuminated by the relativistic and pluralistic analysis of these notions.

The general considerations motivating our enterprise are as follows.

In ordinary single-operator modal logic, the sentence "it is logically necessary that A" is given the following truth-condition.

> "It is logically necessary that A" is true at world w (in model $\mathfrak{M}$) iff
> "A" is true at all logically possible worlds accessible from w (in $\mathfrak{M}$).

But now suppose we consider several logical systems at once, say the da Costa series $\mathscr{C}_n$, $0 \leq n \leq \omega$. In this case, the above truth-condition can no longer be used. For to *which* one of the many systems does "logically necessary" refer now? Any

system from the da Costa series may be meant. And according to *which* one of the various logics are the worlds referred to logically possible? Again, any system from the da Costa series may be meant.

To remove this ambiguity, one has to specify which logic one has in mind in both cases. This is what we will do in the following. Each of the systems we present has denumerably many modal operators $\Box_n$, $0 \leq n \leq \omega$, corresponding to $\mathscr{C}_n$, $0 \leq n \leq \omega$, respectively. The subscripts of the operators indicate the logical systems to which they are relativized; $\Box_n A$ may be read as "it is $\mathscr{C}_n$-necessary that A", or as "according to $\mathscr{C}_n$, it is necessary that A". Semantically, we introduce various sets of worlds W_n, $0 \leq n \leq \omega$, likewise corresponding to $\mathscr{C}_n$, $0 \leq n \leq \omega$, respectively; W_n is the set of worlds which are possible according to logic $\mathscr{C}_n$, or stated otherwise, it is the set of worlds in which $\mathscr{C}_n$ is valid. Having made these distinctions, we are able to give the following disambiguated truth-condition:

> "It is $\mathscr{C}_n$-necessary that A" is true at w (in $\mathfrak{M}$) iff "A" is true at all $\mathscr{C}_n$-possible worlds accessible from w (in $\mathfrak{M}$).

In this way, we explicitly recognize the fact that there is more than one logic around. The truth-condition has the effect that $\mathscr{C}_n$-axioms are $\mathscr{C}_n$-necessary but need not be $\mathscr{C}_m$-necessary if $m \neq n$, which is in accord with our intuitions on logic-relative necessity. Since $\mathscr{C}_n \subseteq \mathscr{C}_m$ if $n \geq m$, we will make the plausible assumption that $W_m \subseteq W_n$ if $m \leq n$ for all m, $0 \leq m \leq \omega$, n, $0 \leq n \leq \omega$ (the stronger the logic, the less it will count as possible). Thus, $\mathscr{C}_n$-axioms are $\mathscr{C}_m$-necessary if $n \geq m$.

The series of the smallest logical systems arising from this semantic condition will be denoted as $\mathscr{C}_n K$, $0 \leq n \leq \omega$. We will study this series and some related ones in §2 below. The applications of the systems will be discussed in §3.

First, however, a preliminary remark. As we have said, it is our goal to apply our systems to the analysis of Vasil'ev's views. Now according to Arruda (1977), an imaginary logic in the sense of Vasil'ev must be adequate to handle at least two sorts of negation, viz., classical (strong) negation, and a weaker negation for which the law of contradiction is not valid. The former type of negation may be defined in $\mathscr{C}_n$, $0 \leq n < \omega$; however, it may *not* be defined in $\mathscr{C}_\omega$ (Arruda mistakenly claims the contrary). In order to remove this difficulty, we have enriched the language with a primitive symbol for strong negation, notated as $\approx$. This has no effect on $\mathscr{C}_n$, $0 \leq n < \omega$, except that conjunction and disjunction may now be defined in terms of $\supset$ and $\approx$. However, it makes our $\mathscr{C}_\omega$ stronger than da Costa's $\mathscr{C}_\omega$. For example, Peirce's law $(((A \supset B) \supset A) \supset A)$ may now be proven in $\mathscr{C}_\omega$ (in the same way as in classical logic), which is impossible in da Costa's original $\mathscr{C}_\omega$.[1]

2. The series $\mathscr{C}_n K$, $0 \leq n \leq \omega$

2.1. The language. Let AT be a denumerable set. The set of formulas WFF is the smallest set such that AT $\subseteq$ WFF and if A, $B \in$ WFF then $\sim A$, $\approx A$, $A \supset B$, $\Box_n A \in$ WFF, for each n, $0 \leq n \leq \omega$.

Definitions: $A \,\&\, B$, $A \vee B$ and $A \equiv B$ are defined as usual. A^o is short for $\sim(A \,\&\, \sim A)$. A^n is short for $A\overbrace{oo\cdots o}^{n}$, i.e., for A followed n times by ^{o}. $A^{(n)}$ abbreviates $\overbrace{A^o \,\&\, A^{oo} \,\&\, \ldots \,\&\, A^n}^{n}$. Finally, $\sim^{(n)} A$ stands for $\sim A \,\&\, A^{(n)}$, and $\Diamond_n A$ for $\approx\Box_n\approx A$.

[1] See da Costa (1974) and Loparić (1977) on the indefinability of classical negation and the unprovability of Peirce's theorem in $\mathscr{C}_\omega$.

2.2. Axiomatization of $\mathscr{C}_nK$, $0 \leq n \leq \omega$. Each $\mathscr{C}_nK$, $0 \leq n \leq \omega$, is axiomatized by adding the following axiom schemes C1–C7 and rule scheme C8 to classical propositional logic (formulated with classical negation, $\approx$):[2]

C1 $A \vee \sim A$.
C2 $\sim\sim A \supset A$.
C3 $\sim A \supset (A \supset B)$, provided that $n = 0$.
C4 $B^{(n)} \supset ((A \supset B) \supset ((A \supset \sim B) \supset \sim A))$, provided that $n \neq \omega$.
C5 $(A^{(n)} \,\&\, B^{(n)}) \supset ((A \supset B)^{(n)} \,\&\, (A \& B)^{(n)} \,\&\, (A \vee B)^{(n)})$, provided that $n \neq \omega$.
C6 $\Box_m A \supset \Box_k A$ if $m \geq k$, for each k, $0 \leq k \leq \omega$, m, $0 \leq m \leq \omega$.
C7 $\Box_m(A \supset B) \supset (\Box_m A \supset \Box_m B)$, for each m, $0 \leq m \leq \omega$.
C8 $\vdash_m A \Rightarrow \vdash_n \Box_m A$, for each m, $0 \leq m \leq \omega$.

Here $\vdash_n A$ is an abbreviation for $\emptyset \vdash_n A$, where $S \vdash_n A$ (for $S \subseteq$ WFF) in turn means that A is derivable from S by means of the axioms and rules of $\mathscr{C}_nK$. Derivability is defined in the usual way.

The distinction between the constant n (of $\mathscr{C}_nK$) and the variables k and m should be especially noted in the above. Furthermore, notice the special form of the denumerably many rules of necessitation (C8). In conjunction with C7, these rules have the consequence that for each sequence Σ of modal operators (including the null-sequence) and each k and m, $\{A : \vdash_k \Sigma\Box_m A\}$ is a $\mathscr{C}_mK$-theory. (A $\mathscr{C}_mK$-theory is a set of sentences containing $\mathscr{C}_mK$ and closed under modus ponens.) This justifies our reading of $\Box_m A$ as "it is $\mathscr{C}_mK$-necessary that A". Following the common doxastic interpretation of modal logic as the logic of rational belief, it allows us to read $\Box_m A$ as "it is $\mathscr{C}_mK$-rational to believe that A" or "a perfect $\mathscr{C}_mK$-logician (adherent of $\mathscr{C}_mK$) believes that A". (See §3.2 below.)

It may be observed that, for any m, $0 \leq m \leq \omega$, the modal fragments $\{A : \vdash_n \Box_m A\}$ of all $\mathscr{C}_nK$ are exactly the same. Furthermore, the $\{\approx, \supset, \Box_m\}$ fragment of each $\mathscr{C}_nK$ is exactly the same as the classical modal system K.

Finally, notice that $\mathscr{C}_nK \subseteq \mathscr{C}_mK$ if $n \geq m$. The strongest logic is $\mathscr{C}_0K$, while $\mathscr{C}_\omega K$ is the weakest one. In $\mathscr{C}_nK$, $0 \leq n < \omega$, strong negation may be defined as $\approx A = \sim^{(n)} A$. In $\mathscr{C}_\omega K$ it cannot be but primitive.

2.3. Semantics. A Kripke-style "possible-worlds" model for $\mathscr{C}_nK$ is a structure

$$\mathfrak{M} = \langle \langle W_m \rangle_{0 \leq m \leq \omega}, W_n, w_0, R, V \rangle,$$

where:

- each W_m, $0 \leq m \leq \omega$, is a set (of $\mathscr{C}_mK$-possible worlds);
- $W_k \subseteq W_m$ if $k \leq m$;
- W_n is the distinguished set of "really possible" (i.e., $\mathscr{C}_nK$-possible) worlds;
- w_0 (the actual world) is a member of W_n;
- $R \subseteq W \times W$, where $W = \bigcup\{W_m : 0 \leq m \leq \omega\} = W_\omega$;
- V : WFF $\times\, W \mapsto \{0, 1\}$ is a function satisfying the following conditions:[3]
  1. $V(\approx A, w) = 1$ iff $V(A, w) = 0$;
  2. $V((A \supset B), w) = 1$ iff $V(A, w) = 0$ or $V(B, w) = 1$;
  3. if $V(\sim A, w) = 0$ then $V(A, w) = 1$;
  4. if $V(\sim\sim A, w) = 1$ then $V(A, w) = 1$;
  5. if $V(\sim$A$, w) = 1$ then $V(A, w) = 0$, provided that $w \in W_0$;
  6. For all m, $0 \leq m < \omega$: if $V(B^{(m)}, w) = V((A \supset B), w) = V((A \supset \sim B), w) = 1$ then $V(A, w) = 0$, provided that $w \in W_m$;

[2] Cf. da Costa (1974).
[3] Cf. da Costa & Alves (1977) for the non-modal conditions.

7. For all m, $0 \leq m < \omega$: if $V(A^{(m)} \,\&\, B^{(m)}, w) = 1$ then $V(((A \supset B)^{(m)} \,\&\, (A \,\&\, B)^{(m)} \,\&\, (A \vee B)^{(m)}), w) = 1$, provided that $w \in W_m$;
8. $V(\Box_m A, w) = 1$ iff $V(A, v) = 1$ for all $v \in W_m$ such that wRv.

For $S \subseteq$ WFF, $S \models_n A$ means: for all $\mathscr{C}_n K$-models in the above sense: if $V(B, w_0) = 1$ for all $B \in S$, then $V(A, w_0) = 1$. For all m, $0 \leq m \leq \omega$, we say that A is *valid* on W_m (in a particular model) iff $V(A, w) = 1$ for all $w \in W_m$.

2.4. Completeness.

Completeness theorem: $S \vdash_n A$ iff $S \models_n A$.

Proof. From left to right: trivial. From right to left: a canonical model may be constructed in the usual way. Let $w_0 \in W_n$ be a $\mathscr{C}_n K$-maximal nontrivial extension of S, let W_n be the set of $\mathscr{C}_m K$-maximal nontrivial sets of the language (this is is the only unusual part of the construction), and let $V(A, w) = 1$ iff $A \in w$. It is not difficult to show that the canonical model satisfies all conditions from §2.3 and that, for any A which is not derivable from S, $V(B, w_0) = 1$ for all $B \in S$ while $V(A, w_0) = 0$. This completes the proof.[4]

2.5. Some correspondence results.

Seriality: If we add the axiom $\Diamond_\omega(A \vee {\sim}A)$ to each $\mathscr{C}_n K$, $0 \leq n \leq \omega$, we obtain a series $\mathscr{C}_n K$, $0 \leq n \leq \omega$, of systems which are complete with respect to the class of serial models (i.e., models in which $\forall w \in W \exists v \in W\, wRv$).

Reflexivity: corresponds to adding $\Box_n A \supset A$ to each $\mathscr{C}_n K$. (Notice that $\Box_\omega A \supset A$ would be too weak and $\Box_0 A \supset A$ too strong.)

Transitivity: corresponds to adding $\Box_m A \supset \Box_\omega \Box_m A$ (for all m, $0 \leq m \leq \omega$) to each $\mathscr{C}_n K$.

Symmetry: corresponds to adding $A \supset \Box_\omega \Diamond_n A$ to each $\mathscr{C}_n K$.

2.6. The logic of the imaginary. In conformity with Vasil'ev's use of the term (see §3.1 below), we say that a world $w \in W$ is *imaginary* from the point of view of $\mathscr{C}_n K$ if $w \notin W_n$. So w is imaginary according to $\mathscr{C}_n K$ if w is possible according to *some* logic, but impossible according to $\mathscr{C}_n K$ itself. Imaginary worlds are the worlds "lying beyond the horizon of the logical space" of $\mathscr{C}_n K$.

Imaginariness may be expressed in the object-language by introducing a new modal operator I_n. $I_n A$ may be read as "according to $\mathscr{C}_n K$ it is imaginary that A" or as "it is $\mathscr{C}_n K$-impossible but ($\mathscr{C}_\omega$-)imaginable that A". Thus:

$$I_n A \stackrel{\text{def}}{=} {\approx}\Diamond_n A \,\&\, \Diamond_\omega A.$$

The stronger the logic, the more will be imaginary according to it. Classical logic gives the verdict "imaginary" most easily, whereas nothing is imaginary according to $\mathscr{C}_\omega K$. (So imaginariness is as logic-relative as possibility and necessity are.)

It may be of some interest to investigate what the logic of the imaginary is like all by itself. Fortunately, the answer is easy, for the case is similar to that of "purely logical (as contrasted to physical) possibility", which has been studied by Bacon (1981).

Bringing Bacon's axiomatization into line with our notation, we may axiomatize the notion of "it is imaginary according to $\mathscr{C}_n K$" by adding the following axiom schemes I1-I4 and rule schemes I5-I7 to $\mathscr{C}_\omega K$ (for all m, n, $0 \leq m \leq \omega$, $0 \leq n \leq \omega$):[5]

[4] One may compare the completeness proofs of classical multiply modal logics which have been given by Fitting (1969) and Rennie (1970). On modal logic, see also Chellas (1980).

[5] Bacon's operator N of purely physical necessity corresponds to our $I_n{\approx}A$. Bacon's relation of physical accessibility S corresponds to our $R \restriction W_n$, while his relation of logical accessibility R corresponds to our R. The main differences between Bacon's systems and ours are threefold. First, we have replaced Bacon's axiom N3 by his derived rule T3. (Both are easily seen to be

I1 $\approx I_\omega A$.
I2 $I_n A \supset I_m A$, if $n \geq m$.
I3 $(I_n A \,\&\, I_n B) \supset I_n(A \vee B)$.
I4 $(\approx I_n(A \,\&\, B) \,\&\, I_n A \,\&\, I_n C) \supset I_n \approx (C \supset (A \,\&\, B))$.
I5 $\vdash_\omega A \equiv B \Rightarrow \vdash_\omega I_n(A \equiv B)$.
I6 $\vdash_\omega A \Rightarrow \vdash_\omega \approx I_n \approx A$.
I7 $\vdash_\omega A \supset B \Rightarrow \vdash_\omega I_n(A \,\&\, C) \supset (I_n B \supset I_n A)$.

Some noteworthy theorems and derived rules are (for any n, $0 \leq n \leq \omega$):

T1 $I_n(A \vee B) \supset (I_n A \vee I_n B)$.
T2 $(I_n(A \,\&\, B) \,\&\, I_n \approx A) \supset I_n B$.
T3 $\vdash_\omega A \Rightarrow \vdash_\omega I_n(A \,\&\, B) \supset I_n B$.
T4 $\vdash_\omega A \supset B \Rightarrow \vdash_\omega I_n A \supset I_n(A \,\&\, B)$.

If $R \restriction W_n$ is serial, we have $\vdash_n A \Rightarrow \vdash_\omega \approx I_n A$. If $R \restriction W_n$ is reflexive we have $\vdash_\omega A \supset \approx I_n A$.

3. Applications

3.1. Vasil'ev's imaginary logics and worlds. The Russian physician Vasil'ev has become famous as one of the first forerunners of paraconsistent logic.[6] His viewpoints are clarified to a great extent by our multiply modal approach.

Inspired by the existence of various imaginary (non-Euclidean) geometries, Vasil'ev envisaged the possibility of constructing a great multitude of "imaginary" logics. These logics would enable us to study a large class of "imaginary worlds" which are impossible to classical logic, but nevertheless quite well imaginable by our minds. According to Vasil'ev, Aristotelian logic is an instrument of knowledge for only a limited class of worlds, the "classical" worlds, in which, for example, the law of non-contradiction holds. However, beyond the classical worlds there is a whole range of imaginary worlds, which obey the laws of various imaginary logics. Vasil'ev did not deny the truth of classical logic: he assumed that experience has taught us that the real world we inhabit is classical. But we can imagine that it could have been otherwise. The truth of classical logic is only an empirical matter; "logic is as empirical as geometry".[7] The idea that classical logic is universally valid is an illusion created by our particular place in logical space and a lack of imagination to look beyond the classical horizon.

Vasil'ev did not give a formal development of his views. However, an attempt to do this has been made by Arruda (1977). Arruda's formalizations indeed capture some of Vasil'ev's basic insights. However, her proposals seem to have two shortcomings. First, they do not capture Vasil'ev's central idea of a *plurality* of imaginary logics, "existing", so to say, side by side; she just presented several isolated systems. And second, she did not clarify the idea of an "imaginary world" at all, let alone the idea of a *plurality* of types of imaginary worlds, each of them possible according to some different imaginary logic. Indeed, the term "imaginary world" did not even occur in her formal exposition.

Our systems do not have these shortcomings. The introduction of several modal operators, each of them corresponding to a different logic from the da Costa series,

interderivable.) Second, we have dropped the condition that R is reflexive (corresponding to Bacon's axiom N1). Third, I1 and I2 have no counterparts in Bacon's system; they are immediate consequences of our definition of I_n and of axiom C6.

[6] Vasil'ev (1912). We follow the exposition of Vasil'ev's views given by Arruda (1984). See also Żarnecka-Biały (1985) and Puga & da Costa (1988).

[7] This famous assertion of Putnam (1968) could have come straightly from Vasil'ev's writings. Putnam (1968), Rescher & Brandom (1980), and various other modern authors not only share Vasil'ev's view that classical logic could be empirically false, they even claim that it has in fact been shown to be false (by quantum mechanics).

enables us to capture the idea of a plurality of logics existing side by side. As we have seen, this has even allowed us to express the notion of "imaginary according to logic $\mathscr{C}_n$" within the language. Likewise, the multitude of types of worlds in the semantics, each type corresponding to one of the da Costa logics, seems to be a fairly direct expression of Vasil'ev's idea of a multitude of worlds described by various logics. Vasil'ev's idea that the actual world is classical may be captured by the condition that w_0 is a member of W_0. But even if we stipulated this we should not overlook the other worlds, and consider $\mathscr{C}_0K$, rather than the classical single-operator modal system K, as the logic of the imaginable or the possible (in a wide sense).

3.2. The logic of belief. Apart from clarifying Vasil'ev's ideas, our systems are also interesting from the point of view of doxastic logic.[8] Classical doxastic logic (which simply is modal logic with $\Box A$ read as "the agent believes that A") has often wrestled with the problem of how to give an account of inconsistent beliefs which does not imply that *everything* is believed. This is a problem for classical doxastic logic, because it has the (doxastic variant of the) theorem $\Box(A \,\&\, {\sim}A) \supset \Box B$. The usual solution is to distinguish between "implicit" inconsistencies of the form $\Box A \,\&\, \Box{\sim}A$ and "explicit" inconsistencies of the form $\Box(A \,\&\, {\sim}A)$ and to deny that the former imply the latter. Thus, the "belief-set" (set of believed sentences) of the agent is generally not closed under conjunction, and it may contain at least one type of inconsistencies (implicit inconsistencies) without collapsing into the whole language.[9] Now this method of fragmentation or compartmentalisation may certainly be applicable in a number of instances, although it may sometimes have the drawback that it is extremely sensitive to the way the belief-set is broken down into internally consistent subsets.[10] But our account is simpler, for we do not have to split up the agent. Even explicit inconsistencies are harmless on our account, since for all $m > 0$, $\Box_m(A \,\&\, {\sim}A) \supset \Box_m B$ is invalid (in all $\mathscr{C}_nK$, $0 \leq n \leq \omega$).

Notice that our approach does *not* involve abandoning classical logic. We may retain $\mathscr{C}_0$ as a valid description of the actual world, but we must resist the temptation to regard the belief-set of an agent as necessarily being a theory of the same logic. (See §2.2 above for the meaning of "theory of a logic".) The belief-set need not be classical; the agent may adhere to another logic than we (the belief-ascribers) do. Just as the ascription of beliefs is, according to Clark (1976),

> mainly a matter of keeping the references and concepts of those of us who are scribes, recording the occurrences of psychical happenings, distinct from those of the agents to whom we ascribe mental events,—

so the ascription of beliefs is a matter of keeping the agents' and our (the scribes') *logics* distinct as well. We should not be so narrow-minded (or conceited) as to foist our own logic on everyone.

Are agents having different logics than ours *ipso facto* irrational? We do not think so. Rationality is as logic-relative as necessity. Whether a particular system of beliefs is rational or irrational just depends on the logic by which this system is judged, just as a sentence may be necessary according to one logic and contingent according to another. (For example, $\approx(A \,\&\, {\sim}^{(n)}A)$ is necessary according to $\mathscr{C}_nK$, but contingent according to $\mathscr{C}_mK$ if $m > n$.) Let us say that a belief-set is *rational* iff it is a theory of *some* logic; it is *rational according to logic* $\mathscr{C}_nK$ iff it is a *theory of* $\mathscr{C}_nK$. So a belief-set containing $A \,\&\, {\sim}^{(n)}A$, for example, cannot be rational

[8] On classical doxastic logic see, e.g., Hintikka (1962), Lenzen (1978) and Lenzen (1980).

[9] See, e.g., Lewis (1982), Lewis (1986); postscript to Lewis (1978) in Lewis (1983), Rescher & Brandom (1980), Stalnaker (1984). The minimal deontic logic D of Chellas (1980) is a good example of a (deontic variant of a) doxastic logic that may be obtained in this way.

[10] This criticism has been expressed by Belnap, Jr. (1977).

according to $\mathscr{C}_n K$ while it may be rational according to $\mathscr{C}_m K$, $m > n$. (Whether it actually is rational according to $\mathscr{C}_m K$ depends, of course, not only on this sentence itself, but also on the rest of the belief-set.)

Thus, if our systems are given a doxastic interpretation, they represent various types of rational belief. For each logic $\mathscr{C}_n$ of the da Costa series there is a corresponding type of believer, whose beliefs are rational with respect to just that logic. The belief-sets of these various types of believers are semantically modelled by different types of worlds. For each type of believer there is a different class of "doxastic alternatives" (as they are commonly called), worlds the believer "mentally lives in"; these worlds may be different from the type of worlds we imagine ourselves to be living in and they may accordingly be merely "imaginary" to us.

The range of forms of rationality we admit is, of course, rather limited: we have not included intuitionists, followers of Łukasiewicz's three-valued logic, etc. But our approach is at least not as parochial as that of the classical doxastic logicians, who see classical rationality as the only form of rationality, by which everyone is to be judged, even if the objects of the judgment themselves explicitly disavow the standards by which the judgment is made (as the intuitionists do).[11]

3.3. Logical relativism. Now, this recognition of a plurality of types of rational belief brings us close to the thesis of "logical relativism", which has received a tremendous amount of discussion within anthropology during the last 75 years. And indeed, we think our account manages to throw some long-needed light on this notoriously unclear thesis.

Logical relativists typically make the following claims.

1. "People of different cultures may have specifically different logics (for example, [there may be] a peculiarly Chinese logic distinct from Western logics)" (Lévy-Bruhl 1949). People of different cultures who follow different logics than ours should not be considered irrational: their "beliefs are on our standards irrational, but on other [...] standards they are about 'real' phenomena and 'logical"' (Lukes 1967). "The standards of rationality in different societies do not always coincide" (Winch 1964).
2. In an "ontological" formulation, logical relativism is the claim "that people of other cultures live in other worlds, so that what is rational in their world may well appear irrational in ours" (Sperber 1982). Sperber elaborates: "The relativist slogan, that people of different cultures live in different worlds, would be nonsense if understood as literally referring to physical worlds. If understood as referring to cognized worlds, it would overstate a very trivial point. [...] If, however, the worlds referred to are *cognizable worlds*, then the claim need be neither empty nor absurd." (Ibid.)
3. To these claims, it was, originally, often added that "the primitive mind is not constrained above else, as ours is, to avoid contradictions. What to our eyes is impossible or absurd, it sometimes will admit without seeing any difficulty." (Lévy-Bruhl 1925). "It does not bind itself down, as our thought does, to avoiding contradiction" (Lévy-Bruhl 1910).

[11] The recognition of a variety of types of rational belief makes our systems different from the non-classical doxastic logics to be found in Routley & Routley (1975) and da Costa & French (n.d.). Similarly, it makes them different from modern "situation semantics" and "discourse representation theory", which are nowadays often put forward as successors to the modal, possible-worlds approach to doxastic logic. Our critique of single-operator "modal" doxastic logic applies with the same force to the latter approaches: even if we consider much weaker systems than classical logic—as these modern analyses do,—we should distinguish between the scribes' logics and the logics of the agents to whom the beliefs are ascribed. (One should distinguish between situations scribes and agents think they live in, or between discourses of scribes and agents, respectively.)

It is of course an empirical matter to decide whether the thesis of logical relativism is true. Current opinion no longer seems to favor it.[12] However, this may at least partially be due to its unclarity: the thesis of logical relativism hinges on such notions as "logic", "rationality", "(cognizable) world", "consistency" and "contradiction", but anthropologists have always ignored the clarification of these notions in logic, while logicians showed no interest in clarifying the anthropological debates either. Therefore the thesis may have been abandoned too early. The merits and defects of a hypothesis cannot be properly judged until the hypothesis is sufficiently understood.

We think our "doxastic Vasil'evean" systems precisely enable us to clarify the three claims of logical relativism. First, we have seen how the claim that different people have "different logics" and "different standards of rationality" may be understood: their belief-sets are theories of different logics. Second, we have seen that theories of different logics describe different types of worlds. People having different logics do not have the same "doxastic alternatives" and may therefore be said to "live in" different kinds of worlds (mentally). Sperber's "cognizable worlds" are just the same as our "imaginable worlds". And finally, we have seen that some of the belief-sets we have considered (viz., the theories of the systems $\mathscr{C}_n K$, $n > 0$) are tolerant of contradictions, which provides a formal underpinning of the third claim. Therefore we think our analysis goes a long way in providing a clear and adequate explanatory model of the central traits of logical relativism.

4. Conclusion

This completes our exposition of multiply modal logics based on da Costa's $\mathscr{C}_n$, $0 \leq n \leq \omega$. We have not indicated all areas to which our systems might be applied. For example, the analysis of "truth in fiction" bears a close resemblance to doxastic logic (Lewis 1978), and our approach may be used to give an account of truth in fictional or non-fictional texts which do not subscribe to the canons of classical logic, but follow, describe or proclaim different logics. Think, for example, of tales written in accordance with paraconsistent logic, or simply of intuitionistic textbooks: it would be unfair, it would not be in accordance with the spirit of the texts, and it may even be seen as a sign of misunderstanding them, to judge such texts by classical logic. Deontic logic (which is also close to modal logic) would be another area of application. Various cultures might not only be pluralistic in their ethical norms (e.g., in the way described by Menger (1974)), but also in the logical standards by which they judge adherence to these norms.

Without doubt, there are more applications to be found. However, we hope the above may suffice to demonstrate the usefulness of the pluralistic, relativistic approach to modal logic. As Lewis (1986) has stated, the realm of possible worlds is "a philosophers' paradise", but he went on to argue that we do not need impossible worlds to carry out any interesting philosophical tasks. We hope to have shown that impossible worlds are as useful as possible worlds, and, moreover, that we do not need just one type, but lots and lots of varieties of them.[13]

Note added in print (to the original article)

As has already been pointed out in the text, the restriction to the da Costa series is inessential: our account may be extended to other systems of logic. Indeed,

[12] As the textbook by Cole & Scribner (1974) states, "The most firmly based [...] conclusion we can reach [...] is that [...] there is no evidence for different kinds of reasoning processes such as the old classic theories alleged—we have no evidence for a 'primitive logic'." By the way, the thesis was also repudiated by its originator towards the end of his life (Lévy-Bruhl 1949).

[13] The author wishes to thank professors N. C. A. da Costa and K. Sadegh-zadeh for their stimulating comments.

we have also constructed a system consisting of (1)the formalization of Vasil'ev's imaginary logic by Arruda (1977), (2) intuitionistic logic and (3) Łukasiewicz's three-valued logic.[14]

Professor N. C. A. da Costa has indicated how the above construction may be made completely general in one fell swoop. Loparić and he have demonstrated that any system of logic whatsoever has a two-valued semantics of valuations relative to which it is sound and complete.[15] This has the consequence that all logical systems may be treated in exactly the same way as the da Costa systems have been treated here.

[14]Lokhorst (1985 *c*).

[15]Loparić & da Costa (1984).

CHAPTER 6

The Modal Status of Antinomies

Abstract

In order to study the modal status of antinomies (provable contradictions), we present two modal extensions of the antinomic calculus proposed by F. G. Asenjo and J. Tamburino in their "Logic of antinomies", *Notre Dame Journal of Formal Logic* **16** (1975), pp. 17–44. Both systems are proved to be absolutely consistent and to be sound and complete with respect to certain Kripke-style models. It is shown that antinomies are both necessary and impossible in any case. They are provably contingent as well when serial accessibility relations between possible worlds are assumed.

First published in *The Notre Dame Journal of Formal Logic*, Volume **29** (Winter 1988), 102–105. ISSN 0029–4527.

What is the modal status of antinomies?[1] Classical modal logic provides no interesting answer to this question because it lets antinomies turn all well-formed formulas (including all modal formulas) into theorems. In the present note, we propose two nonclassical modal systems which do not suffer from this defect. Both systems are obtained by supplementing the semantics of Asenjo's and Tamburino's antinomic propositional logic L (see Asenjo & Tamburino (1975), familiarity with which will be assumed in this article) with a very natural-sounding truth condition for modal formulas. The surprising result is that antinomies are in any case both *necessary* and *impossible*: according to the second system we propose, they are both *non-necessary* and *possible* as well. It may be doubted whether these results are in accord with our intuitions. However, it should be remembered that our intuitions were formed during centuries of classical slumber; acquiring the right intuitions in antinomic thinking may simply be a matter of time.

1. The systems

1.1. The language. The language is as in Asenjo & Tamburino (1975, p. 19), but add to formation rule 2: if $\mathscr{B}_1$ is a statement form, $\Box\mathscr{B}_1$ is a statement form. Definitions:

$$\neg^*\mathscr{B}_1 \stackrel{\text{def}}{=} \mathscr{B}_1 \supset (A_1 \,\&\, \neg A_1);\ \Diamond\mathscr{B}_1 \stackrel{\text{def}}{=} \neg\Box\neg\mathscr{B}_1.$$

1.2. Semantics. An antinomic model is a triple $\langle W, R, V\rangle$, where W is a set (of "possible worlds"), $R \subseteq W \times W$, and $V : AT \times W \mapsto \{0, 1, 2\}$. (Here AT is the set of atomic statements.) $V(A_i, w) = 0$ or 1, whereas $V(B_i, w) = 2$.

The interpretation function I is defined as follows:

1. $I(A_i, w) = V(A_i, w)$, $I(B_i, w) = V(B_i, w)$.
2. $I(\neg\mathscr{B}_1, w)$, $I(\mathscr{B}_1 \,\S\, \mathscr{B}_2, w)$, where § is a truth-functional connective: as given in the tables in Asenjo & Tamburino (1975, p. 18), suitably relativized to the world w.

[1] An antinomy is, syntactically speaking, a provable statement whose negation is also provable; semantically, it is a statement that is both true and false at all possible worlds in all models.

3. $I(\Box\mathscr{B}_1, w) = \begin{cases} 0 & \text{if } \forall w'(wRw' \Rightarrow\ I(\mathscr{B}_1, w') = 0) \\ 1 & \text{if } \exists w'(wRw' \text{ and } I(\mathscr{B}_1, w') = 1) \\ 2 & \text{otherwise} \end{cases}$

The motivation for the latter clause is straightforward. Like Asenjo & Tamburino (1975), we read "$I(\mathscr{B}_1, w) = 0(1, 2)$" as "$\mathscr{B}_1$ is true and not false (false and not true, true and false) at w". Hence clause 3 is merely another way of stating the familiar and intuitively plausible condition 3':

3′. $\Box\mathscr{B}_1$ is $\begin{cases} \text{true at } w \text{ (in a model } M) \text{ iff } \mathscr{B}_1 \text{ is true at} \\ \text{all } w' \text{ accessible from } w \text{ (in } M) \\ \text{false at } w \text{ (in } M) \text{ iff } \mathscr{B}_1 \text{ is false at some} \\ w' \text{ accessible from } w \text{ (in } M).^2 \end{cases}$

We say that $\mathscr{B}_1$ is *valid* in $\langle W, R, V\rangle$ iff for all $w \in W\, I(\mathscr{B}_1, w) \neq 1$ (i.e., iff $\mathscr{B}_1$ is true at all $w \in W$ in the model).

A *serial* antinomic model is an antinomic model satisfying the condition that $\forall w \exists w' wRw'$.

1.3. Axiomatization. $\mathscr{A}$-formulas are determined as in Asenjo & Tamburino (1975, pp. 20–21), but add to C2a: if $\mathscr{A}_1$ is an $\mathscr{A}$-formula, $\Box\mathscr{A}_1$ is an $\mathscr{A}$-formula.

The axioms of M are as follows. M1–M13 are the same as L1–L13 (Asenjo & Tamburino 1975, p. 21). To these we add:

M14: $\Box(\mathscr{B}_1 \supset \mathscr{B}_2) \supset (\Box\mathscr{B}_1 \supset \Box\mathscr{B}_2)$.
M15a: $\neg\Box\neg^*\mathscr{B}_1 \supset \Diamond\mathscr{B}_1$.[3]
M15b: $\Diamond\mathscr{B}_1 \supset \neg\Box\neg^*\mathscr{B}_1$.

The axioms of MD are those of M plus:

D: $\Box\mathscr{B}_1 \supset \Diamond\mathscr{B}_1$.

The rules of both M and MD are modus ponens (R1) and $\mathscr{B}_1/\Box\mathscr{B}_1$ (R2).

1.4. Soundness and completeness.

Theorem: $\mathscr{B}_1$ is valid in the class of all antinomic models (all serial antinomic models) iff $\mathscr{B}_1$ is derivable in M (MD).

Proof: from right to left: trivial.

From left to right: we define the canonical model $\langle W, R, V\rangle$ for M (MD) as follows:

1. W is the set of all subsets of the language absolutely consistent and complete with respect to M (MD).
2. $R = \{\langle w, w'\rangle \in W \times W :$ for all $\mathscr{B}_1 : w \vdash \Box\mathscr{B}_1 \Rightarrow\ w' \vdash \mathscr{B}_1\}$. (Here and in the following, $\vdash$ stands for M- (MD-) derivability.)
3. $V(\mathscr{B}_1, w) = \begin{cases} 0 & \text{if } w \vdash \mathscr{B}_1 \text{ and } w \nvdash \neg\mathscr{B}_1 \\ 1 & \text{if } w \nvdash \mathscr{B}_1 \\ 2 & \text{if } w \vdash \mathscr{B}_1 \text{ and } w \vdash \neg\mathscr{B}_1. \end{cases}$

Lemma: . For all $\mathscr{B}_1$ and all $w \in W$ in the canonical model:

$$I(\mathscr{B}_1, w) = \begin{cases} 0 & \text{if } w \vdash \mathscr{B}_1 \text{ and } w \nvdash \neg\mathscr{B}_1 \\ 1 & \text{if } w \nvdash \mathscr{B}_1 \\ 2 & \text{if } w \vdash \mathscr{B}_1 \text{ and } w \vdash \neg\mathscr{B}_1. \end{cases}$$

[2] Using the definition of $\Diamond$, the corresponding condition for $\Diamond\mathscr{B}_1$ turns out to be: $\Diamond\mathscr{B}_1$ is true at w (in M) iff $\mathscr{B}_1$ is true at some w' such that wRw' (in M); $\Diamond\mathscr{B}_1$ is false at w (in M) if $\mathscr{B}_1$ is false at all w' such that wRw' (in M). Similar truth conditions (for tensed instead of modalized formulas) are to be found in Priest (1982, Section 3.2).

[3] M15a is interderivable with $\neg\Box(\mathscr{B}_1 \supset \mathscr{B}_2) \supset \Diamond\mathscr{B}_1$.

Proof of lemma: The proof is by induction on the length of $\mathscr{B}_1$. In case $\mathscr{B}_1$ is atomic, the lemma holds by definition. Inductive Hypothesis (I.H.): the lemma holds for $\mathscr{B}_1$, $\mathscr{B}_2$. (i) Then it holds for $\neg\mathscr{B}_1$, $\mathscr{B}_1 \S \mathscr{B}_2$, where § is a truth-functional connective: see Asenjo & Tamburino (1975, proof of Proposition 4.12, pp. 33–37). (ii) Then it holds for $\Box\mathscr{B}_1$. There are three subcases.

Subcase 1. Suppose $w \vdash \Box\mathscr{B}_1$ and $w \nvdash \neg\Box\mathscr{B}_1$. The first conjunct implies $\forall w'(wRw' \Rightarrow\ w' \vdash \mathscr{B}_1)$ by definition of R. The second conjunct implies $w \nvdash \neg\Box\neg^*\neg\mathscr{B}_1$ by M15a and R1, hence $w \vdash \Box\neg^*\neg\mathscr{B}_1$ by completeness of w, hence $\forall w'(wRw' \Rightarrow\ w' \nvdash \neg\mathscr{B}_1)$ by definition of R and absolute consistency of w'. Combination of both consequents and application of I.H. and definition of I yields $I(\Box\mathscr{B}_1, w) = 0$.

Subcase 2. Suppose $w \nvdash \Box\mathscr{B}_1$. Consider $N(w) = \{\mathscr{B}_2 : w \vdash \Box\mathscr{B}_2\}$. Suppose $N(w) \cup \{\neg^*\mathscr{B}_1\} \vdash A_1 \,\&\, \neg A_1$. Then $N(w) \vdash \neg^*\neg^*\mathscr{B}_1$ by deduction theorem, hence $N(w) \vdash \mathscr{B}_1$ by antinomic propositional calculus L, which means there are $\mathscr{B}_3, \ldots, \mathscr{B}_n \in N(w)$ such that $\vdash \mathscr{B}_3 \supset (\mathscr{B}_4 \supset \ldots \supset (\mathscr{B}_n \supset \mathscr{B}_1)\ldots)$ by definition of derivability and deduction theorem ($n-2$ times). R2 yields $\vdash \Box(\mathscr{B}_3 \supset (\mathscr{B}_4 \supset \ldots \supset (\mathscr{B}_n \supset \mathscr{B}_1)\ldots))$, whence $\{\Box\mathscr{B}_3, \ldots, \Box\mathscr{B}_n\} \vdash \Box\mathscr{B}_1$ by M14 and R1 ($n-2$ times), whence $w \vdash \Box\mathscr{B}_1$ by completeness of w—a contradiction. Hence $N(w) \cup \{\neg^*\mathscr{B}_1\}$ is absolutely consistent. Therefore $\exists w' \in W(wRw'$ and $w' \vdash \neg^*\mathscr{B}_1)$ by Lindenbaum's lemma (compare Asenjo & Tamburino (1975, Lemma 4.11)), whence $\exists w'(wRw'$ and $w' \nvdash \mathscr{B}_1)$ by absolute consistency of w', whence $I(\Box\mathscr{B}_1, w) = 1$ by I.H. and definition of I.

Subcase 3. Suppose $w \vdash \Box\mathscr{B}_1$ and $w \vdash \neg\Box\mathscr{B}_1$. The first conjunct implies $\forall w'(wRw' \Rightarrow\ w' \vdash \mathscr{B}_1)$ by definition of R. The second conjunct implies $w \vdash \neg\Box\neg^*\neg\mathscr{B}_1$ by M15b and R1. $\neg\Box\neg^*\neg\mathscr{B}_1$ is an $\mathscr{A}$-formula by C1b and C2a, hence $w \nvdash \Box\neg^*\neg\mathscr{B}_1$ by absolute consistency of w. By the same reasoning as in Subcase 2 we have $\exists w'(wRw'$ and $w' \nvdash \neg^*\neg\mathscr{B}_1)$, whence $\exists w'(wRw'$ and $w' \vdash \neg\mathscr{B}_1)$ by completeness of w'. By I.H. we have not $\exists w'(wRw'$ and $I(\mathscr{B}_1, w') = 1)$ and *not* $\forall w'(wRw' \Rightarrow\ I(\mathscr{B}_1, w') = 0)$, whence $I(\Box\mathscr{B}_1, w) = 2$ by definition of I.

This completes the proof of the lemma.

Completeness follows (compare Asenjo & Tamburino (1975, p. 39)): Suppose $\nvdash_M \mathscr{B}_1$ ($\nvdash_{MD} \mathscr{B}_1$). Then there is a $w \in W$ in the canonical model for M (MD) such that $w \nvdash_M \mathscr{B}_1$ ($w \nvdash_{MD} \mathscr{B}_1$) by Lindenbaum's lemma, whence $I(\mathscr{B}_1, w) = 1$ by our Lemma. The canonical model for M (MD) is an antinomic model (serial antinomic model), hence $\mathscr{B}_1$ is not valid in the class of all antinomic models (all serial antinomic models).

2. The modal status of antinomies in M and MD

Observation 1: If $\mathscr{B}_1$ is an antinomy, the schemas $\Box^k\mathscr{B}_1$ and $\neg\Diamond^k\mathscr{B}_1$ are theorems of M and MD for every $k \geq 0$ ($k \in \mathbb{N}$). (Proof: by induction on length of formula, using R2.)

Observation 2: If $\mathscr{B}_1$ is an antinomy, every instance of the schema $\Sigma\mathscr{B}_1$, where Σ is any (!) sequence of occurrences of $\neg$, $\Box$, and $\Diamond$, is a theorem of MD. (Proof: by induction on length of formula, using R1, R2, and D.)

The latter observation does not imply, however, that every statement concerning the modal status of antinomies is provable in MD. In fact, there are infinitely many of such statements which are unprovable; for example, each statement of the form $\neg^*\Sigma B_1$, where Σ is any sequence of occurrences of $\neg$, $\Box$, and $\Diamond$, is invalid in the class of all serial models and therefore neither provable in M nor in MD. Hence:

Observation 3: M and MD are absolutely consistent.

The author wishes to thank Professors F. G. Asenjo, N. C. A. da Costa and K. Fine for helpful comments.

Bibliography

Anderson, J. A., Pellionisz, A. & Rosenfeld, E., eds (1990), *Neurocomputing 2: Directions for research*, MIT Press, Cambridge, Mass.

Anscombe, G. E. M. (1959), *An Introduction to Wittgenstein's Tractatus*, Hutchinson, London.

Aquinas (1894), *Summa Theologica*, Vol. 1, second edn, Forzani et S., Rome.

Arruda, A. I. (1977), On the imaginary logic of N. A. Vasil'ev, *in* A. I. Arruda, N. C. A. da Costa & R. B. Chuaqui, eds, 'Non-Classical Logics, Model Theory, and Computability', North-Holland, Amsterdam.

Arruda, A. I. (1984), 'N. A. Vasil'ev: A forerunner of paraconsistent logic', *Philosophia Naturalis* **21**, 472–491.

Asenjo, F. G. & Tamburino, J. (1975), 'Logic of antinomies', *Notre Dame Journal of Formal Logic* **16**, 17–44.

Bacon, J. (1979), 'The logical form of perception sentences', *Synthese* **41**, 271–308.

Bacon, J. (1981), 'Purely physical modalities', *Theoria* **47**, 134–141.

Bacon, J. (1988), 'Four modal modelings', *Journal of Philosophical Logic* **17**, 91–114.

Bacon, J. (1989), 'A single primitive trope relation', *Journal of philosophical logic* **18**, 141–154.

Barnes, J., ed. (1984), *The Complete Works of Aristotle: The Revised Oxford Translation*, Vol. 1, Princeton University Press, Princeton, N. J.

Barwise, J. (1981), 'Scenes and other situations', *The Journal of Philosophy* **78**, 369–397.

Bäumker, C. (1877), *Des Aristoteles Lehre vom den äussern und innern Sinnesvermögen*, Leipzig.

Belnap, Jr., N. D. (1977), A useful four-valued logic, *in* J. M. Dunn & G. Epstein, eds, 'Modern Uses of Multiple-Valued Logic', Reidel, Dordrecht.

Berger, G. (1989), On self-awareness, *in* S. Lindström & W. Rabinowicz, eds, 'In So Many Words: Philosophical Essays Dedicated to Sven Danielsson on the Occasion of His Fiftieth Birthday', University of Uppsala, Uppsala, pp. 215–224.

Berghel, H. L. (1978), Harman's Tractarian thoughts, *in* Leinfellner, Leinfellner, Berghel & Hübner (1978).

Beth, E. W. (1959), *The foundations of mathematics*, Amsterdam.

Black, M. (1964), *A Companion to Wittgenstein's 'Tractatus'*, Cornell University Press, Ithaca, N.Y.

Block, I. (1964), 'Three German commentators on the individual senses and the common sense in Aristotle's psychology', *Phronêsis* **9**, 58–63.

Brentano, F. (1924), *Psychologie vom empirischen Standpunkt*, Vol. 1, Felix Meiner, Hamburg.

Brentano, F. (1973), *Psychologie vom empirischen Standpunkt, Vol. 1 (1874)*, Meiner, Hamburg.

Carnap, R. (1956), *Meaning and Necessity: A Study in Semantics and Modal Logic*, 2nd edn, University of Chicago Press, Chicago.

Charlton, W. (1981), 'Telling the difference between sweet and pale', *Apeiron* **15**, 103–114.

Chellas, B. F. (1980), *Modal Logic: An Introduction*, Cambridge University Press, Cambridge.

Chellas, B. F. (1991), 'Time and modality in the logic of agency', *Studia logica* . to appear.

Chisholm, R. M. (1981), *The first person: An essay on reference and intentionality*, Brighton.

Chisholm, R. M., Marek, J. C., Blackmore, J. T. & Hübner, A., eds (1985), *Proceedings of the Ninth International Wittgenstein Symposium*, Hölder-Pichler-Tempsky, Vienna.

Churchland *née* Smith, P. (1986), *Neurophilosophy*, M.I.T. Press., Cambridge, Mass.

Clark, R. (1976), 'Old foundations for a logic of perception', *Synthese* **33**, 75–99.

Cocchiarella, N. B. (1984), Philosophical perspectives on quantification in tense and modal logic, *in* D. Gabbay & F. Günthner, eds, 'Handbook of Philosophical Logic, Vol. 2', D. Reidel, Dordrecht.

Code, A. (1976), 'Aristotle's response to Quine's objections to modal logic', *Journal of philosophical logic* **5**, 159–186.

Cole, M. & Scribner, S. (1974), *Culture and Thought: A Psychological Introduction*, Wiley, New York.

Copi, I. M. (1958), "Tractatus' 5.542', *Analysis* **18**, 102–104.

Czermak, J. (1979), Ein mengentheoretisches Modell der *Tractatus*-Ontologie, *in* H. L. Berghel, A. Hübner & E. Köhler, eds, 'Proceedings of the Third International Wittgenstein Symposium', Hölder-Pichler-Tempsky, Vienna.

da Costa, N. C. A. (1974), 'On the theory of inconsistent formal systems', *Notre Dame Journal of Formal Logic* **15**, 497–510.

da Costa, N. C. A. & Alves, E. H. (1977), 'A semantical analysis of the calculi C_n', *Notre Dame Journal of Formal Logic* **18**, 621–630.

da Costa, N. C. A. & French, S. (n.d.), The logic of self-deception. Unpublised typoscript.

da Costa, N. C. A. & Marconi, D. (1987), 'An overview of paraconsistent logic in the 80s', *Monografias da Sociedade Paranaense de Matemática* **5**.

Dawkins, R. (1976), *The selfish gene*, Oxford.

de Montaigne, M. E. (1580), Apologie de Raimond Sebond, *in* 'Essais II', Simon Millanges, Bordeaux.

Descartes, R. (1662), *Renatus des Cartes De homine, figuris et latinitate donatus a Florentio Schuyl*, Leiden.

Eberle, R. A. (1970), *Nominalistic systems*, D. Reidel, Dordrecht.

Fagin, R. & Halpern, J. Y. (1988), 'Belief, awareness, and limited reasoning', *Artificial Intelligence* **34**, 39–76.

Favrholdt, D. (1964), *An Interpretation and Critique of Wittgenstein's Tractatus*, Munksgaard, Copenhagen.

Fechner, G. T. (1860), *Elemente der Psychophysik, Vol. 2*, Breitkopf und Härtel, Leipzig.

Field, H. (1978), 'Mental representation', *Erkenntnis* **13**, 9–61.

Fitting, M. (1969), 'Logics with several modal operators', *Theoria* **35**, 259–266.

Fogelin, R. J. (1976), *Wittgenstein*, Routledge and Kegan Paul, London.

Foster, K. & Humphries, S., eds (1951), *Aristotle's De Anima in the version of William of Moerbeke and the Commentary of St. Thomas Aquinas*, London.

Freudenthal, H. (1960), LINCOS*: design of a language for cosmic intercourse*, North-Holland, Amsterdam.

Gabbay, D. M. (1976), *Investigations in modal and tense logics with applications to problems in philosophy and linguistics*, D. Reidel, Dordrecht.

Garson, J. W. (1978), 'Completeness of some quantified modal logics', *Logique et analyse* **21**, 153–164.

Garson, J. W. (1984), Quantification in modal logic, *in* D. M. Gabbay & F. Günth ner, eds, 'Handbook of philosophical logic', Vol. 2, D. Reidel, Dordrecht, pp. 249–307.

Goodman, N. (1951), *The structure of appearance*, Cambridge, Mass.

Griffin, J. (1965), *Wittgenstein's Logical Atomism*, Oxford University Press, Oxford.

Gunderson, K. (1975), Asymmetries and mind-body perplexities, *in* C. Y. Cheng, ed., 'Philosophical aspects of the mind-body problem', University of Hawaii Press, Honolulu, pp. 99–130.

Hamlyn, D. (1968), *Aristotle's De Anima, books II and III*, Clarendon Press, Oxford.

Harman, G. (1973), *Thought*, Princeton University Press, Princeton.

Haugeland, J. (1982), 'Weak supervenience', *American Philosophical Quarterly* **19**, 93–103.

Hintikka, J. (1958), 'On Wittgenstein's 'solipsism'', *Mind* **67**, 88–91.

Hintikka, J. (1962), *Knowledge and Belief*, Cornell University Press, Ithaca and London.

Hintikka, J. (1969), On the logic of perception, *in* N. Care & R. Grimm, eds, 'Perception and personal identity', Case Western Reserve University Press, Cleveland, pp. 140–173.

Hintikka, J. & Hintikka, M. B. (1983), 'Some remarks on (Wittgensteinian) logical form', *Synthese* **56**, 155–170.

Hughes, G. E. & Cresswell, M. J. (1972), *An Introduction to Modal Logic*, second printing edn, Methuen, London.

Humberstone, L. (1986), 'Extensionality in sentence position', *Journal of Philosophical Logic* **15**, 27–54.

Kahn, C. (1966), 'Sensation and consciousness in Aristotle's psychology', *Archiv für Geschichte der Philosophie* **48**, 43–81.

Keisler, H. J. (1971), *Model Theory for Infinitary Logic*, North-Holland, Amsterdam.

Kenny, A. (1984), *The Legacy of Wittgenstein*, Basil Blackwell, Oxford, chapter Wittgenstein's Early Philosophy of Mind.

Kosman, L. (1975), 'Perceiving that we perceive: On the Soul, III.2', *Philosophical Review* **84**, 499–519.

Küng, G. (1963), 'Ontologie und logistische Analyse der Sprache'.

Lee, D., ed. (1980), *Wittgenstein's Lectures, Cambridge 1930–1932*, Blackwell, Oxford.

Leinfellner, E., Leinfellner, W., Berghel, H. L. & Hübner, A., eds (1978), *Proceedings of the Second International Wittgenstein Symposium*, Hölder-Pichler-Tempsky, Vienna.

Lenzen, W. (1978), 'Recent work in epistemic logic', *Acta Philosophica Fennica* **30**.

Lenzen, W. (1980), *Glauben, Wissen und Wahrscheinlichkeit*, Springer, Vienna.
Lévy-Bruhl, L. (1910), *Les fonctions mentales dans les sociétés inférieures*, Alcan, Paris.
Lévy-Bruhl, L. (1925), *La mentalité primitive*, 4th edn, Alcan, Paris.
Lévy-Bruhl, L. (1949), *Les carnets*, Presses universitaires de France, Paris.
Lewis, D. (1978), 'Truth in fiction, with postcript', *American Philosophical Quarterly* **15**, 37–46.
Lewis, D. (1982), 'Logic for equivocators', *Noûs* **16**, 431–441.
Lewis, D. (1983), *Philosophical papers I*, Oxford University Press, New York and Oxford.
Lewis, D. (1986), *On the plurality of worlds*, Basil Blackwell, Oxford.
Lokhorst, G. J. C. (1982), 'An ancient Greek theory of hemispheric specialization', *Clio medica* **17**, 33–38.
Lokhorst, G. J. C. (1985 *a*), A formalization of Wittgenstein's early theory of belief, *in* Chisholm, Marek, Blackmore & Hübner (1985).
Lokhorst, G. J. C. (1985 *b*), Ontologie, semantiek en filosofie van de geest in Wittgensteins Tractatus. Unpublished thesis, Erasmus University, Rotterdam.
Lokhorst, G. J. C. (1985 *c*), Pluralistic doxastic logic. Appendix of *Ontologie, semantiek en filosofie van de geest in Wittgensteins Tractatus*, unpublished thesis, Erasmus University Rotterdam.
Lokhorst, G. J. C. (1988 *a*), 'Ontology, semantics and philosophy of mind in Wittgenstein's *Tractatus*: A formal reconstruction', *Erkenntnis* **29**, 35–75.
Lokhorst, G. J. C. (1988 *b*), Psychofysische superveniëntie, *in* W. Callebaut & P. Mostert, eds, 'Acta filosofiedag Maastricht 1987', Eburon, Delft, pp. 161–165.
Lokhorst, G. J. C. (1989), 'Multiply modal extensions of Da Costa's C_n, $0 \leq n \leq \omega$, logical relativism, and the imaginary', *The Journal of Non-Classical Logic* **5**(2), 7–22.
Loparić, A. (1977), 'Une étude sémantique de quelques calculs propositionnels', *Comptes Rendus de l'Académie des Sciences de Paris, Série A* **284**, 835–838.
Loparić, A. & da Costa, N. C. A. (1984), 'Paraconsistency, paracompleteness, and valuations', *Logique et analyse* **27**, 119–131.
Lukes, S. (1967), Some problems about rationality, in *Rationality* (Wilson 1970).
Menger, K. (1974), *Morality, Decision, and Social Organization: Toward a Logic of Ethics (1934)*, Reidel, Dordrecht.
Menger, K. (1980), Language and mathematics, *in* R. Haller & W. Grassl, eds, 'Proceedings of the Fourth International Wittgenstein Symposium', Hölder-Pichler-Tempsky, Vienna.
Modrak, D. K. W. (1981), 'Koinē aisthēsis and the discrimination of sensible differences in De Anima III.2', *Canadian journal of philosophy* **11**, 405–423.
Modrak, D. K. W. (1987), *Aristotle: The power of perception*, Chicago.
Mudersbach, K. (1968), Versuch einer Axiomatisierung der Sätze 1–2.063 des *Tractatus Logico-Philosophicus*. Unpublished.
Mudersbach, K. (1978), A tentative axiomatization of the ontology of the *Tractatus*, *in* Leinfellner et al. (1978).
Natsoulas, T. (1988), 'Is any state of consciousness self-intimating?', *Journal of mind and behavior* **9**, 167–204.
Natsoulas, T. (1989), 'An examination of four objections to self-intimating states of consciousness', *Journal of mind and behavior* **10**, 63–116.
Neuhäuser, J. (1878), *Aristoteles' Lehre von dem sinnlichen Erkenntnisvermögen und seinen Organen*, Leipzig.
Osborne, C. (1983), 'Aristotle, De Anima 3.2: How do we perceive that we see and hear?', *Classical Quarterly* **33**, 401–411.
Perzanowski, J. (1985), Some observations on modal logics and the *Tractatus*, *in* Chisholm et al. (1985).
Priest, G. (1982), 'To be and not to be: dialectical tense logic', *Studia Logica* **41**, 249–268.
Puga, L. Z. & da Costa, N. C. A. (1988), 'On the imaginary logic of N. A. Vasil'ev', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **34**, 205–211.
Putnam, H. (1968), Is logic empirical?, *in* R. Cohen & M. Wartofsky, eds, 'Boston Studies in the Philosophy of Science', Vol. 5, Reidel, Dordrecht.
Rennie, M. K. (1970), 'Models for multiply modal systems', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **16**, 175–186.
Rescher, N. & Brandom, R. (1980), *The logic of inconsistency*, Basil Blackwell, Oxford.
Rodier, G. (1900), *Aristote, Traité de l'âme*, Vol. 2, Leroux, Paris.
Ross, W. (1961), *Aristotle: De Anima*, Clarendon Press, Oxford.
Ross, W. D. (1949), *Aristotle*, 5th edn, London.
Routley, R. & Routley, V. (1975), 'The role of inconsistent and incomplete theories in the logic of belief', *Communication and Cognition* **8**, 185–235.
Royce, J. (1912), *The world and the individual, first series*, New York.

Rucker, R. (1984), *Infinity and the mind*, London.

Russell, B. (1918), The philosophy of logical atomism (1918), *in* R. C. Marsh, ed., 'Logic and Knowledge', Allen and Unwin, London.

Russell, B. (1922), Introduction, in *Tractatus Logico-Philosophicus, ed. and tr. by D. F. Pears and B. F. McGuinness* (Wittgenstein 1971).

Russell, B. (1927), Appendix C, *in* 'Principia Mathematica', 2nd edn, Cambridge University Press, Cambridge.

Russell, B. (1984), *Theory of Knowledge: The 1913 Manuscript, ed. by E. R. Eames and K. Blackwell*, Allen and Unwin, London.

Ryle, G. (1949), *The Concept of Mind*, Hutchinson, London.

Sanford, D. H. (1975), 'Infinity and vagueness', *Philosophical Review* **84**, 520–535.

Schiller, J. (1975), 'Aristotle and the concept of awareness in sense perception', *Journal of the History of Philosophy* **13**, 283–296.

Shields, C. (1991), The first functionalist, *in* J. C. Smith, ed., 'Historical foundations of cognitive science', Dordrecht.

Simons, P. (1986), 'Tractatus Mereologico-Philosophicus?', *Grazer Philosophische Studien* **28**, 165–186.

Skyrms, B. (1981), 'Tractarian nominalism', *Philosophical Studies* **40**, 199–206.

Slakey, T. (1961), 'Aristotle on sense perception', *Philosophical Review* **70**, 470–484.

Soames, B. (1983), 'Generality, truth functions, and expressive capacity in the *Tractatus*', *Philosophical Review* **92**, 573–589.

Sorabji, R. (1974), 'Body and soul in Aristotle', *Philosophy* **49**, 63–89.

Sperber, D. (1982), Apparently irrational beliefs, *in* M. Hollis & S. Lukes, eds, 'Rationality and Relativism', Basil Blackwell, Oxford.

Stalnaker, D. (1984), *Inquiry*, MIT Press, Cambridge, Mass.

Stalnaker, R. (1976), Propositions, *in* A. MacKay & D. Merrill, eds, 'Issues in the Philosophy of Language', Yale University Press, New Haven.

Stalnaker, R. (1991), 'The problem of logical omniscience, i', *Synthese* **89**, 425–440.

Stegmüller, W. (1966), 'Eine modelltheoretische Präzisierung der Wittgensteinschen Bildtheorie', *Notre Dame Journal of Formal Logic* **7**, 181–195.

Stekla, H. (1970), *Der Regressus ad Infinitum bei Aristoteles*, Meisenheim am Glan. Beihefte zur Zeitschrift für phi-losophische Forschung, Heft 22.

Suszko, R. (1968), 'Ontology in the *Tractatus* of L. Wittgenstein', *Notre Dame Journal of Formal Logic* **9**, 7–33.

Thomason, R. H. (1973), Perception and individuation, *in* M. K. Munitz, ed., 'Logic and Ontology', New York University Press, New York, pp. 261–285.

Urmson, J. O. (1956), *Philosophical Analysis*, Basil Blackwell, Oxford.

Urquhart, A. (1991), Russellian propositions. Lecture at the Ninth International Congress of Logic, Methodology and the Philosophy of Science, Uppsala.

van der Eijk, P. J. (1991), Aristoteles: Über Träume, Über die Weissagung im Schlaf, PhD thesis, University of Leiden, Leiden.

Vasil'ev, N. A. (1912), 'Воображаемая (неаристотелева) логика', *Журнал Министерства Народного Просвещения* **40**, 207–246. Reprinted in (Vasil'ev 1989).

Vasil'ev, N. A. (1989), *Воображаемая логика: Избранные труды*, Издательство Наука, Москва.

von Wright, G. H. (1982), Modal logic and the *Tractatus*, *in* 'Wittgenstein', Basil Blackwell, Oxford.

Wansing, H. (1990), 'A general possible worlds framework for reasoning about knowledge and belief', *Studia logica* **49**, 523–539.

Weiskrantz, L. (1986), *Blindsight: A case study and implications*, Oxford.

Wilson, B. R. (1970), *Rationality*, Basil Blackwell, Oxford.

Winch, P. (1964), Understanding a primitive society, in *Rationality* (Wilson 1970).

Wittgenstein, L. (1964), *Philosophische Bemerkungen, ed. by R. Rhees*, Basil Blackwell, Oxford.

Wittgenstein, L. (1971), *Tractatus Logico-Philosophicus, ed. and tr. by D. F. Pears and B. F. McGuinness*, 2nd edn, Routledge and Kegan Paul, London.

Wittgenstein, L. (1979), *Notebooks, ed. and tr. by G. H. Von Wright and G. E. M. Anscombe*, 2nd edn, Basil Blackwell, Oxford.

Wolniewicz, B. (1968), *Rzeczy i fakty*, Panstwowe Wydawnictwo Naukowe, Warsaw.

Wundt, W. (1874), *Grundzüge der physiologischen Psychologie*, Engelmann, Leipzig.

Żarnecka-Biały, E. (1985), Truth and logic in Vasil'ev's imaginary worlds, *in* 'Proceedings of the Bulgarian-Polish Symposium on Logic', Bulg. Akad. Nauk, Sofia, pp. 11–19.

Samenvatting

In de zes artikelen waaruit dit proefschrift bestaat, worden formeel-logische analyses gegeven van de volgende onderwerpen op het gebied van de filosofie van de geest:

(1) Aristoteles' opvattingen over perceptueel zelf-bewustzijn, het fenomeen dat zich voordoet wanneer we "waarnemen" dat we iets waarnemen. In Hoofdstuk 1 laten we eerst zien hoe de ogenschijnlijk tegenstrijdige theorieën die te vinden zijn in *De anima* III.2 en *De somno* II met elkaar verzoend kunnen worden. (Geen commentator is hier tot nu toe in geslaagd.) Vervolgens geven we een axiomatisering van Aristoteles' fundamentele ideeën over de zintuiglijke waarneming. We laten zien dat de—tot dusver veelal onbegrepen—argumenten in beide werken gereconstrueerd kunnen worden als correcte bewijzen binnen de resulterende axiomatische stelsels. We eindigen met het geven van bewijzen voor de consistentie van deze stelsels. Aristoteles' ideeën zijn niet slechts van historisch belang, maar kunnen gebruikt worden om de theorieën van zulke uiteenlopende moderne auteurs als de bioloog Richard Dawkins, de filosoof Keith Gunderson en de wiskundige Rudy Rucker—volgens wie compleet zelf-bewustzijn buiten het bereik van eindige wezens ligt, omdat het noodzakelijkerwijze een "oneindige regressie" van mentale fenomenen met zich meebrengt—te weerleggen.

(2) Wittgensteins vroege opvattingen over mentale representatie. (Hoofdstuk 2, 3 en 4.)

(2a) Wittgensteins "taal van het denken" theorie. Volgens Wittgenstein kan de geest worden opgevat als een verzameling zinnen, die de werkelijkheid op dezelfde manier weerspiegelen als zinnen dat doen. In Hoofdstuk 2 en 4 presenteren wij een precieze theorie die beschrijft hoe dit in zijn werk gaat.

(2b) Wittgensteins analyse van propositionele attitude toeschrijvingen (zinnen zoals "A gelooft dat p", "A weet dat p", "A denkt dat p", en "A ziet dat p"). Volgens Wittgenstein is de zin "A denkt dat p" vergelijkbaar met de uitspraak "zin 'p' zegt dat p". Wanneer we zeggen dat A gelooft dat p, beweren we dat A's geest minstens één gedachte g bevat die zegt dat p; omdat gedachten vergelijkbaar zijn met zinnen is de tweede helft van deze bewering (gedachte g zegt dat p) vergelijkbaar met "'p' zegt dat p". De laatste uitspraak is echter *unsinnig*. We kunnen niet zeggen wat een zin betekent maar dat uitsluitend tonen; "'p' zegt dat p" behoort niet tot de taal en heeft geen waarheidswaarde. Zijn propositionele attitude toeschrijvingen ook *unsinnig*? Dit zou een eigenaardige opvatting zijn, en er zijn geen aanwijzingen dat Wittgenstein haar zelf aanvaardde. Een theorie waarin propositionele attitude toeschrijvingen vergelijkbaar zijn met uitspraken zoals "'p' zegt dat p" maar niettemin gewone zinnen van de taal zijn, zou aantrekkelijker zijn. De grootste moeilijkheid bij het opstellen van een dergelijke theorie is dat we zullen moeten laten zien dat propositionele attitude toeschrijvingen geen inbreuk maken op Wittgensteins waarheidsfunctionaliteitsprincipe, dat stelt dat de waarheidswaarden van alle zinnen volledig worden bepaald door de waarheidswaarden van de zogenaamde "elementaire zinnen". Niemand is er tot dusver in geslaagd om een theorie op te stellen die aan deze *desiderata* voldoet, maar in Hoofdstuk 2

en 3 presenteren wij er een. Onze theorie is in feite een superveniëntie-theorie van dezelfde soort als tegenwoordig zo in de belangstelling staat.

Omdat Wittgensteins ideeën over mentale representatie en propositionele attitude toeschrijvingen niet begrepen kunnen worden zonder aandacht te besteden aan zijn fundamentele semantische en ontologische opvattingen, bespreken en formaliseren we deze ook, althans voor zover dat mogelijk en nuttig is.

(2c) Wittgensteins opmerkingen over het "metafysische" of "transcendentale" subject. In tegenstelling tot het "empirische" zelf, dat door de psychologie bestudeerd wordt (en dat in Hoofdstuk 2 en 3 aan de orde komt), is dit geen verzameling zinnen: het representeert niet en vervult ook geen enkele andere functie. (Hoofdstuk 4.)

(3) Lévy-Bruhls theorie van het "logisch relativisme"—de theorie dat verschillende volkeren verschillende logica's kunnen aanhangen, hun eigen maatstaven van rationaliteit kunnen hebben, en in verschillende "denkbare werelden" kunnen leven. In Hoofdstuk 5 laten wij zien dat deze these binnen de logica ondergebracht kan worden: wij gebruiken Da Costa's paraconsistente (inconsistentie-tolererende) calculi C_n, $1 \leq n \leq \omega$, om een oneindig aantal systemen te construeren die het allemaal mogelijk maken om over de overtuigingen van volkeren te spreken die er andere logica's op na houden dan wij.

We hebben een kort technisch artikel toegevoegd (Hoofdstuk 6) waarin volledigheidsbewijzen worden gepresenteerd van twee systemen die lijken op de systemen uit Hoofdstuk 5; net zoals de laatste, zijn het paraconsistente modale logica's.

De formele analyses die we van al deze verschillende filosofische opvattingen geven voldoen aan alle eisen van precisie, objectiviteit en controleerbaarheid die gewoonlijk aan formele theorieën worden gesteld. We denken dat een filosofische analyse pas voltooid is als ze in een dergelijke logische vorm is gegoten. We betrekken alle verkregen theorieën op moderne discussies in de filosofie van de geest—en sommige zelfs op het gebied van de *kunstmatige* intelligentie.

Hoewel de logische benadering die we volgen op een traditie van dertig jaar kan bogen, zijn alle formele theorieën die we opstellen nieuw. Dit toont aan dat de filosofie niet het enige gebied is dat profiteert van de logische benadering: deze activiteit kan ook een stimulerend effect op de logica zèlf hebben. Dit is niet verwonderlijk, omdat de verdiensten en tekortkomingen van de theorieën die zij heeft ontwikkeld pas aan het licht kunnen treden als we er een nuttig gebruik van proberen te maken.

We denken daarom dat het een betreurenswaardige zaak is dat oude filosofische theorieën slechts zelden vanuit een modern logisch perspectief worden beschouwd. We hopen met dit proefschrift enige verbetering in deze situatie te hebben gebracht.

Curriculum vitæ

Gerrit Jan Cornelis Lokhorst (roepnaam Gert-Jan) werd op 16 februari 1957 te Gouda geboren. In 1975 legde hij het eindexamen gymnasium-β af aan het Christelijk Lyceum te Delft. In 1980 behaalde hij het doctoraalexamen geneeskunde en in 1985 het doctoraalexamen wijsbegeerte (*cum laude*), beide aan de Erasmus Universiteit Rotterdam (EUR).

Van 1979 tot 1989 vervulde hij diverse aanstellingen aan de Faculteit der Wijsbegeerte van de EUR, zowel in de wijsgerige antropologie als in de logica. In 1989 was hij toegevoegd onderzoeker in de theorie van neurale netwerken aan het Instituut voor Taal– en Kennistechnologie van de Katholieke Universiteit Brabant. Vanaf 1989 is hij universitair docent filosofie en informatica en universitair docent logica aan de Faculteit der Wijsbegeerte van de EUR.

Hij is de auteur van *Brein en bewustzijn: de geest-lichaam theorieën van moderne hersenonderzoekers, 1956–1986* (Rotterdamse filosofische studies, deel 5; Delft, Eburon, 1986). Bovendien publiceerde hij artikelen in *The Behavioral and Brain Sciences*, *Bulletin of the History of Medicine*, *Clio Medica*, *Erkenntnis*, *Intermediair*, *Journal of Non-Classical Logic*, *Mens en Melodie*, *Neurology*, *Nieuwstribune*, *Notre Dame Journal of Formal Logic*, *NRC-Handelsblad*, *Personal Computer Magazine*, *Philosophical Investigations*, *Piano Bulletin* en *Tijdschrift voor Marketing*.

Zijn doctoraalscriptie wijsbegeerte, getiteld *Ontologie, semantiek en filosofie van de geest in Wittgensteins Tractatus: een formele analyse* (1985), werd bekroond met de Prijs van de Internationale School voor Wijsbegeerte 1986.

Stellingen

1 Aristoteles' opmerkingen over "waarnemen dat we zien en horen" in *De anima* III.2 en *De somno* II zijn niet in tegenspraak met elkaar, maar vullen elkaar aan.

Dit proefschrift, hoofdstuk 1.

2 Brentano, Husserl, Chisholm en vele anderen hebben beweerd dat mensen niet kunnen zien zonder zich ervan bewust te zijn dat ze zien. De "blind sight" experimenten tonen aan dat zij ongelijk hebben.

Dit proefschrift, hoofdstuk 1.

3 Het is mogelijk om een waarheidsfunctionele analyse van propositionele attitude toeschrijvingen te geven die in overeenstemming is met wat Wittgenstein in zijn vroege geschriften over dergelijke toeschrijvingen beweert.

Dit proefschrift, hoofdstuk 2 en 3.

4 In *Tractatus* 5.5421 neemt Wittgenstein stelling tegen Leibniz's bewering dat de ziel een spiegel van de wereld is.

Dit proefschrift, hoofdstuk 4.

5 De these van het logisch relativisme kan met logische middelen worden geëxpliciteerd.

Dit proefschrift, hoofdstuk 5.

6 De weerstand tegen niet-klassieke logica's berust grotendeels op onwennigheid.

Dit proefschrift, hoofdstuk 6.

7 De tot dusver voorgestelde formele dialectische logica's zijn niet geschikt om het systeem van Hegel te formaliseren.

G. J. C. Lokhorst, "Het formaliseren van Hegels dialectische logica", in G. Vandenakker, C. Leijenhorst en J. Prinsen, red., *Filosofiedag Utrecht 1989*, Delft, Eburon, 1989, blz. 116–125.

8 De theorie dat de twee helften van de grote hersenen van de mens functioneel van elkaar verschillen werd reeds ver vóór de negentiende eeuw verkondigd.

G. J. C. Lokhorst, "An ancient Greek theory of hemispheric specialization", *Clio Medica* **17** (1982), 33–38; "Hemisphere differences before 1800", *Behavioral and Brain Sciences* **8** (1985), 642.

9 De filosofische interpretatie van de hersenwetenschappen dient niet te worden overgelaten aan gepensioneerde hersenonderzoekers.

G. J. C. Lokhorst, *Brein en bewustzijn: de geest-lichaam theorieën van moderne hersenonderzoekers, 1956–1986*, Delft, Eburon, 1986.

10 De materialistische opvatting van het geest-lichaam probleem is bevorderlijker voor het verantwoordelijkheidsgevoel van de medicus practicus dan de dualistische.

G. J. C. Lokhorst, *Arts en Fiets* **7** nr. 5 (1990), 4–8.

11 De gezondheidsraadcommissie "Neurochirurgie op psychiatrische indicatie" houdt er een mythisch beeld van de psyche op na.

Gezondheidsraad, Commissie neurochirurgie op psychiatrische indicatie, *Neurochirurgische behandeling van patiënten met zeer ernstige psychische aandoeningen*, 's-Gravenhage, Gezondheidsraad 1990, publicatie nr. 90/21.
G. J. C. Lokhorst, "Psychochirurgie en het geest-lichaam probleem", in M. J. van den Hoven en M. Verkerk, red., *Snijden in het brein: ethische*

en juridische aspecten van psychochirurgie, Leuven, Garant, 1991, blz. 17–20.

12 De "philosophy of mind" zou zich minder door de computerwetenschappen en meer door de natuurwetenschappen moeten laten inspireren.

G. J. C. Lokhorst, "Analog automata and the foundations of cognitive science", lezing op het *Ninth International Congress of Logic, Methodology and Philosophy of Science*, Uppsala 1991, te verschijnen.

13 Eduard Wette heeft zijn stelling dat 1 gelijk is aan 0 niet overtuigend bewezen.

G. J. C. Lokhorst, "1 = 0: de logica van Eduard Wette", *NRC-Handelsblad* 29 december 1987, Bijlage Wetenschap en Onzin, blz. 4.

14 De manier waarop getallen in het Arabische schrift en in sommige computergeheugens worden gerepresenteerd (de tekens geven oplopende machten van het grondtal weer), is rekenkundig gezien handiger dan de bij ons gebruikelijke schrijfwijze. In plaats van "234" zouden we "432" of, om verwarring te voorkomen, "٢٣٤" moeten schrijven.

De notatie heeft al vaker een rotatie ondergaan: zie G. Beaujouan, "Étude paléographique sur la 'rotation' des chiffres et des *apices* du X^{e} au XII^{e} siècle", *Revue d'Histoire des Sciences* **1** (1947), 301–313.

15 De Nederlandse benaming van het getal ٢٤٢, waarin eerst de honderdtallen, vervolgens de eenheden en dan pas de tientallen worden genoemd, is ondoordacht en verwarrend. Het is beter om "vierendertig en tweehonderd" te zeggen.