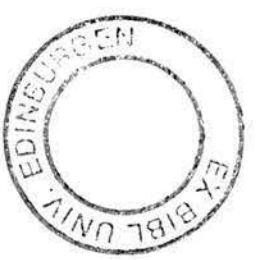
Semantics and the Stratification of Explanation in Cognitive Science

Philip L. Kime

Doctor of Philosophy

University of Edinburgh

1996



Abstract

This work is concerned with a pervasive problem in Cognitive Science which I have called the "stratificational" approach. I argue that the division into "levels of explanation" that runs as a constant theme through much work in Cognitive Science and in particular natural language semantics, is in direct conflict with neuroscientific evidence. I claim it is also in conflict with a right understanding of the philosophical notion of "evidence". The neuroscientific work is linked with the philosophical problem to provide a critique of concrete cases of research within the natural language semantics community. More recent neuroscientifically aware research is examined and it is demonstrated that it suffers similar problems due to the same deep running assumptions as those which effect traditional formalist theory. The contribution of this thesis is thought to be that of a demonstration of the essential nature and indeed the *ubiquity* of the basic assumptions in the field. Also, a new link is forged between the concerns of the formalists and certain seemingly more abstract philosophical work. This link enables us to see how much philosophical problems infect research into cognition and language. It is argued that practical research in Cognitive Science simply cannot be seen to be independent of the philosophical basis of the entire subject. The resulting picture of Cognitive Science and its place is outlined and explored with special emphasis on what I have called the "Principle of Semantic Indistinguishabliity" which says that the contribution of what can be broadly termed "environment" is epitemologically opaque to our cognition. The importance of this principle is discussed.

Declaration

I declare that the work and composition involved in this thesis were undertaken and carried out respectively by myself and no other.

Philip L. Kime

September 1996

Contents

Chapter 1

Introduction

The purpose of this work is to draw out a fundamental thread of reasoning and methodology that underlies most traditional work, and some not so traditional work, in Cognitive Science. It will be argued that this line of reasoning is at odds with the implications of modern neuroscience and cannot base a reasonable claim to "explain" human cognition. The picture I shall identify is that which I shall call "stratified". This, in general, is an attempt at explanation that divides into "levels of explanation", each with its own concepts that are said to be essential to the explanation of a phenomenon. There are specific and pragmatic manifestations of this, I discuss these in Chapter 3 and 7 in particular. There are also more abstract expressions of the same tendency which I examine mainly in Chapter 6. One of the principle tasks is to demonstrate the links between the assumptions of the more abstract formulations of this approach and their pragmatic instantiations in work in Cognitive Science. This allows it to be made clear that certain methodological problems are ubiquitous within the field and are not simply a result of the particular pragmatics of a particular research area.

In Cognitive Science as a whole, it is generally appreciated today that there are problems to do with integration of traditional formal systems and the evolutionary and biological aspects of human cognition. One aim of

this work is exactly to give an argument, supported from work in the brain sciences, that a certain methodology – particularly that enshrined within formal systems in language semantics – is strongly denied its evidential basis as a result of certain empirical considerations. It is also denied much of its basis as a result of the incongruity between the original motivations of logical formalism and the use to which this formalism is put today. The conclusion of this is that Cognitive Science's role in certain areas is severely limited and it crucially relies on an amount of empirical brain research in places thought usually to be completely separate from the "low-level" evidence from neuroscience. Part of my thesis is that stratified systems and particularly systems of formal logic within linguistics and semantics, cannot possibly be independent in the way imagined. There is also exploration of a general point regarding the character of the relation between strata in a stratified theory. There is, I shall argue, an irresolvable tension between the desire to have separate strata which are both independent but related. We shall see this both in concrete terms in the discussion of Fodor and in the abstract in the discussion of McDowell.

George Lakoff has expressed agreement with this particular premise:

> "... linguistic results ... indicate that human reason uses some of the same mechanisms involved in perception and ... human reason can be seen as growing out of perceptual and motor mechanisms."[1]

If this is correct, then I think that there are enormous implications for Cognitive Science in its practise of semantics since the mechanisms of motor and perceptual systems impose radical constraints when applied in the area of semantics.

Given this, my aim is to demonstrate that certain seemingly theory-independent areas of research in Cognitive Science such as linguistics and natural language semantics are actually infected with damaging assump-

[1](Lakoff 1988) p. 301

tions from certain misguided philosophical positions. The idea that we can simply model things in Cognitive Science and wait for someone else to sort out the theoretical structure into which all of the models will fit is not tenable. I shall demonstrate this in several concrete cases and couple this with a critique from neuroscience which is crucially related to a more philosophical critique of fundamental assumptions. The structure of the work is as follows. Firstly, I give an overview of foundational issues in Cognitive Science by discussing central works. Then, I introduce the main problems in concrete form by way of an examination of certain approaches to inference in formal semantics. Chapter 4 expands on this in an analysis of the notion of "compositionality" with reference to the "stratificational" approach I find apparent in traditional work in Cognitive Science and the assumptions it disguises. Chapter 5 introduces the themes from neuroscience and the relations they have to the philosophical critique in Chapter 6. In Chapter 7, I demonstrate that the assumptions I have identified are present even in work motivated by a desire to leave behind the formalist program. I explain why this is the case and the implications this has for a correct view of "evidence" in Cognitive Science. At this point, I deal with pertinent objections to my view stemming from the parts of the discipline I have mentioned. Chapter 8 condenses the problem and shows the fundamentals of the whole problem in relief, suggesting what all of the preceding means for Cognitive Science.

Chapter 2

Placing the Approach to Cognitive Science

It is as well to place the ensuing approach within the landscape of Cognitive Science as this reveals the fundamental assumptions that I make about the subject. These fundaments are important since the position I advocate is based heavily on rather critical foundational notions about the role of explanation and evidence in Cognitive Science. In this section, I shall outline the relevant major assumptions in Cognitive Science and shall try to say where I find them lacking, why I do so and what general philosophical strategy I advocate. This will aid in setting the scene for the more detailed arguments to come. I refrain from providing a history of Cognitive Science and AI since this has been done many, many times before and I am more concerned with the conceptual issues here than the historical placement of ideas. One might consult (Dreyfus & Dreyfus 1988) for a clear history of the two main competing Cognitive Science programmes. (Nuallain 1995) has a very comprehensive history of most things relevant to the development of Cognitive Science.

In recent years, few books have been more thorough in providing a highly detailed view of the practice of Cognitive Science than von Eckardt's

"What is Cognitive Science?"[1]. Not only does it attempt to outline the methodology in terms of aims, methods and approaches but also tries to point out philosophical assumptions and canonical examples of research. As an example of a recent survey of the foundational material of mainstream Cognitive Science, it is a useful work to focus discussion around. By examining a wide variety of approaches to standard research in the field, von Eckardt promotes a position which aims to establish the validity of certain quite traditional assumptions about Cognitive Science. The point of the book is to defend Cognitive Science against the idea that the field is too loosely connected to constitute a real subject, in the sense of conforming to some respectable methodological criteria. The issue is confused by von Eckardt freely (but explicitly, thankfully) mixing normative with descriptive criteria and so the picture that emerges is one which describes some combination of how Cognitive Science actually is with a little of how von Eckardt thinks it should be. This, as it happens, is useful since I shall highlight the central features of my normative claims by offsetting them against those that von Eckardt extracts from the major positions in the field.

Cognitive Science has inherited many different concerns from many different fields. The interesting thing is that these concerns have largely been different ways of approaching similar things and thus a coalescence of aims has resulted in certain problems becoming paramount in Cognitive Science. One of the first, coming out of Newell and Simon's "Physical Symbol System Hypothesis" in AI was an interest in the role of *symbols* as a medium for representation of the world. An interest in how to develop systems of symbols in order to be able to economically and satisfactorily mirror relations in the world became a major research programme. This combined very early on with the linguistics community who were attempting to utilise tools from the practise of logic in order to provide models

[1](von Eckardt 1993)

of language. Chomsky and Montague were pioneers in this area and, inspired by influential psychologists working in the new "cognitive" paradigm on the relationship between language and thought[2], came to be seen to be relevant not only to language but to the study of the whole structure of cognition. The most well-known modern inheritors of this tradition are Fodor and Pylyshyn who still defend a notion of symbolic representation. The computational thrust was provided partly by the advent of more powerful computers and the initially impressive results in famously limited chess programs and "block-world" research. It was also partly fuelled by the functionalist philosophy of mind that arose at the time. Together, this computational element has bound Cognitive Science quite closely to computer metaphor and simulation.

The most significant division has been that between the traditional approach and the connectionist research that has spread since the "perceptron" work of McCulloch and Pitts. Still an essentially computationalist paradigm, the connectionists famously eschewed the symbolic and formalist sides of the early Cognitive Science and advocated a more biologically realistic approach to understanding cognition. We shall return to these themes later.

2.1 Cognitive Science as a Science

It is quite common to defend the scientific status of a science by defending the nature of the basic assumptions. Popper's famous "falsificationist" approach was essentially this and Kuhn's paradigmatic model was a description of how this works. Now, it is very common for Cognitive Scientists to do the same, presumably to be in order with the rest of that which is called science. Marr's widely-known model of the visual system that prompted the most famous "levels of description" approach was exactly this: an attempt

[2]I am thinking here particularly of Vygotsky, Sapir and Whorf and Skinner

to regiment the assumptions of a field. von Eckardt is very clear on this and argues that the "foundational consensus" that the field displays is a result of coherency of basic assumptions. It is this which firmly establishes Cognitive Science as a science[3]. It is not always like this, however. One of the figures central to significant sub-fields of Cognitive Science, Noam Chomsky, was quick to realise that what really matters in the status of a science is rather the *data* that it employs. Chomsky was very keen for his brand of linguistic theory to have a solid body of data as a result[4]. In agreement with Chomsky, I think it is really the nature of the data that Cognitive Scientists of the traditional school appeal to that is responsible for any suspicions about the coherency of the field. As to whether these suspicions can be justified, I leave this question for the body of the work. Suffice to say that I think, by and large, they can. A concentration on the relationship between assumptions and practice allows one to *say* that something is a science by virtue of its connection with the world but to deny this in almost every concrete situation because one's data is of a sort that is really nothing like that of the natural sciences. This is exactly what von Eckardt allows:

> "In particular, one would expect that, *ceteris paribus*, if the foundational assumptions are wildly off base, the research program will eventually fail no matter ingenious or strenuous the research effort."[5]

The trouble is that the *ceteris paribus* clause can encompass as many assumption-denying implications as one likes as long as the data is flexible enough to allow an almost unlimited number of reinterpretations. This is a central part of my thesis and occupies a considerable part of the material following. The very tentative way in which von Eckardt approaches the issue of the "domain of enquiry" (i.e. the data of the subject) makes

---

[3](von Eckardt 1993) p.4
[4]This is explicit in Chomsky's earlier work
[5]ibid p. 4

this clear. In a telling descriptive/normative shift, von Eckardt first says that the "low-level" assumptions of the field posit phenomena that "can be assumed" to exist. Then, immediately after this, von Eckardt says that it is "important to preserve" the low-level nature of these assumptions. This is a characteristic normative strategy in Cognitive Science and displays the wisdom of Chomsky's early attempts to secure some data for linguistic research[6]. von Eckardt uses this normative requirement of "preservation of low-level data" to ensure that one can compare approaches to *the same phenomena*. Of course, this comparison does indeed require such preservation but is certainly not a good argument, or indeed any argument at all that there *is* such a secure domain of "low-level phenomena" data. It merely establishes a weaker, hypothetical claim that *if there are not such things*, then there are severe problems for comparison of approaches within Cognitive Science. Indeed, I would argue that a desire to make theoretical underdetermination more tractable does not provide evidence in any way for one's ontological assumptions, rather it completely depends on them. This is, I should hope, obvious. What is not obvious is that this mistake is consistently made in traditional Cognitive Science because of mistakes about the nature of the data. von Eckardt does allow that the ontological assumptions of Cognitive Science can change as this is apparent as a feature of every mature science. This is not enough to establish anything however as the history of science also demonstrates that a field whose ontological assumptions change does not thereby have a claim to sciencehood. History is riddled with examples of change from one ontological fiction to another.

An element of the contribution this thesis makes to the subject is to make clear a type of argument used to defend methodology in traditional Cognitive Science. It is a variation on a theme highlighted briefly above in the case of the avoidance of underdetermination. As a paradigm of foun-

[6]See Chapter 4

dational Cognitive Science research, von Eckardt displays this argument form clearly

> "...it is far easier to adopt a "divide and conquer" strategy [to explaining human cognitive capacities]. And only [this strategy], I submit, makes any methodological sense. Our capacities are difficult enough to explain one by one. If we are faced with the task of describing them in interaction at the outset, the job becomes completely daunting."[7]

Again the pattern emerges that certain foundational decisions made by Cognitive Science are based not on anything as secure as is suggested, but rather on a hypothetical of the form "if not this assumption, then no Cognitive Science". A central aim of this thesis is to demonstrate that the antecedent of these conditionals, often rushed over hastily in introductions to Cognitive Science books and papers, can be seen to be satisfied in significant cases of Cognitive Science research.

A general example of the problem with the data of the field is given by von Eckardt's "Domain Specifying Assumptions" for Cognitive Science. Take, for example:

> D2 (PROPERTY ASSUMPTION) Pretheoretically conceived, the human cognitive capacities have a number of important properties. I shall refer to these as the *basic general properties* of cognition
>
> (a) each capacity is *intentional*; that is, it involves states that have content or are "about" something. :[8]

Given that there is much debate in Cognitive Science regarding what "intentionality" actually is and when it manifests, it cannot be reasonable for a science to include this as a "property assumption". In physics, for example, it is not argued about whether something has the property of being "hot" or "cold" or "boiling" since these are exactly the sort of things that are

---

[7]ibid p. 64
[8]ibid p. 47

fit to play the part of such "property assumptions". "Being intentional" is not suitable since there is no accepted criteria of what counts as such. This is an important point and one which is addressed specifically in Chapter 4. It is a specific example of the nature of the data in Cognitive Science being so significantly different from that in the mature sciences that it threatens to enlarge the *ceteris peribus* clause mentioned above in order to ensure that no foundational questions arise. If we hardly know what "being intentional" means, we can reinterpret this in cases when our work contradicts what we currently think it means. In Chapter 4, I discuss this as the problem that Cognitive Science has little established dogma to "push against" in order to gain leverage on its problems. Indeed, I shall argue that the nature of the subject is in general, as traditionally conceived, conducive to making it almost impossible for it to have the leverage necessary to enable it to become a mature science. von Eckardt's first assumption reads:

> D1 (IDENTIFICATION ASSUMPTION) The domain of Adult Normal Typical Cognition consists of the human cognitive capacities.

As a basic datum "cognitive capacity" is hardly illuminating since in a foundational enquiry, part of that in which we are interested at this juncture is exactly to find out just what *Cognitive* Science is. I think we have little "pretheoretic" idea what "cognitive" is or exactly what is meant by a "capacity" here. In fact, telling us what these words mean seems to be a part of what Cognitive Science is trying to do. The situation with the physical sciences as examples of mature science is that although they eventually may alter and define what is designated as their domain, they famously *start out* by addressing properties and objects which we are pretheoretically happy with. This is exactly Quine's famous dictum that science is an extension of common sense.

2.2 Undermining the Foundation

It is necessary, after these rather general points, to examine specifically some arguments regarding the foundations of Cognitive Science, particularly those which I will have cause to question later on. I am specifically interested in those questions regarding the role of neuroscience since I will explicitly address these in the main body of the work. The bedrock of much Cognitive Science is formed of a belief that there is a certain "level of description" or "realm of explanation" that can be investigated essentially separately from the facts about its actual biological realisation. A weaker version of this attitude is that *something* can be learned about the subject matter by this assumption, leaving in question whether a realistic explanation can completely assume this. My view is that we cannot even seriously hold this weaker position for reasons that will become apparent in Chapters 5 and 6.

Fodor has been a major proponent of a view concerning the foundations of Cognitive Science which sees an explanatorially independent realm from the neurological. His argument is that the correlations between neural states and the cognitive phenomena that we are desirous to explain are too vague to really warrant the belief in any real constraint on Cognitive Science by neuroscience. von Eckardt sees this as a version of the "multiple-realisability" argument that seeks to keep the explanations of Cognitive Science separate because the data involved could arise in many possibly ways; thus, this independence of any *particular* possible realisation of the subject seems to show that no *particular* constraint source is relevant. This is obviously a fallacy[9] since we need not resort to modal arguments in a case where we know what the source of realisation is. There may be many possible sources of a "cognitive" ability but we are presumably really only concerned with one: the brain. So, given this, how could

[9]von Eckardt too argues against this. Ibid pp. 322–327

neuroscience not be a constraint upon Cognitive Science as long as it is concerned with human beings? Of course, von Eckardt's explicit formulation of the aims of Cognitive Science is to the study of "human cognitive capacities" and so she is quite right and indeed compelled to reject the simple multiple-realisability argument. von Eckardt realises though that Fodor's argument is a little different but her reply is, I think not wholly satisfactory as evidenced by the position she adopts later on. This point is important to discuss as it makes clear where my position fits in with the foundational literature at a crucial point of division.

So, Fodor's view challenges the non-modal correction of the situation above by arguing that there is still multiple-realisability in the *actual* supposed substrate of cognition. That is, we have little or no evidence for *specific* constraints on, say, semantic processing and thus no clear idea of constraints at all. At best, this is a merely epistemological point about lack of knowledge from which it is dangerous to draw philosophical conclusions. This is really as far as von Eckardt goes:

> "The fact that thus far only gross psycho-neural correlations have been discovered does not mean that there are no detailed correlations to be had."[10]

I think we can do better than this as a rebuttal and I think we *need* to since this leads von Eckardt into error later on (see below). It is my belief that the kinds of specific "cognitive" functions that interest Fodor, however one models or characterises them, must depend on evolutionary pre-dating capacities such as core motor skills[11] and that it is highly unlikely that these would have the sort of disparate neural realisation necessary to give Fodor's result. Nature can afford to experiment with relatively insignificant (in terms of survival) filigree such as the niceties of semantics and language but this is simply not possible with skills central the survival

[10]Ibid. p. 326

[11]The detailed argument for this claim can be found in Chapter 4

of a certain species. This assumption is, in fact, central to the biological sciences since there it is never assumed that one person might coordinate vision with one neural story and another person with a different one. Even across species we find similar realisations of core skills; this is the basis of comparative anatomy. The reason that neuroscience appears so vague and non-specific in its constraints is that the implications of its more interesting models have not been explored. It is exactly the examiniation of this that constitutes part of the task of this thesis.

Now, it is von Eckardt's weak reply to Fodor's argument that prompts a critical concession to the traditional conception of Cognitive Science. There are then, for von Eckardt three possibilities regarding the role of neuroscience and the foundations of Cognitive Science.[12]

- A "top-down" approach whereby Cognitive Science proceeds with psychology and model building at the level of "cognitive capacities" and only later tries the resultant theories against the evidence from neuroscience.

- A "bottom-up" strategy which attempts to work towards information processing models of such capacities from work in neuroscience.

- The "co-evolutionary" method which allows research in both domains as seems "fruitful"[13] and influence in both directions.

Against the top-down idea, von Eckardt argues convincingly that there is no reason to wait until we have a completed information processing theory before we can utilise the obviously relevant neuroscientific data from, for example, impairment studies. Against the bottom-up approach, it is merely stated that answers to basic questions about how people do certain (canonically "cognitive") things are dependent on an answer to the question

[12] pp. 327–328

[13] A rather colourful and vague methodological dictate given von Eckardt's careful and lengthy opening theoretical characterisation of Cognitive Science methodology.

about their information-processing implementation. This is not so much an argument as a clear example of question begging I think but so as to be as constructive as possible, I shall instead turn to what I see to be the problem with von Eckardt's final choice: the "co-evolutionary" methodology.

It is my view that a real appreciation of the perfectly good argument against the top-down approach conclusively legislates against the co- evolutionary tactic too. That is not to say that I therefore have complete faith in the bottom-up method but more of this point later. Strangely, the very fact that the neuroscientific constraints *are* so general and vague means, contra Fodor, that they undermine the very fabric of basic assumptions that the major functionalist, information-processing, symbolic and formal approaches to Cognitive Science depend critically on. This is discussed at length through that which follows whilst here, I will make a more general point. The simple historical fact is that co-evolution of approaches *did* not occur and the disparity of theoretical evolution evident as a result of this means that co-evolution is too much to hope for in the future. The motivations of the traditional approach to Cognitive Science that Fodor represents long predate the more modern science of the brain. The basis of logic and formal systems which forms the heart of the information-processing approach is comparatively mature compared with the burgeoning field of neuroscience[14]. As a result, the criteria for a "good" theory are set by the pre-dating paradigm and we have the current situation where much of the neuroscientifically motivated research is trying to address problems that have been forced onto it by the established traditional approach. This point is addressed in detail later but it serves to show that a co-evolutionary approach is an ideal, the main premise of which cannot be satisfied. Co-evolution requires a fairly equal start and I shall argue later that the requirements for success had already become so ingrained by the time neuro-

[14](Pellionisz 1984) makes the point clearly that neuroscience has no clear paradigm since it is so new.

science began to make inroads into Cognitive Science, that it was bound to "fail" according to the standards of the existing paradigm. So, von Eckardt's co-evolutionary strategy is, at best, a disguised top-down situation which is unsatisfactory for the same reasons. Essentially the problem is this – and this is a major theme for this whole work: that co-evolution of areas with completely different notions as to what constitutes "evidence", "explanation" and "good theory" is not possible. In trying to co-evolve with such underlying problems, one view always falls foul of the other's methodological criteria. Part of what I will say is that the neuroscientific view is prone to fall foul of the traditional cognitivist's methodological dictates and if there should be any falling foul, it should be in the other direction.

Another full-scale attempt to outline the structure and direction of Cognitive Science is Sean O Nuallain's "The Search for Mind"[15]. The central claim of this work is that language use consists of exactly this sort of co-evolution of different levels of description:

> "... use of language involves exploitation of a formal symbol system, interaction of this system with operational knowledge, and intersubjective knowledge of oneself as an object in the world. Neuroscientific evidence currently exists for the first two points."[16]

It is clear that O Nuallain is concerned with the contributions of these levels of knowledge rather than a co-evolution of methodology but the idea is clearly similar: the symbolic should be seen to be a partner in a Cognitive Science that proposes to tackle the issue of language. Now, O Nuallain is very much aware of the crucial importance of the notion of context. He is quite clear that its essentially non-linguistic and non-formal nature guarantees that a purely formal account of language will fail[17]. However, there seems to be a quiet acceptance of the fact that the formal description of

---

[15](Nuallain 1995)
[16]Ibid p.175
[17]See especially Ibid, Chapter 3

language is compatible with this non-formal notion of context. If formal linguistics fails to capture the non-formal aspect of context, then, as is the thrust of my arguments below, the attempt to make them co-operate in a "hybrid" or "co-evolutionary" manner will fail also. We return to this theme properly in Chapter 6 where we investigate the inherent trouble in trying to blend levels of explanation that are, by design, complementary. An obvious point to raise here is O Nuallain's claim that neuroscientific data exists supporting the idea that a formal system plays a part in language use. Interestingly, after a thorough survey of neuroscience and work in PDP, the only really supportive material that is discussed in not really neuroscience but work such as Smolensky's PDP systems[18] which are explicitly designed to tackle the issue of a symbolic/non-symbolic interaction. So, no neuroscientific support is adduced as such and Smolensky is accurately reported as not allowing formal grammars any *causal* role anyway; something I wholeheartedly agree with and develop further below. At the very least, I note here that if one assigns no causal role to formal grammars, then one is not really seriously considering a hybrid explanation or methodological co-evolution at all.

O Nuallain is very concerned, unlike von Eckardt, to class as central the notions of "consciousness", "self" and "environment". I think this is a praiseworthy tendency indeed but I have severe reservations about Cognitive Science being able to do this. While I do not address these issues explicitly, this thesis might be taken as an argument why this cannot happen. In respect of O Nuallain's wishes, I will note here a problem I find with his approach that is exactly the problem I argue later is central to this whole debate. On p. 236[19], O Nuallain outlines the detail of his approach towards including these famously omitted notions mentioned above.

"Social factors can be handled by informational characterisa-

[18]Particularly (Smolensky 1988)
[19]Ibid

> tion of subject-environment relations (as done in, for example, situation semantics), affect by studying its informational role, and consciousness by examining projection of informational distinctions."

Firstly, I have little faith in situation semantics as being able to address these issues as I describe in Chapter 4: I think the new generation of "informational" semantics theories have given us little more than a vocabulary with which to fool ourselves that we really are incorporating "context" into our endeavours. In general, I think this quote depicts the mistake of trying to maintain a real difference between formalism and areas like "context" whilst also attempting to link them. McDowell's attempt at this is used as a descriptive case of this problem in Chapter 6. O Nuallain's approach is certainly a lot more mature and appealing than most in that he recognises the major omissions in Cognitive Science; the issue I take is with the means proposed to repair this and I think that these means are flawed in a way that runs as a common thread through attempts to provide a coherent subject matter and methodology for Cognitive Science and its attendant philosophy. The issue is illuminated when O Nuallain says:

> "The viewpoint of this book is that many of [the problematic semantic] arguments can be resolved with a clearer model of context."[20]

I would agree with this if it were not for the fact that "a clearer model" is, I will argue, inevitably elaborated upon in terms which concede everything to an essentially symbolic formalist picture which thereby contradicts the whole notion of a non-symbolic and pervasive idea of environmental context. Crucially,

> "... context relates to the interaction of the symbol system with other types of knowledge."[21]

---

[20]Ibid p. 262
[21]Ibid p. 327

I would have no quarrel with this notion of context were it not for the fact that I find good reason to suggest that it is exactly this notion of "interaction" that we fail to specify. We fail to do this not for any reasons to do with contingent lack of effort or ingenuity but rather because of how "context", "symbolic" and "formal" are defined in the first place.

2.3 Ineliminibility

To return to von Eckardt's treatment, the conflict of explanatory interests revealed above is displayed further in her argument for the ineliminibility of the information-processing level of description. She argues that in order to see this, one needs to see that the level of neural realisation cannot help in explaining that which Cognitive Science seeks to explain. The argument for this is simply one by definition:

> "How is this possible given that I have barred representations *per se* from the neural level?"[22]

Of course, the methodological demands of the cognitive information processing view require representations on von Eckardt's approach and so the neural level could not possibly meet this objection. The interesting thing that seems to have been missed though is the consequences of this rather extreme position. If the neural level is barred in principle, indeed by definition, from contributing towards that which is necessary for a good explanation in Cognitive Science, then what can really be said, except in the rather vague and hopeful current idiom of "interdisciplinary research", about the future of a "co-evolutionary" strategy? This is exactly the same problem as faces MacDowell that I discuss later in which he tries to keep approaches far enough apart to be still distinguishable but close enough

[22]Ibid p. 333

together to be intercommunicable. The tension is, I shall argue, too much for the subject to bear.

In general, we have a common situation in much foundational work in Cognitive Science and AI. With the early interest in, for example, chess playing programs, it was decided first what constituted the essence of the explanatory domain under question, and then research was encouraged under that methodological umbrella. von Eckardt is no different in that she has decided what is wanted from Cognitive Science in terms of explanation and then proceeds to judge the approaches on that basis. However, as the history of science shows, one does not always get what one wants in terms of explanation. Part of the enterprise of science is exactly to find out what sorts of explanation are appropriate rather than to explore amongst ideas for ones which fit the preconceived idea of relevance. No-one would suppose that the history of science is a paradigmatically rational enterprise in which there were no biases regarding explanatory relevance but Cognitive Science is in an unusually tender situation, being so young, where any crystallisation of explanatory desiderata could be quite fatal for a healthy growth of the subject. Once thing we should learn from science is that the form of an explanation need not be, and often is not, obviously related to the terms in which the phenomena under investigation are described. In this case, so-called "cognitive capacities" need not have an explanation that invokes a cognitive or information processing level of description. Of course, they *might* have to invoke such things but in order to do so, one needs a better argument than that which merely states that other possibilities are ruled out since they do not happen to use the right vocabulary. Who would, for example, have thought that heat would be explained by mean aggregate velocity of molecules or similar since the vocabulary is so different to that which is used to describe temperature!

What happens, I think, when one allows a diversity of explanational possibilities is not that the answers are necessarily more diverse but that

the *questions* are. Here is a crucial example that von Eckardt gives to elucidate her conclusion that "explaining intentionality exclusively at the neural level is a hopeless enterprise."[23]. von Eckardt imagines asking an imaginary neuroscientist who claims a solution to neuroscientific explanation in Cognitive Science

> "Tell me, Professor X, when your subject Joe images a small pine tree at an angle in the centre of his imagistic 'field', what is going on at the neural level that explains the intentionality of Joe's imaging?"[24]

Not surprisingly, the answer about movement from neural state N624 to N1009 is derided. If one asks a question about *intentionality* then, by definition on von Eckardt's account, there *is* no neural answer and so this example is really a peculiar exercise in setting up straw men that one has decided cannot exist by definition. The desirable explanatory categories of one research program are being used to ask questions of another which is at the very least, startlingly unfair. If one really *had* developed a neuroscientific answer as Professor X is supposed to have done, then the question asked would need to be different as "intentionality" is the central fulcrum of a completely different research program. One cannot move between radically different research programs without bringing along their methodological concerns[25] too and this is a radically underappreciated fact in Cognitive Science and one which lies in the way of all interdisciplinary and co-evolutionary approaches. Kuhn recognised this in his insightful observation that science, for progress by any realistic criteria, needs dogma. It needs to fundamentally restrict questions about, crucially, methodology.[26].

If the neural level contributes nothing to an explanation of intentionality, then what use is it even in tandem with something (information pro-

---

[23]Ibid p. 335
[24]Ibid p.335
[25]and vocabulary
[26]A classic statement of this is (Kuhn 1963)

cessing research etc.) that can? That is, if the neural level is so useless in this respect, what is the point of urging co-evolution? At the same time as the fear that, let in at all, the neural level would change the questions involved (which of course it would), the prevailing psycho-physical monism requires that lip-service be paid to physical and empirical research. Again, this is a manifestation of the irresolvable tension within the subject which forms the focus of this thesis.

Now, I might be criticised at this point for being unduly negative: I have outlined many problems without really suggesting what might help in fixing them. As O Nuallain says of philosophers of mind:

> "If they criticise a research program, they should propose a substitute one as rich in its place."[27]

In many cases, this is indeed true but, as a general principle and in the case of Cognitive Science, I must disagree. If one criticises a research program on foundational issues, it may well be that one's argument *is* that, even conceived broadly, the subject has boundaries that it cannot cross – limits that it cannot transcend. This is close to my view of Cognitive Science and I demand the right to not put anything "as rich" in its place if the meaning of "as rich" is largely dictated by the very subject I find fault with. Indeed, I think that this has a lot to do with the problem in Cognitive Science: the methodological dictates have dictated so successfully that one can hardly propose anything else without being accused of having missed out a lot of data i.e. one is accused of not being "rich" enough. This thesis is really about the very concept of "rich" and why requests to supply "rich" replacements and theories cannot and *should* not – on the criteria demanded by the subject – be fulfilled.

---

[27]Ibid p.32

Chapter 3

A Central Problem for Formal Semantics

The aim of this chapter is to outline, discuss and attack what I consider to be a central problem for the logico-formalist programme in natural language semantics. It is intended as a concrete introduction to the typical problems of what I will call a "stratified" theory. This requires an examination of the notions of meaning postulates and lexical decomposition: core features of formal natural language semantics. I shall argue that these two elements of formal semantics are the primary stumbling blocks to a plausible account of natural language and will go on to suggest why this is the case.

Formal semantics of natural language reached an impasse, familiar since Carnap, with the realisation that purely formal accounts were inadequate in dealing with certain constructions that seem to rely on the meaning of lexical items. For example, the most oft-quoted case is of analytic but not logically true statements such as

"All bachelors are unmarried"

As an analytic sentence, we would like it to be true by necessity, true in all possible worlds. However, its logical form dictates that, if it is true at all, its

truth is at least a *formal* contingency i.e. it just so happens that the extension of "bachelor" is a subset of the extension of "unmarried". Any necessity of truth is such by virtue of extra-syntactic factors. Further, this is not just a problem for traditionally analytic sentences. It is a general problem for any formal approach that posits lexical atomic constituents. For example, we cannot determine, by logical form, that "if you know something, then you believe it" is a good inference on formal grounds. We need something like a classical epistemology – knowledge is "justified, true belief" – to give us the necessary premise. The problem reaches its sharpest when you are forced to deal with inference that rests on properties of actions and objects: "If you are hit on the head with a hammer, you will be in pain". The general difficulty is that the atomic predicates and objects are represented by arbitrary and logically simple symbols, concealing the rich set of associations that seemingly allow us to perform these inferences. "All bachelors are unmarried" looks like "$\forall x Fx \rightarrow Gx$" and *that* is not a logical truth.

It is clear that this is a very important problem. It amounts to the inability to capture practically the entire semantic aspect of human languages. The relationships between concepts, the inferences based on these relationships and the knowledge we derive in virtue of said relationships constitute an almost all-engulfing lacuna in the formalist programme if there is no way to deal with this problem in a manner true to formalist principles given that their whole motivation is to account for human information processing as syntactic manipulation.

3.1 The Formalist Solutions

It is generally thought that Carnap was the first to worry about this sort of problem, although in a different context. He proposed what has become one of the standard methods of dealing with this difficulty: meaning postulates. These are formal constraints on the class of models that can be invoked to

account for certain constructions. For example, we might have a meaning postulate like the following to constrain our models to ones in which "All bachelors are unmarried" is necessarily true:

$$[\![bachelor]\!] \subseteq [\![unmarried]\!]$$

That is, a *meta*-theoretical constraint on the models we are allowed to use to account for the data. In more familiar terms, it narrows the range of possible worlds down to those obeying the criteria embodied in the meaning postulate.

An alternative approach that has received wide support within linguistics and formal semantics is the method of lexical decomposition. Here, instead of giving constraints on the possible models, one supposes that the lexical items one is dealing with are actually not atomic but rather complexes that disguise lower level internal structure which renders troublesome inferences formally valid. For example, a solution along these lines to the "All bachelors are married" problem might suppose that "bachelor" is decomposable into "unmarried man" and thus the sentence *really* says "All unmarried men are unmarried". This is fine as it is *logically* true and thus an account of understanding the analyticity of the sentence can be given along purely formal lines.

I now turn to an exposition of the supposed merits/demerits of these two opposing methods of dealing with this crucial problem, with specific attention to drawing out their common assumptions. These will form the basis of the criticisms to come in section 3.2.

Later on, I shall examine the techniques employed by an example of the more recent "information theoretic" approaches – Situation Theory – which claims to have overcome many of the limitations of classical approaches. It will be evident that the underlying stratified approach, even here, renders this new type of semantic theory ineffectual in tackling the shortcomings of its implicit assumptions.

3.1.1 Lexical Decomposition

Lexical decomposition tactics are motivated by a desire to directly bring aberrant data under the formalist's wing. If a particular construction is not formally treatable lexical decomposition provides a way to analyse it further until it is. Katz[1] explicitly likens the method of lexical decomposition to a finer level of granularity of analysis. The general idea is that, in terms of granularity, a decompositional approach stands to predicate logic as predicate logic stands to propositional logic[2]. This is an intriguing idea that derives much of its appeal from analogy with the manifest advantages of predicate over propositional logic. Unfortunately, it must be noted that Katz is glossing over an essential difference between the two relationships he wishes to draw analogies between. Predicates are part of sentences in a fairly straightforward way. They are syntactic components of sentences that have a number of independent motivations e.g. an account of properties or a way of speaking about classes. The components of a decomposition however, are *semantic* components: their role is to give us semantic links with other decomposed lexical items. "Unmarried" is not syntactically tokened within "bachelor". This makes the step from lexical items to their decompositions crucially unlike the step from sentences to predicates. We do not have a clear notion of what counts as a "correct" decomposition as we do not have a syntactic tokening to guide us. So, the discontinuity in the case of decomposition is a problem due to an absence of any motivation for considering decompositional elements as "parts" of a lexical item in a sense that the traditional structural relationship between predicate and sentence would support. The move from propositional to predicate logic is a kind of "decomposition" of propositions but is quite a harmless sort as we have many independent arguments in favour of the sub-propositional elements. The motivation for lexical decomposition is purely teleological: we need it

[1](Katz 1972)
[2](Katz 1972) p. 185

in order to be able to *do* certain things. The decomposition of lexical items is a decomposition of *meaning* which is altogether a more slippery thing to have to dissect. It is one thing to look at the parts of a *formal* complex but it is entirely another to look at the parts of a complex of *meanings*.

Nevertheless, the supporters of lexical decomposition may be justified in some way in their endeavours quite independently of scruples against justifying the approach by analogy with the well-accepted syntactic decomposition strategies in formal logic. The idea persists that a finer level of analysis of the data – another 'strata' in their explanation – will reveal the logical structure sufficient to render the problem inferences valid and the problem sentences intelligible. Lakoff's influential paper[3] "Linguistics and Natural Logic" embodies a detailed advocacy of the decompositional over the meaning postulate approach. As in most decompositional treatments, he begins by arguing that meaning postulates are *ad hoc*. We shall return to this point later. There are two main areas in which Lakoff thinks that meaning postulates are sorely lacking. The first is their seeming inability to capture some linguistic regularities that one would like to demand of one's models. For example, it is suggested that linguistic forms involving the lexical item "come" are related to the item "bring" along with a verb of causation in a fairly regular manner. For example:

1. bring = cause to come

2. bring about = cause to come about

3. bring up = cause to come up

Lakoff argues that you would need separate meaning postulates to cover each of the separate "come" constructions in 1–3 above and this would gloss over the regularity made apparent by the common containment of the element "bring". The source of this trouble, for Lakoff, is that meaning pos-

[3](Lakoff 1972)

tulates contain logical forms that "do not contain phonological shapes"[4]. Here we have, I think, a misunderstanding as to the nature of the solutions that are being proposed. Whether or not meaning postulates are sensitive to phonological regularity is hardly a principled objection. Meaning-postulates are created exactly to fit the troublesome data and so there is nothing in principle stopping us from continuing this for phonological regularity. In effect, we should just allow the logically atomic predicates in a postulate to be decomposable, thus exposing the regularity we need in order to effect treatment exactly in the way that Lakoff lauds the lexical decomposition approach for doing. One may object that this is rather perverse as it makes the meaning postulates dependent on a lexical decomposition but this is beside the point. Lakoff is mistaken that meaning postulates *cannot* account for the regularity.

This confusion about the status of the constructs designed to solve the problems is apparent in another of his arguments for the decompositional approach. Lakoff argues that only a decompositional picture can account for a principled ban on bizarre constructions that we would not want to license. As an example, he supposes there is a verb "accusate" where "x accusated y that S_1" is subject to the decomposition

"x said that S_1 and that y was guilty".

It can be demonstrated that a decompositional account of this can employ various coordinate structure constraints to ensure that "accusate" is not an allowable verb. Again, Lakoff argues that meaning postulates cannot do this as one could easily have a meaning postulate like

$$\text{accusate}(x, y, S_1) \equiv \text{say}(x, \text{and}(\text{innocent}(x)), \text{guilty}(y))$$

There are no restrictions and so constructions of this sort are not disallowed under the meaning postulate treatment. However, even Lakoff notes that

[4](Lakoff 1972) p. 610

> "The only way to keep the meaning postulate hypothesis from permitting such possible lexical items would be to impose something corresponding to Ross's coordinate structure constraint on meaning postulates."[5]

The appropriate response here is, I think, "why not?". After all, one has imposed such a constraint on the decomposition and not been tempted to use this fact as a rejection of that approach. No, the unstated source of Lakoff's worry here is a tacit acceptance of the division of the task into theoretical and meta-theoretical parts; a stratification of explanation. Lakoff believes that a decomposition takes advantage of the theoretical machinery of linguistics quite naturally whilst a meaning postulate would need to have such machinery translated into the meta-theory to take effect: meaning postulates are meta-theoretical constraints on possible models. This is an important point and, strangely, one which meaning postulate advocates have regimented to defend their position too. It will be addressed more fully below.

3.1.2 Meaning-Postulates

The motivation behind a meaning postulate view is that the way to capture relationships in lexical meaning is to restrict the class of possible models to those which respect our intuitions about such relationships. The proponents of this view regard lexical decomposition as *ad hoc* as it must posit a non-decomposable base at some point and the point at which this occurs is arbitrary. In addition, it is argued that we soon reach an analysis that is far from complete yet has no obvious decomposition to take it further. For example, a stock example of decomposition is that "kill" is defined as "cause to die". This does not help much as to analyse "Peter killed Paul $\rightarrow$ Paul died" as "Peter caused Paul to die $\rightarrow$ Paul died" does not render the former as a logical truth. It is not obvious at all how to proceed with

[5](Lakoff 1972) p. 614

a decomposition of "cause" that would render the inference formally valid. Here, the meaning postulate account seems to offer an advantage in that a meaning postulate need not pretend to be a decomposition into the "underlying meaning" of a lexical item: a practice that is dubious from the start. However, accompanying this is often a feeling that meaning postulates overcome the requirement that revealing logical form is all that is necessary to render certain inferences valid. A representative example of this is (Fodor, Garrett, Walker & Parkes 1980)[6] where it is argued that a decompositional approach *is* basically the assumption that logical form is all there is to validity – the suggestion being that a meaning postulate approach is not so limited. For example, it is said that a principle that would allow us the inference from "cause to die x" to "die x"

> "would have to be sensitive to the meaning of 'cause' ... and would thus have precisely the character of a meaning postulate."[7]

It is not clear to me here in what way a meaning postulate can be sensitive to a "meaning" where a decomposition cannot. A meaning postulate can put a restriction on a class of models and this seems to avoid the problem of having to posit an underlying "real" meaning. The argument runs: after decomposition, there are certain "residual inferences" that depend on meaning that are still unresolved. So, the decomposition of "kill" into "cause to die" leaves the inference from "cause to die" to "die" unresolved. The way that we choose to resolve this must be "sensitive" to the meaning of "cause" as it must be able to discriminate it from, for example, the inference from "wish to die" to "die". The mistake lies in talking of the end "after" decomposition. The motivation for a decomposition is the desire to render inferences formally valid; so, one which does not is simply not a complete decomposition. If you end up with "cause to die" as an analysis of

---

[6] p. 271

[7] (Fodor, Fodor & Garrett 1975) p. 526

"kill" and this is not enough to render you an inference to "die", then what is to prevent you from analysing "cause" in such a way as to guarantee the inference? Fodor *et al* assume that there is no "plausible" analysis of "cause" that would help. But as plausibility is secondary here to adequacy, I do not see this as a relevant objection. This is all inextricably bound up with methodological problems with the programme underlying both the decompositional and meaning postulate approaches and will be taken up in section 3.3.2.

There are still other reasons why we might be suspicious of the claims to "sensitivity" by proponents of the meaning postulate approach. We are aiming to give an account of the meaning relations between lexical items. This is the root of the whole problem as initially defined. Now, as noted above, the sort of relations we are interested in are not simply the relations between analytic *definiens* and *definiendum*. Take the following sentence:

Because the sun was blazing, Peter was hot.

A meaning postulate to deal with this would simply tell us about what models we can develop given that they contain "blazing" and "hot". It would tell us, for example, that everything is hot when the sun is blazing or similar. However, this is a very impoverished notion of the semantic link between "blazing" and "hot". There is an element of metaphor here: an "internal" link one might say, as opposed to the the purely "external" link that meaning postulates grant us in terms of coincidence of extensions in models. So, to say that meaning postulates are more "sensitive" to meanings is at the very least over-stating the case. Take the example

Because the sun was blazing, Peter was inside.

The connection between "blazing" and "inside" seems to be less metaphorical, less "internal" than the above. Perhaps one could plausibly have a meaning postulate to deal with this (although, really, I think not; see below) as the metaphorical relationship does not appear as strong in this

case. Anyway, whether deep, metaphorical, contingent or analytic; meaning postulates treat all meaning relations the same way.

Addressing the battle between decomposition and meaning postulates, (Fodor et al. 1975) raise the question of sentence comprehension and the startling speed with which it occurs. They seem to think that the mere fact that we manage to perform this task so rapidly, legislates against a view that divorces "semantic representation" from surface structure. The assumption is that such a divide would require more time than empirical results demonstrate is feasible. Lexical decomposition does indeed imply that the semantic representation of a sentence is something that a complete decomposition of any pseudo-lexical terms delivers as output. Therefore, a sort of "decoding" must take place before inference rules can begin to play a part. This rests on the assumption that all inference is, at base, purely formal. Fodor *et al* argue that we should prefer to keep the difference between the surface and semantic forms as small as possible so that we "reduce the load on processes that must be assumed to be performed on-line"[8]. I think this is a very much mistaken attitude, fundamentally the same underneath but superficially the opposite of Lakoff's view regarding the advantages of lexical decomposition. I shall expand on this point as I now turn to an examination of the commonalities that exist between the two approaches now outlined.

3.2 Common Assumptions and Problems

The central thing to note with respect to these two attacks on the problem of lexical meaning, is that they both argue for a level at which the 'real' work is done. Decompositional approaches posit a sub-lexical stratum containing the elements of meaning underlying the lexical items. Meaning-postulates suppose a rich meta-theoretical level of constraints that provide

[8] *Ibid* p. 526

restrictions on the allowable models of the lexicon. Broadly, the benefits are as follows: a decompositional approach can exploit syntactic structure constraints as it allows the breaking of the lexical items to occur at the same level as the operation of said constraints. A meaning postulate tack minimises the amount of time taken to process as it keeps the solution at the level of unanalysed sentence structure. If we envision the "levels of processing" picture that both of these models encourage, then we might say that they consider it an advantage to *localise* processing. By this, is meant the desire to keep the levels of representation one has to proceed through to a minimum. If we imagine that the model-theoretic constraints are "above" the lower level lexical decompositions in terms of depth of processing[9], then it is reasonably clear how the two approaches differ: one localises processing at the "top" end of the model and the other at the "bottom" end. Both, I think are mistaken.

The problem lies in the reification of the explanatory levels that the traditional accounts posit. There seems to be an assumption that the different levels of processing are independent and disjoint, thus you can gain real advantages by attending to the level at which you perform most of the processing. It is surely an appropriate attitude to be amazed at the speed of sentence comprehension but this needs to be informed by a realisation of the real-world embodiment of such comprehension. We are, I take it, supposing that the brain performs the necessary processing. Let us consider the argument that meaning postulates are more plausible as they prevent us from having to perform time-expensive decomposition "online". The solution is to capture the effect of a decomposition in the "inference module" by employing model constraints. If we reflect that we are, after all, assuming that the brain performs all of the functions necessary to implement this picture, we can see that this is simply an example of the inappropriateness of the model-theoretic approach. The choice is between

[9]See e.g. (Fodor et al. 1980) p. 273

decomposition — a process that manipulates forms of the object language and meaning postulates — a mechanism that guarantees the inference in the meta-language. However, as the brain must implement *all* of the operations we posit, there simply is no time-saving to be made by pushing the processing into the meta-language. There is simply no "off-line" to be taken advantage of. We are used to thinking this way because this is exactly we what we can do with formal languages: we can easily reduce complexity in the object language if we are happy about making the meta-language more complex. Notice, though, this division is not something that transfers well to real-world systems. A computer programmed to perform inference has no processor independent meta-language governing its operation: it must perform "meta" operations too. When it comes to implementation, a meta-theory necessarily comes part and parcel of a theory; thus, it is misleading to suggest that time savings are to be made by reducing the operations in the object-language. The real system must implement the meta-language too and thus pay all time costs. Take, for example, the following inference:

$$P \wedge Q \vdash R \wedge T$$

Suppose that a decomposition reveals that Q has the internal structure $P \rightarrow S$. Suppose further that S has internal structure $R \wedge T$. These two decompositions were required before it became obvious that the inference was valid. To "save time" we might like to say that, instead, we had the following inferential rules:

$$\frac{Q \vdash P \rightarrow S}{S \vdash R \wedge T}$$

In this case, do we not manage the inference in two less steps as we miss out the "on-line decomposition"? The answer must surely be negative. Given that the inference rules are implemented by the same system that implements the steps mediated by such rules, you gain in one area and lose in another. You gain time leaving out the explicit decomposition yet

lose it in having to search through more rules to find one to apply. Not only this, by pushing the processing into the rules, you very quickly develop very strange logics with many, many rules. Imagine this machine implemented. The search time and the handling of the model-theory to deal with a logic with so many non-standard rules would equal the time "saved" in performing less inference steps.

This, I think, is a simple consequence of deductive equivalence. This ensures that there is a link between $P \vdash Q$ and $\vdash (P \rightarrow Q)$. When the model of deduction and the deduction itself inhabits the same physical system, the link is, I think, also one of processing speed. Given the deductive equivalence, there can be no relative time savings, for the human conceived as logic processor, between any formalist theories. If you do not have to pass through as many inferential steps, you have to pass through more meaning postulates. Fodor *et al* seem to recognise this fundamental link but fail to grasp its significance. They say that changing the notion of the lexical form requires

> "... compensating adjustments of the inferential system ..."[10]

The essential problem with this is that the "compensating adjustments" are enough to make no difference overall for a real, embodied system. The brain does not care at what abstract "level" one might care to think of it operating at, it still has to *do* everything. We should *still* find the speed of sentence comprehension surprising and thus we *still* need a theory to address this.

A further objection to this traditional formalist enterprise is that it gets the whole problem of meaning backwards. Both decomposition and meaning postulates get their motivation from particular problematic inferential relationships. They are a tactic we may use when we encounter a troublesome case. This assumes that the picture we converge upon will be

[10] *(Fodor et al. 1975) p. 525*

derivable from the inferential relations our surface language sanctions: it is these that throw up the problems we create decompositions and postulates to solve. We have no reason whatever to suppose such a picture is so derivable however. Indeed, we have a good deal of evidence to suppose that such a picture is *not* derivable in *any* way that depends on taking intuition and language as its evidence. As Dreyfus[11] and Winograd have previously urged, this was exactly the program undertaken by Husserl and Hegel: which they both declared to be intractable after many years of detailed study. The attempt to specify meaning relationships by either a decomposition or a restricting postulate is precisely the attempt to specify the relations between every concept we employ in an explicit manner: by enumerating all of the links by either embodying them all in postulates or instructions to decompose. The set of meaning postulates or decompositions will be a specification of all meaning relationships. So, we are considering something of the size of the cartesian product of the set of possible concepts. In traditional formal approaches, we have to do this in order to allow the logical paradigm to continue: if we do not specify what relations a lexical item bears to others in some way appropriate to logic, we cannot proceed formally. It seems very unlikely that something as large as the cartesian product of the set of concepts we employ is a realistic thing to suppose the brain entertains. The reason for the enormity of information is that traditional formalism requires that this information be explicit in either a translation into simpler terms or as a rule restricting inference, both of which are something the brain has to *do*. It will be objected that there will actually be a significant reduction in complexity as we come to determine the *underlying* features of our concepts. It is thought indeed that we will be able to determine common elements, as in the case of "bring" above (section 3.1.1), that will reduce the complexity of meaning interrelationships to some plausible level. This hope is, in fact, a central aspiration for tra-

[11] For example, see (Dreyfus & Dreyfus 1988)

ditional formal semantics. At the very least, it requires that we show why early phenomenological programmes failed to accomplish this. I think it an unsatisfactory approach, nevertheless, to develop a notion of the underlying meaning relationships out of solutions to many isolated and technical problems for reasons to be addressed in section 3.3.2.

3.3 "Information" Based Approaches: Situation Theory

The early 80s saw the emergence of "Situation Theory" amidst a desire for generalising linguistic semantics to incorporate much of the significant work done by analytic philosophers such as Searle and Austin. Situation Theory, given its first substantial exposition in (Barwise & Perry 1983), was an attempt to incorporate notions of partiality and perspective into semantics. It was an attempt to realise the importance of *context* in a formal setting. As such, it was an attempt to circumscribe semantic relevance in things called "situations" and thus to elucidate meaning relations. It is crucial to consider then, how the Situation Theoretic approach fails to deal with the problem outlined above. I think it falls foul of the same unjustified assumptions regarding implementation-level independence that besets the other views so far presented.

The basic elements in a Situation Theoretic approach are "situations". They are that which delineates the sphere of the epistemically relevant for a given semantic task. This is a modern version of the well-known AI tactic of model circumscription and restricted quantification. It is a little unclear what a situation actually *is* however. A definition of *situation* would be a start in helping to make a classical treatment of inference manageable by restricting the domain. Traditional ways of doing this are troubled by the problem of having to *a priori* determine the domain restrictions and end

up doing it in such a way as to guarantee the inferences required. For example, in the sentence "It was a sunny day and Jane thought everything looked beautiful", the quantification is obviously not over *everything* in the world but only a small section. The problem is to say what. It is not merely things within Jane's view since buses and drains may well be part of the scene but Jane is not really saying anything about *them*. The sentence has overtones of speaking about biota but how does one know this? "Context" is the usual answer but, as we have seen, it is unclear how to express this "context" as it is not clear what should be in it or, far more importantly, how to conveniently and preferably algorithmically specify what is in it at a given time. Specifying the context *a priori* is about all we can do in the face of this. We say certain facts, usually those that will help as premises in inferences we are interested in, are "part of the context" and thus need not actually take part in the inference as such. Situations are meant to limit the domain to manageable sets we can quantify over and to allow us a method of expressing "presupposed" context. This is the implementation-level confusion. So what is a situation? Cooper tells us that "Situations are thought of as situation-theoretic objects ..."[12] which strikes at first blush as being a mite circular. More promisingly, we find that "It is standardly assumed that situations are identified by the infons they support ..."[13]. However, given that infons are a superset of "facts": the "possible facts"[14], then we are perplexed to learn that "Facts can ... be thought of as invariants across situations"[15] and we are back where we started. Infons are "basic units of information"[16] but are composed of a relation, an argument list and a polarity. Relations are famously taken as basic in Situation Theory and so this does not give us much of a clue as to what an infon is.

---

[12](Cooper 1988) p. 50
[13](Cooper 1991) p. 13
[14](Cooper 1991) p. 9, (Barwise 1989) ps. 205,264
[15](Cooper 1988) p. 56
[16](Cooper 1991) p. 9, (Barwise 1989) p. 205

Leaving these basic definitional problems aside, what are the supposed *roles* of the Situation Theoretic posits? Situations are *partial* models. They settle only a fraction of the semantic questions about the world. This is how they delimit the domain of enquiry. They are motivated by a purposefully naïve realism derived from Austin and Gibson which I reconstruct in the following way. Possible word semantics provided a spur to Situation Theory to move away from the "unanalysed blobs"[17] posited by said theory. It was felt that they settled *too much* for a given agent: a world was a model of all the facts it contained. This was essentially a crisis of epistemology with Situation Theory declaring an omission of fallibilism in semantics. A central feature was held to be the *partiality* of semantic context: the semantic model an agent uses (for Situation Theory is still model-theoretic) was felt to settle only a narrow range of facts and not all of them, as possible worlds demand. Hence the terminology "situation" was adopted to indicate that the *size* of the model was considerably smaller than a typical possible worlds model. Great pains were taken to distance this view from one which simply rendered situations as "small worlds"[18]. Now, this partiality was partiality in the situations themselves and not in the perception an agent might have of a situation. The latter view is something possible world semantics has to employ to model fallibilism given the nature of such a model. So here is the glue that binds the disparate assumptions of Situation Theory together ... in order to accommodate naïve realism *and* fallibilism, the objects perceived have to be partial. The situations are themselves partial. One naïvely perceives partial things rather than partially perceives all-encompassing objects. Indeed, Barwise and Perry[19] tell us that "Reality consists of situations, individuals having properties and standing in relations at various spatio-temporal locations". This is a

[17](Cooper 1988) p. 54
[18](Barwise 1989) p. 79
[19](Barwise & Perry 1983)

move to be appreciated for its technical leverage rather than its philosophical motivations[20], since it merely stipulates as basic certain things which would be problematic. The more you assume, the more you can use your assumptions to do but then the entirety of your system is subject to doubt when the assumptions come to look suspicious.

The overall effect of this approach is to push epistemology out of the account by postulating simple access between meaning and reality. The fallibilism that proved the undoing of classical AI in the 60s and 70s is accounted for by partiality in reality. This neglect of epistemics is, I think, a serious flaw in the Situation Theoretic approach. Since Kant there has been a growing realisation that concepts central to semantics may well be more determined by our internal structure than the real world. More recently, neuroscience has been demonstrating how the kinds of categories posited by Kant could arise in the neural circuitry of the brain. A naïve realism is, I take it, not something that a serious account of semantics can abide if it is to be philosophically or empirically adequate. Neither can it claim to have relevance to an *explanation* of the semantics of human communication. See Chapter 5 for a discussion of this point.

3.3.1 Reappearance of the Central Problem

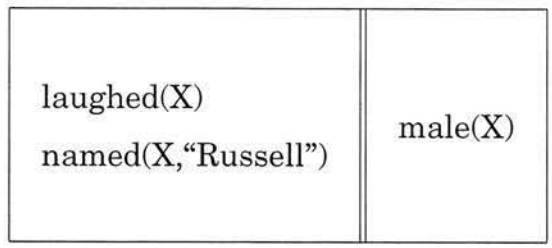
In Situation Theory, relations are taken as basic. This concurs with the realism of the theory but allows the central problem for traditional semantic theory, discussed above, to manifest. As relations are basic, meaning relations between them must be established in an extensional manner. Lexical decomposition is forbidden as relations are, *ex hypothesi*, basic. The connections are accomplished by what are termed in the literature "constraints". These are conditionals that hold between types of situation. So one might have something like "If F holds in a situation of type T then F' holds in a

[20]Terry Winograd also makes this point well in (Winograd 1985)

situation of type T'". In concrete examples, these constraints look familiar "If X is an unmarried man in a situation of type T then T' is a type of situation in which X is a bachelor". Given suitable anchoring of parameters, situation types of type T are usually said to contain those of type T'. It should be clear that these constraints are meaning postulates under a different name. They ensure a link between basic relations in the same manner as meaning postulates do for relations and predicates of traditional logic and thus fall foul of the same objections. Inference in Situation Theory takes place by means of constraints but proponents of the view are keen to point out that the type of inference is not merely "formal" in the sense that has been seem to damn traditional logics. Barwise argues that the "situated inference" espoused by Situation Theory is not purely "formal" in that it allows a dependency on "embedding circumstances"[21]. That is, he thinks it is contextually sensitive. If this were so, things would be rosier for Situation Theory as it could exploit context to avoid having to make explicit deductive inferences (whether supposedly in the meta-language or not). However, this "context sensitivity" is delivered in the usual ways by "exploiting environmental constants" or "making implicit parameters explicit". Naturally, this involves , in practise, *a priori* specification of parameters, tacit constants etc. thus reducing "situated inference" to traditional formal inference with a few more *a priori* specified premises. Context sensitivity turns out to be a chimera within all logico-formal approaches to semantics as the only way to achieve it is to explicitly and formally specify the context and that is a task that is implausible as an element of theory realistic to human information processing. Certainly, *if* you can specify context, then you can have clean formal inference. However, the antecedent of this conditional is the crux of the whole of Cognitive Science to some extent. It is the old problem of AI under yet another different guise. In recent years, Situation Theory has developed a notion of "presupposition" which is meant to

[21](Barwise 1989) p. 146

do justice to the promises of "situated inference". In the current graphical idiom, we may express "Russell laughed" as



where the material to the right of the double vertical lines is said to be "presupposed" as part of the context. Usually, anything that would guarantee an inference you would like to make is put here. Context sensitivity is reduced to putting more formal sentences to the right of two vertical lines in the belief that this renders them part of the "context". Presumably, part of the desire to do so is that elements of the "context" are not really "there" and thus explicit inferential steps are not needed to accommodate their import in a particular case. This, again, is exactly the mistake that Fodor makes with respect to meaning postulates. The suggestions of "different levels" of information and therefore processing are quite strong in graphical notations such as the above but are none the better off implementation-wise. Simply put, you might call a formal sentence anything you like – "context" is a favourite. However, if it plays a role in determining the outcome of an inference, the steps which use it must be traversed and thus time spent. As a result of these observations regarding Situation Theory, I am forced to view it as varying little from the basic assumptions of traditional symbolic formal approaches to semantics, thus it suffers the same problems.

3.3.2 Methodological Concerns

The idea that traditional formalist approaches to the problems of meaning-relation and inference constitute an *explanation* of human semantic dependencies seems strange as they are motivated by a desire to accommodate

surface manifestations of such dependencies. It is assumed that the way we speak about inference gives us real evidence for the way we actually do it. As a result one ought to be concerned about the methodological aspects of the traditional approach. In an area that aspires to explanation, we must aim to say how things are in a way that does not merely describe the surface effects. To say that we are allowed the inference from "bachelor" to "unmarried" because we have a decomposition of "bachelor" seems to be little more than a redescription of the fact that we accept the inference because we can *always* generate a decomposition for every case. We may argue about which is the simplest decomposition etc. but we can never *fail*. In general, if you have a inference you would like to capture from P to Q but which is not logically valid, simply provide a decomposition by stating P in terms of Q. How could we ever fail to do this? Similarly, there are no real constraints on the generation of meaning postulates. If we need one to secure an inference, we simply make one. The trouble at the root of this is that formal considerations alone (consistency, completeness etc.) are not restrictive enough on model-generation in the traditional programme. As Richard Gregory puts it "[But] it is all too easy to postulate things having just the right properties"[22]. The explanation we are afforded as to why there is a semantic link between two lexical items and why we permit an inference between them is that they are linked in a certain way (postulate or decomposition) and this guarantees the inference. However one does not explain an event by simply incorporating a prediction of it into ones theory. Whatever explanation is, it requires something a little deeper than the regimenting of an event into current knowledge. Enterprises that can never fail are worrisome because they make us suspicious that they are irrelevant to the real world: reality has no bearing on them.

Decomposition and meaning postulates are designed to *cover* the data. They work by regimenting an anomaly into the class of accounted-for cases.

[22](Gregory 1966) p. 10

Thus, to borrow a distinction from Quine[23], these approaches are designed to *fit* the behaviour of human speakers. There is, however, no reason to suppose that, even if we *could* have a model that fitted the totality of meaning relationships in a natural language, that this would be an *explanation* of what guides humans in their linguistic habits. As Quine says, "Fitting is a matter of true description; guiding is a matter of cause and effect."[24]. Decomposition and meaning postulates are designed to render an iteratively more accurate description of the class of allowable inferences and meaning relations. That the formal outcome is a model of what *causes* us to infer as we do is neither a logical nor plausible consequence of its descriptive adequacy. How do we infer and impute meaning relationships between supposedly lexical items? Because there are rules/decompositions that license this. How do we know there are such rules? Because these rules can account for the data. Of course, many sets of rules can account for the data so how to choose between them? We ask people what they consider the correct decomposition/constraint is. Here, as Quine notes for linguistics in general, we have argued in a full circle. We start by asking why people e.g. allow an inference from "bachelor" to "unmarried" and end by concluding that the reason is that they consider it reasonable to do so, which is just to say that they *do* it.

The weak link in the chain is the traditional linguistic notion of "evidence". Unlike evidence in the natural sciences, the sort of thing taken in linguistics contains an essentially *modal* ingredient. We ask native speakers, in effect, "Can you *possibly* make sense of this utterance?" or "*Could* you make out three scopal readings for this particular quantified sentence?". A supposedly empirical basis that has a modal aspect is worrying because it stands or falls with the susceptibility of the evidence to the ravages of personal, cultural and temporal idiosyncrasy. In short, it is evi-

[23](Quine 1972)
[24](Quine 1972) p. 442

dence shot through with a relativism that severely damages its evidential import.

What is required is a *general* mechanism that can serve as an explanation of the tendency towards lexical meaning relationships in natural languages. One that does not simply *translate* the relationships into the logical machinery and thereby claim to have accounted for the phenomenon in general. This might also save us from the embarrassment of having to decide between extensionally equivalent sets of decompositions/postulates by begging the question from the population whose behaviour we are trying to explain.

3.4 Themes to be Addressed

There are, then, two sorts of objections to the leading theories of lexical meaning. The first is an objection in practice: the logical apparatus misleads us into solutions that are heavily dependent on assumptions concerning orthogonality of processing tasks between the logical and meta-logical machinery. This is, as I shall demonstrate, the essence of the stratification problem. These assumptions are unfounded when implemented in a single, real system. The second objection is a scruple against the defining of the notion of meaning relationship in a piecemeal fashion by taking the superset of all specific solutions to individual problems of lexical meaning relation. Given that meaning relationships between lexical items are ubiquitous, the following idea suggests itself. It has been very fruitful in the past to incorporate ubiquity of effect into a *structural* treatment of the effect. For example, the ubiquity of gravity in its interactions with all objects motivated Einstein to reduce gravity to a warping in the *structure* of spacetime. This is in direct contrast to the old models that treat gravity as a force acting within spacetime: a separate thing to be described by separate laws. The beauty of this is twofold; it simplifies and satisfies

our desire for explanations that do not simply collect all specific solutions together and it provides a way to address worries about real world implementation and processing speed. This seems to me to be exactly what we require in our present case. I propose that we should treat meaning relationships as features of the structure of a *semantic space*, thereby reducing them to an analogical status as that of gravity in general relativity. This prevents us from worry about the speed of sentence comprehension and inference as the relations we would otherwise need explicit logical steps for become structural features which do not need explicitly inferring. A model of this sort also provides a general underlying account of meaning relationship that seems to fit our expectations of what an "explanation" should look like. This is further explained in Chapter 5 and is fundamental to all that is to follow. As an example of what is intended, if we have some eggs in a box with the lid closed, we can say little about the relations between their positions ... the space they inhabit is fairly orthogonal[25] However, put them in an egg-box and the situation is very different. We know automatically *from the structure of the space they now inhabit* certain things about them. For example, we know that none of them lie on their sides. We know that the distance relations between eggs obey transitivity as egg-boxes are made to be all alike (so they stack well) and this defines a metric on the "egg-box space". None of these "facts" have been derived by any sort of inference: they are products of the structure of the space the eggs now inhabit. So it is, I think, with concepts. Our concepts inhabit a highly structured and, up to a point, malleable semantic space that defines how we see relations between lexical items. Just as the egg-box defines how we see the relationship between egg positions. Malleable "up to a point" as an arbitrarily malleable semantic space would contradict the differences between species. Having certain constraints on the limits of ones conceptual malleability is a defining feature of what it is to see the world from the per-

[25]That is, movement in one direction does not entail movement in another.

spective of a certain type of creature. This sort of view removes emphasis from the individual meaning relationships between lexical items and such and puts semantic significance squarely in the lap of the semantic space that embodies the relationships.

There are some rather radical consequences of this position which I shall mention in later chapters. It embodies a move towards a Cognitive Science more concerned with neuroscience as the only hope of providing information regarding the point where we are now led to believe the crux of the matter lies. Now let us look in more detail at an instance of a more abstract example of the stratification methodology. After this, we shall build up to Chapter 6 after we have undertaken an examination of some crucially relevant recent considerations arising from neuroscience. In Chapter 6, we shall encounter a very general account of the stratification viewpoint and this will enable us to see the issues stripped to the bone and apparent in their identity with the problems in formal natural language semantics and Cognitive Science.

Chapter 4

Compositionality and Inference

Compositionality is very much a central feature of much work in natural language semantics. Its inception was famously occasioned when Frege wished to reduce the truth-values of linguistic units to functions of the truth-value of their parts. Explicitly a normative expedient historically developed along with certain formal languages, those wishing to investigate the semantics of natural languages adopted it along with the rest of the formal tools; true indeed that:

> "In fact compositionality is so basic a starting point for the logical way of doing semantics that in logic proper it almost always goes unnoticed."[1]

It is useful, then, to cast an eye back over the history of semantics in order to determine the intentions of the originators of such a "basic starting point". It may seem obvious to the current generation of formalists that such tools were intended for the use to which they are now put but a look at history does not bear this out. Frege, the most abused progenitor in this respect with talk of modern "Fregean compositionality" and the like as applied to natural language, believed that:

> "Language is not governed by logical laws in such a way that

[1](Gamut 1991)

> mere adherence to grammar would guarantee the formal correctness of thought processes."[2]

This is largely in direct contrast to the Montagovian tradition that supposedly followed directly from many of Frege's principles. Frege intended his formalism and the semantics thereof to be characteristic of the thoughts underlying language and not of language itself. Indeed, he explicitly states that language gets in the way of semantics.

> "I am not in the happy position here of a mineralogist who shows his hearers a mountain crystal. I cannot put a thought in the hands of my readers with the request that they should minutely examine it from all sides. I have to content myself with presenting the reader with a thought, in itself immaterial, dressed up in sensible linguistic form. The metaphorical aspect of language presents difficulties. The sensible always breaks in and makes expression metaphorical and so improper. So a battle with language takes place and I am compelled to occupy myself with language although it is not my proper concern here."[3]

Another rather historically maligned character is Tarski. He was explicit in his rejection of the application of his theories to natural languages. This does not, however, prevent much talk of "Tarskian" model theory in semantics today.

> "At the present time the only languages with a specified structure are the formalised languages of various systems of deductive logic, possibly enriched by the introduction of certain non-logical terms. However, the field of application of these languages is rather comprehensive; we are able, theoretically, to develop in them various branches of science, for instance, mathematics and theoretical physics."[4]

This looks quite optimistic about the role logic can play, but only as far as modelling already very formal languages within the logic. It does not extend as far as using them as models of any *natural* language. Indeed, he explicitly states that (italics his):

---

[2](Frege 1882)
[3](Frege 1918)
[4](Tarski 1944)

> *"The problem of the definition of truth obtains a precise meaning and can be solved in a rigorous way only for those languages whose structure has been exactly specified.* For other languages — thus, for all natural, "spoken" languages — the meaning of the problem is more or less vague, and it's solution can only have an approximate character."[5]

Tarski was an advocate of *replacing* natural language by approximate formal languages rather than extending the formal towards the natural:

> "Whoever wishes, in spite of all difficulties, to pursue the semantics of colloquial language with the help of exact methods will be driven first to undertake the thankless task of a reform of this language ... It may however be doubted whether the language of everyday life, after being 'rationalised' in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalised languages."[6]

Forgetting this leads to problems:

> "... the concept of truth (as well as other semantical concepts) when applied to colloquial language in conjunction with the normal laws of logic leads inevitably to confusions and contradictions."[7]

Thus, difficulties in applying certain formal theories to natural languages are hardly something we might encounter without, as it were, warning. As regards the concept of compositionality, this "Frege's principle" has a somewhat murky history. It is not terribly clear that we can derive the principle from Frege without some distortion[8] and the clues that are present (particularly in (Frege 1892)) are more concerned with creating new, canonical languages rather than fitting old, natural ones.

One might be tempted to say that such historical evidence is of secondary importance to the use to which the concept is put today; the inventors of the idea might have been wrong about the scope of its application.

---

[5](Tarski 1956)
[6](Tarski 1931) in (Tarski 1956)
[7](Tarski 1931) in (Tarski 1956)
[8](Janssen 1983)

This seems to be the sort of line that Davidson takes in his classic (Davidson 1967). This is a reasonable point but the lesson of the above is merely to prevent us from using arguments from intellectual authority in justifying a principle as "Fregean" or "Tarskian" etc. One finds cold comfort there. More important certainly are arguments that seek to show inadequacies and confusions in the use of the concepts today. It is to this that we turn after a brief look at some considerations to do with the application of prescriptive formalism to natural languages arising from the comments above.

4.1 Prescription and Description

The basics of the formal logics that are employed even today and which are said to be radically new approaches[9] are fundamentally derived in spirit from the work of the aforementioned pioneers. These systems, designed using concepts desirable in a canonical language, have come to be applied to a natural one. The result is that the formal systems are capable of modification and extension in various ways, constrained only by specific things such as consistency and certain meta-logical properties such as completeness. These sorts of considerations are, however, not severe enough when attempting to explain a natural language. They allow for an almost unlimited variety in and unprincipled extension of a given formalism. The goal for formal semantics and linguistics is to cover natural constructions with the formal system. I argued in Chapter 3 that we can never *fail* to do this when the constraints on our formal systems are meant to be constraints on the interrelation of the concepts of the system and not constraints on how well they apply to their ostensible subject matter. Matters of consistency are not enough to ensure that the formal approach is doing anything at all. The only factors involved in accounting for a certain natural language

[9]See for an example, the discussion of Situation Theory in Chapter 3.

construction are whether one *can* find a formal construction to do the work – this is always possible as there is simply nothing to stop us inventing something with exactly the right properties – whether it coheres formally with the rest of our system – again, we can simply modify everything so that this is the case since our formalisms were designed to be flexible – and whether we find the solution "plausible". This latter is such a weak requirement as to be of no use at all. Plausibility in formal semantics is closely linked to being simply able to do something in a formalism and this turns on the too liberal constraints of formal systems in general. For example, a common criticism of Montague semantics was that it all became horribly complicated as if this were a reason, on its own, to doubt anything. So, this, I think, is a very important and often overlooked point. A *normative* formalism is designed to force data to behave itself. This is all very well but one cannot then find well-behaved patterns in data presented from within a formal model and claim that these patterns indicate something about the causes of the patterns; the nature of our cognition, language and thought. I shall turn, later on, to a more basic and important reason why the patterns in our manifest outputs are not a good guide for theorising in Cognitive Science. This consideration under examination presently is more to do with systems that put a heavy emphasis on the actual formalism used rather then the subject matter itself. There are varying degrees of emphasis in Cognitive Science on the nature of the evidence and the identification of systems to present the evidence. For some, the systems that *present* the evidence, particularly in the case of linguistics, are so tied up with the notion of evidence itself that the contribution of the formal systems to the notion of the evidence is severely blurred and misunderstood. It is exactly in this sort of case that the problem described above occurs. If one is dependent – as seems inevitable in some degree – on a certain formal system to even begin to *describe* the patterns in the evidential data, then one had better be sure that the patterns are not a *result* of the

formalism. One had also better be sure that the formalism does not have certain prescriptions about data regularity built into it due to its original intended use as this will almost certainly lead to spurious patterns and the impossibility of ever appearing to be mistaken about whether or not one should be taking a formal approach in the first place. A good example of this is the lambda calculus. Designed as an abstract method of function specification, it was adopted by many semanticists as a device for dealing with problems of well-formedness during the composition of complex sentences in formal natural language analysis. This helps us to be able to give a completely compositional treatment of, for example, sentences containing definite articles like "the". We stipulate that a certain lambda expression binds otherwise unbound variables that would render any representation for such a word as ill-formed. This is, I think, a purely formal expedient; we do it because we can and because it makes the formalism work. However, this gives us a certain pattern in our formalisms: they are more compositional and thus we reify this and come to see language as compositional as a result. The formalism, motivated by orthogonal concerns, has "given" us a pattern. It has given us a something we may be fooled into thinking indicates something essential about our cognition. Thus are the consequences of the liberal constraints on formal theories.

4.2 Contemporary Compositionality

Compositionality has severe pragmatic implications. If one is to make a property of a construction dependent on properties of its constitutive parts, then any use of the construction in a computational role – inference, comprehension or whatever – requires a decomposition into those parts to obtain the necessary data. In model theory, the truth of a sentence requires an examination of smaller and smaller units of the sentence until the "atomic" properties are reached and certain meta-relations are estab-

lished between them. This is a central feature of all model theory, the mechanical aspect of this process being part of its appeal. However, this puts a heavy burden on the pragmatics of such theories when we consider the amount of computation involved. The clarity and formal attractiveness of compositionality derives largely from the simplicity of its formal instantiation; coupled with recursion, the power of this simple idea is obvious. The basic formal process of composition is simple but as a result, one has to do a lot of it to complete the task. Recursive theories are often like this; they will scale to more complex data but at the cost of much computation. The simplicity of the operation also requires well-defined ingredients. One must specify domains, extensions, sets and the like in advance and these must be adequate to the task in hand. Famously in the history of Artificial Intelligence, the computational explosion resulting in the searching necessary and the size of the sets of atoms for such operations demonstrate the computational problems involved and point to deeper problems with the conception of the basic operations. It is, however, rather another factor that seems to cast doubt on the traditionally supposed ubiquity of compositionality and its role in non-formal languages that we take up.

Natural spoken language is littered – smothered rather – in platitudes, cliches and banalities. Much of everyday vocal intercourse is glued together by such ejaculations. There are certain phrases that are so common or colloquial that they are hardly regarded by speakers as part of the essential meaning or import of their speech. "How are you?", "Very sensible", "Never in a month of Sundays" are all phrases suggesting a certain tone or general opinion but not at all determined by the particularities of the words that comprise them. We know rather what is meant *by* them rather than what they mean. There are also phrases in which component words have very different meanings compared to those found in more "ordinary" usage. The word "away" in "put it away" has a rather strange sense. "Get out of here" has "get" in a role incongruous with much of its more obvious uses. These

sorts of phenomena – and it should not be underestimated just how prevalent they are – are not something that sits well with compositionality. The problem is that one must have separate and rather different atomic definitions of the features of most words in order to accommodate their roles in these constructions. This adds even more overhead to the pragmatic issue of computation in real time. The common response to this is simply that it is indeed true that much of language is idiomatic and we would not hope for a theory without exceptions. This depends on the point of view we take. The classical post-Chomskian theorist sees rules with exceptions or rules with extra clauses. The post-Husserlian theorist sees a theoretically heterogeneous mass with some vague regularities. Compositionality is a result of a logical model that was designed to have certain features and it is not easy to see how it can be divorced from those features, the most obvious of which is regularity of form. Formal logic is, eponymously, concerned with form; compositionality as an idea and formal process was designed hand in hand with this and thus it is not easy to see how one can maintain that merely *some* of human natural language and thought is compositional. Thus the sorts of non-technical cases mentioned above which do not seem to fit such a picture result in two difficulties. Firstly, divorcing compositionality from formal regularity of the data is theoretically objectionable and secondly, attempting to maintain the regularity of data requires multiple definitions of the "atoms" of the compositional framework and thus much more computation. For example, to maintain the semantic outline of the verb "get" in a formal system would generally require multiple definitions in order to accommodate cases like "get out of here". The semantics of this special form of "get" would need to, for instance, be something like the semantics for "move". This sort of duplication of semantics for lexical entries is the sort thing that one must do if engaged in a formalist program since the different uses of the word are associated with different types of other words and it is these patterns of relations that *define* the semantics of a

word in an extensional system. Thus one needs as many different specifications of semantics for a word as it has different combinations with other types of words in the language. This means more searching for possible lexical entries to take part in inference. Which "get" does one need in a given situation?

The second problem has been addressed considerably before, notably in terms of the "Frame Problem" and the explicit phenomenology needed to accommodate such a picture[10] The theoretical problem is less considered and is of importance in forcing modern day formalists into the position where arguments like Dreyfus' take effect. A formal theory which had a piecemeal approach to compositionality would be strange simply because it would be, in effect, holding that a formally defined quality X is a function of elements, some of which were not addressed by the formalism and *this* is to say that X is not defined by the formalism either. So a formal theory of X that aims to define X without being thoroughly compositional is not a formal theory of X. Thus, one wants to be thoroughly compositional and this leads into the sorts of problems with computational complexity that have been famously intractable in the history of semantics and Artificial Intelligence.

We are able to understand everyday, colloquial and casual language very rapidly. Technical discussions and complex prose require more time, rereadings and going back over sentences. If one were decomposing in real time; if compositionality governed our actual practise with language, why would this be? It is not simply a factor of length and complexity; that possible reply is insufficient as if the *words* are unfamiliar but the sentence no more complex, we still take longer to reason with, understand and compare. Familiarity and ubiquity of a phrase are far more important in our ease and speed of use; those who use technical and complex language enough learn to attenuate the initial difficulties. The question arises, then,

[10](Dreyfus 1992) is the updated version of one of the classic works in this area.

what is it that allows a seeming bypass of decomposition?

There are analogies – perhaps more than mere analogies given the propensity of evolution to reuse existing structures – for exactly this pattern. Physical skills "crystallise" when practised enough. A good sportsman will come to perform "automatically". A feature of this is that the movements that are performed in the practise of the sport are done so in a different way that a similar movement might be done outside the context of the sport. The movements are "isolated" in the context which gave rise to their isolation. The stroke of a tennis player or the pull of an oarsman are movements which have little in common with either the type or manner of execution of movements outside this specialised context. This is what makes the sportsman a sportsman; there are movements which are excelled at and which are noted for their difference. This is a stark phenomenon as sports and the general blur of normal actions are often sharply separated by conventions. We "do" sports at certain times and places, wearing special clothes and often on certain occasions. Thus the disparity between the use of the body in "normal" life and during sports is quite apparent. The history of AI has given us a superb example of this same phenomenon in the form of computer chess research. It became apparent that expert chess players soon dropped explicit rules in favour of heuristics and intuitive "feels" once their playing had reached a certain level. Moves are no longer explicitly licensed by rules or strategies explicitly composed from moves. Certain moves become "strategically isolated" in the sense that they play roles quite different to the roles the ordinary and non-expert rules allow. This feature of isolation is, I think, an important concept in the understanding of the way in which real-time action can occur.

In language use, we might say that certain parts of constructions become "semantically isolated" from much of the rest of language in the way that the action in expert sport is "physically isolated" from much of the

body's normal movement. Evolution provides further support for this. Nature tends, as is well known, to reuse existing structures for new purposes in the process of adapting to environments. Given the plausible assumption that physical skills preceded linguistic skills, we would not be surprised to see similar patterns in evidence. We think it obvious that certain attainments are only possible when particular skills have become "second nature". Now, "second nature" here implies at least that the skills so called are not decomposed in a manner common to less specialised action. The decomposition would be anathema to something being "second nature" at all. Every teacher of any skill knows that we learn by rules but excel by leaving them behind. In terms of a compositional approach, how would this be? The steps are rigidly defined and so how could proficiency increase? One possibility is that we could simply put this down to hardware. Our brains simply find more economical ways to do the same things, if we repeat them enough. However, the steps in a compositional model are all necessary; that is the nature of recursion. So what would "more economical way" mean here? We might simply say that the brain gets faster at doing common things without stipulating that the way that common operations are performed actually changes; the steps are simply performed more quickly in expert action. It is, however, as mentioned above, not simply speed that differentiates second nature from inchoate action and so this reply will not do. Compositionality is rigid in its operation, this being a desirable feature for formal reasons. Its rigidity comprises an explicit demarcation of operation at every stage and this is not a feature of expert skills or second nature. Further, there is more to being an expert than simply doing things faster than usual. If language and thought are second nature to us and second nature is not simply a matter of doing things faster then we would not expect compositionality to be a central feature of any plausible theory.

The "isolation" of common strategies regarded as second nature is a recurring pattern in organisms that are required to adapt to complex en-

vironments. Now, with language and the use of words, the boundaries of isolation between colloquial and creative usage – between second nature and new skills – is hardly so delimited. Colloquial language constructions might be said to be "crystallised" and thus semantically isolated in the way that expert physical skills are but they enter into the stream of linguistic practise so seamlessly that the special role of their components goes unnoticed. This is quite important as the isolation involved in language is therefore less obvious. There are indeed technical and specialised forums that are conventionally distinct from normal usage and these are distinguished by recognition of their quite different language uses. Weddings, speeches, funerals all have certain language use quite isolated from the rest of the language. The latter moves on but they remain the same. However, this "semantic isolation" is ubiquitous, the lurid conventional barriers of specialist discourse being special cases of a general crystallisation of common usage. This is one of the reasons that one might be tempted to apply a theory of compositionality across natural language; the barriers that would make such an endeavour seem misconceived in the case of, say, giving a model of the components of all physical movements across specialised and everyday cases, are not apparent. The barriers of semantic isolation appear, slowly, as the language changes around such phrases. We are then left with phrases whose component words have little in common with contemporary usage. The jargon of subcultures is an instance of this. Words come to have very idiosyncratic connotations for groups who use them in certain ways. Expletives are another good example. In such cases, the "meaning" of constituent terms is opaque; the phrases themselves are effectively, in the model-theoretic idiom, "atomic" just as certain physical skills become an irreducible part of the expert's movements.

The implication of this is that treating natural language in a compositional manner is a mistake made possible due to the barriers of semantic isolation being invisible in the stream of ordinary discourse. Common

phrases are rather like ice fragments in a stream; in the rush of the water, everything seems to be homogeneous. I think this way of treating the phenomenon of semantic isolation is justified further by an analogy with Kant's treatment of categories as all-pervasive. We could simply say – this is analogous to the traditional idiom of formal semantics – that perceptual illusions are exceptions to normal perception; when the mind interferes. Or we could say, following Kant, that the mind interferes constantly; everything is, to a naïve realist view, an "exception". Now, it may be that there is a certain amount of compositionality involved in the *learning* of a skill; we are taught by building up in layers and practising smaller components, combining them into larger. However, the argument above is intended to show that this is simply not efficient enough to apply in expert performance. Thus we need to realise a division between *learning* and *application*; compositionality is, at best, relevant to the former. We compose new phrases based on the "meaning" of atoms during the process of learning which crystallise during practise into isolated units we rapidly use in application and which sometimes are left behind to become fossils of language. We might account for the theories which depend upon universal linguistic rules and the like by noting that analysis of language is a relatively new task and, like all new tasks, it is approached with a desire to learn. A desire to learn will be a desire to compose new skills and thus it is not surprising that we impose a compositional structure on language when we come to examine it. It seems reasonable to say that the deconstructive and compositional approach to learning and formulating knowledge is something that the human brain has evolved as a useful technique. Of course, that does not imply in any way that the subjects and tasks thereby learned are in themselves compositional. It merely implies that this is the best way to get to grips with certain aspects of them. We can understand the role of compositionality in formal logic as making explicit this process of analysis of new information. There is no implication by this that the information is

of an area that has a combinatorial treatment as a *causal* element. Compositional conceptions of a subject area are powerful and can be applied to most tasks, given ingenuity. This is because we tend to learn tasks in this manner and thus have little difficulty in applying the concepts; we might have an (evolution-based) propensity to view all subjects and tasks – including language and thought – in a certain way but this is quite independent of the question of how we *do* them. Indeed the way we examine things might positively mislead us in our explanations of how we do them. This notion of crystallisation of language and thought has a basis in modern neuroscience. The work of Pellionisz and Llinas described in Chapter 5 shows that the output of systems in the brain governed by empirically implied geometrical models becomes isolated from the units that comprise it due to particular independencies between these units. Semantic isolation is, in my view, a manifestation of the isolation we find between output and its causes when we base our models only on the patterns displayed in that output.

There are two famous objections which apply to this picture; those of productivity and systematicity. Language and thought, it is suggested, have as central feature a productive nature which allows the construction of an infinity of novel utterances and thoughts. That only a combinatorial semantics can account for this is a central tenent of the traditional formalist picture. This has never been a particularly strong argument for compositional formalism however [11]. Given suitably fast mechanisms in the brain, a simple finitely bounded iterative account rather than an infinitely productive one will do. Infinite productivity is rather beyond what is required even if it does happen to comfortably fit with the recursive nature of compositionality. Systematicity is a more important objection. This requires that a theory be able to account for systematic regularities such as understanding or uttering "Bill kicks John" being, in some way related

[11](Fodor 1987) p. 294 (page number refers to Blackwells reprint)

to also understanding or uttering "John kicks Bill". It would be peculiar, it is argued, to be able to understand or utter one without being similarly capable of understanding or uttering the other. It is thought that the relation is *intrinsic* in a sense that only a combinatorial semantics could deal with. Thus, it will be objected, how can a view which denies the use of formalist concepts – indeed one like that suggested above which renders phrases largely opaque to the rest of language – during actual language use account for the recognised systematicities? One would not want to deny that language seems to display systematic features: understanding and being able to utter "Bill hit Bob" seems to be related in some way to understanding and being able to utter "Bob hit Bill". However, there are certain biases that might lead us to believe that systematic features can only be explained in certain ways. The common approach is to characterise the systematic features in terms of the explanation "level" that one then proposes to account for them. Pylyshyn's famous argument is a good example of this[12]. If we have a model E using concepts that are insufficient to constrain the model to reality, then we require another "level" of explanation E' in order to provide the necessary constraints. The problems come when ones "constraints" are formulated in terms of the the "higher level" model E'; the necessity of E' becomes a simple consequence of definition rather than real explanatory need. If we require compositionality as a constraint on models, then obviously a model that enshrines compositionality satisfies the constraint. The quote opening this chapter is evidence that this sort of trickery would go unnoticed given the largely unquestioned merits of compositionality. Often, the desire for systematicity *is* just the desire for compositionality and so it is not difficult to see how the desire for the former could be satisfied by the latter. Fodor:

> "[Systematicity] ... depends on the idea, more or less standard in the field since Frege, that the sentences of a natural

[12](Pylyshyn 1984)

language have a combinatorial semantics.[13]

Having one's cake, in this case, *is* eating it.

The task is to indicate how we might have some form of systematicity without compositionality; a task which, if arguments like Fodor's are convincing, is traditionally like trying to disprove a logical truth. A common tack, following Pylyshyn, has been to show the former as a consequence of a certain level of explanation; the semantically combinatorial sort of the traditional formalism programme. It is quite important at this point to examine briefly the source of the explanational biases of formal semantics. Given the accepted ubiquity of compositionality in work in the area, we might well expect that the type of explanations acceptable there would mirror aspects of this principle. Compositionality is a method that makes clear how something is built out of smaller units, these smaller units generally being built in the same way and frequently being tokens of the same type; propositions are often parts of propositions etc. This pattern allows us to see exactly how a construction is built, the general principle being the same at all stages. In terms of general preferences for types of explanation formalists tend to favour those where the principle of explanation is the same at all stages; it is obvious how something is explained and that the way in which it is so is the same at all stages of the total explanation. A good example is the strategy of the Language of Thought model where the parts involved in the explanation of, say, intending a complex proposition are tokens of the same type as the proposition itself i.e. other propositions. So, it could be maintained, I think, that it is not simply that compositionality is desirable to formalists as it satisfies certain criteria of good explanation, but rather that its ubiquity in the field actively *creates* the desire for certain styles of explanation. The form of the strategy of compositionality is the form of the explanational schema that it is designed to

[13](Fodor 1987)

satisfy[14].

4.3 The Language of Thought and Explanation

Jerry Fodor has held a considerably more sophisticated view of the matter which needs to be examined. In an effort to vindicate the syntactic nature of the Language of Thought, Fodor attempts to banish the influence of intentional properties when explaining mental causation. That is, the *contents* of thought are not themselves deterministic of their causal roles. This is important as it puts the work of language and semantics firmly in the remit of syntax and structured, systematic formalism. This, naturally, puts a heavy emphasis on compositionality since the *form* of thoughts is now of supreme importance in determining causal roles. Fodor's argument for this is mainly that identical (propositional) contents do not imply identical causal roles – the content P and the content $\neg\neg P$ need not have identical causal roles.

The emphasis on pure form and hence on compositionality that this argument aims to establish is avoided simply by saying that the causal role is not *solely* established by intentional properties. The gap in discrimination between causal properties and intentional properties is bridged by the influence of some other factor. There seems to me to be no reason to suppose that we must have only one source of the causal powers of thoughts. It is simply an aspect of the desire, noted above, for explanations whose elements are all tokens of similar types. It is premature to argue from the observation of a disparity between causal role and intentional properties, when we assume that the latter is the sole determinant of the former, to the conclusion that intentional properties have *no* part in the explanation of causal roles. I think this is evidence of the influence of the general pattern of explanation engendered by the form of the accepted notion of composi-

[14]See (Kime 1996) for further exposition of this point.

tionality. The recursive nature of this suggests that we employ concepts of a similar nature at all levels of explanation; this is exactly the topology of recursion. So, it would be very hard for Fodor to allow a more moderate position where intentional properties determine *some* of the causal role as this would be to allow in a very different element of explanation. It would, in effect, break the recursive nature of the explanation and thus break the link with compositionality.

Interestingly, Fodor anticipates part of the argument used above which questions the need for a certain sort of explanation of systematicity. It is argued that even though "psychological laws" are formulated in terms of intentional content, the explanation of these laws need have no reference to this content. The aim of this argument is to show that the prevalence of intentional content in talking about language and thought implies nothing about how it is to be explained, thus leaving the way open for an account – the Language of Thought hypothesis – that leaves intentional content completely out of the picture. This is a perfectly reasonable tactic, as Fodor suggests, the pattern is common throughout the physical sciences; that which is explained takes no part in the explanation. What is puzzling is why Fodor does not allow this to apply to the systematicity that rival theories are said to lack. The only explanation possible, it is held, of systematicity is to have a compositional formalism of language and thought. Given that it was argued above that systematicity *is* compositionality in many respects, then Fodor is, here, requiring that that which is explained *takes the main part in the explanation*. This is completely contrary to the strategy with which he aimed to establish that a certain formalism was alone sufficient to account for causal roles. This is not, I think, a conscious use of explanatory level distinctions to insidiously serve polemic ends. Rather it is a consequence of the ubiquity of compositionality as an element of formalist thought which renders some oblivious to the enormous *petitio principii* which results.

Fodor has some interesting things to say regarding the sort of argument proposed above concerning "isolation" of behaviours. In (Fodor 1987) we have the following argument:

> Suppose there is a kind of event c1 or which the normal effect is a kind of event e1; and a kind of event c2 of which the normal effect is a kind of event e2; and a kind of event c3 which is the normal effect is a complex event e1 & e2. Viz.:
> c1 → e1
> c2 → e2
> c3 → e1 & e2
> Then, ceteris paribus, it is reasonable to infer that c3 is a complex event whose constituents include c1 and c2.

Fodor thinks that anyone who denies the above, which is clearly a consequence of a certain attitude to compositional formalism, is ignoring the "untendentious" principle that says one should "minimise accidents". For, if we denied the conclusion of the above argument, we would be forced to say that c3 was a unique event whose similarity to the coincidence of c1 and c2 was, in effect, an "accident". This is, I think, entirely unreasonable as a methodology for a science that claims to be concerned with human beings. It would require us to ignore evolution which is possibly the most messy and "accidental" account of capacities there is. Fodor even takes this to extremes in arguing that this applies to behaviour. That actions, when performed as complexes, are tokened in the same way as when they are performed singly. To deny this – as was explicitly done above – is, again, to maximise "accidents" and to ignore the "elegant" syntactic theories that result otherwise. We are not in the business, I take it, of elegantly reconstructing the inelegant, opportunistic and messy proclivities of evolution however. Fodor is quite acidic regarding those who posit "Unknown Neurological Mechanisms", charging them with preferring no theory to a computational theory such as his. This, I think, is hardly a charge worth avoiding since the implications of a computational theory employing certain tools is that restrictions are thereby set on what might count as a possible ex-

planation of phenomenon whose relevance is dependent on the use of these tools. At least "no theory" would leave the area a little more open to diverse investigation.

The concept of "accident" here is rather tendentious. An accident is only such compared to a certain way of looking at things. If one believes that logical rules causally govern language and thought then features that are not accountable in a model with such assumptions are naturally seen as accidents. In the above example, calling the case where c3 $\rightarrow$ e1 & e2 did not imply that c3 internally tokened c1 and c2 an accident, requires certain conditions. It is only possible when there is a body of theory that supports claims such as:

- Uniqueness (in some manner) of events of type cn
- Events of type cn are decomposable
- Events of type cn can combine in ways that produce effects recognisably the combination of separate events

and so on. Without this sort of background, it is *not* reasonable to suppose that the simple explanatory structure elucidated by Fodor in his example is one that necessitates a compositional approach. In the case of formalist theory, the necessary types of background claims needed are enshrined exactly in the theory the argument is designed to vindicate and thus embody a circular argument. This point is extremely important and is discussed in more detail below.

The best defence I think one can give for Fodor here is a pragmatic one. Surely, one might say, without a recursive, compositional syntactic theory, the number of "accidents" would be such that we could not possibly suppose that this could be implemented in real time by real humans? If c3 in the above example were a *sui generis* event, then this implies that all events would be unique; all sentences would be "isolated" in the previously

expounded sense and thus there would be no regularity that would facilitate real-time processing. This point is well taken and the preliminary to meeting it is to note an implication of the preceding discussion. Regularity need not be explained in terms common at the level at which it is manifest. Isolated events may only be isolated for a certain way of looking at things. Ice cream sales and stickiness of road tarmac would seem to be isolated phenomena until one realises that outside temperature connects the two. The formalists mistake is simply an aspect of the correlation fallacy; it is assumed that correlations between structures must be explained in a manner that allows for – often obvious – causal roles. Every budding statistician knows that correlation implies nothing about causality. This is the general case of thinking that the systematicity of language must be explained by a language; if not, compositionality is unnecessary. This common pattern – the belief in the necessity of features of a certain type of explanation is a feature of the general stratificational view – will be explored more deeply later.

4.4 A Disanalogy with the Natural Sciences

A very common feature of the formalist approach is to draw seemingly appealing and intuitive parallels with the natural sciences. This is sometimes done in retaliation to the question "what evidence do we have for the atoms of composition; what does combinatorial semantics *combine*?". The tack is to justify certain ontological commitments by appeal to standards which detractors should deem desirable.

One strategy, mentioned above, is to appeal to parsimony; combinatorial formalism is simpler, in terms of its ineliminable ontology, than any rival. Fodor, for example, is happy to quantify over parse trees. This is itself a misleading criterion which fails to appreciate the differences between laws and systems governed by them. We, might have, for example, reason

to expect the basic "laws of nature" to be simple but not their manifestations. Chaotic deterministic systems are "simple" in one sense but their manifestations are not. It seems to be that formalists often mistake the ontological parsimony recommended by post-Duhemian science for a parsimony of the *manifestations* of ones ontological commitments. The fundamental laws governing the brain might well be "simple" – we might expect this from analogy with the accepted simplicity of related laws in physics and chemistry – but this does not mean that any model of semantics must be.

It is clear from physics that the constraints on the laws of matter are quite severe given the necessary conditions for stability thereof. Given this, we might expect laws to be quite rigid, simple and clear. However, the brain is a biological phenomenon, which sorts of things are governed more in their style of operation by evolution. This is constrained by the environment. The environment is not particularly rigid in the space of possible adaptations it allows, evidenced by the huge numbers of species and more particularly, their variation. Thus, it is a mistake to suppose that simplicity is a good methodological guide in *all* areas of science. Sometimes we have good reasons for supposing the laws to be simple, sometimes the constraints are such that we have no reason to suppose either way. Now, it may still be true that as a pragmatic guide, simplicity may be all we have to guide theory choice, even in an area where there is no reason to suppose it is a good guide. However, there is still a difference between the cases where simplicity is a theoretically justified guide and those where it is a wholly pragmatic expedient. We should suspect the latter sort of more errors when it comes to *explaining*. Formalists tend to gloss over this important point in attempt to forge links with paradigms of good methodology; a strategy that draws heavily from Chomsky's earlier work.

Chomsky was eager to point out that claims to the "psychological reality" of certain linguistic rules are exactly the same as in the case of claims

to "physical reality" in the natural sciences. Famously, giving a model of an abstract principle that covers some data in linguistics is logically the same as proposing a theoretical structure for data regarding, for example, heat emissions of black bodies. This is really the heart of the analogy that has been used since to justify the whole formalist enterprise in linguistics. It is quite clear, however, that the analogy is a bad one. There is a distinction, made quite clear by Quine, between fitting data and explaining it. We may be able to account for the regularity in some data by a model that posits an underlying mechanism but this is not the same as giving an explanation of the *causes* of the regularity. The difference is in the way in which the claim is related to other more or less fixed points of accepted theory. We can, to examine a model of ion exchange, look through electron microscopes and thus tie in with other areas of knowledge. If the areas we tie into are firmly held, we may say we have "established" the new theory. In linguistics, there is simply not the body of related theory that would provide any leverage for claims to explanation over and above mere models of data. This is not to say that there is an obvious distinction between fitting data to models and "really explaining"; merely that the latter, if used sincerely at all, is the result of a relation to a large body of firmly held theory. This is a relation that formal semantics does not have to anything. The more modern concern over how to link formal theories up to lower-level neurology or neural net research is exactly an attempt to provide such leverage. In a dialogue concerning his views on Language and Thought, Chomsky commented that to reject formalist theories as merely fitting data rather than explaining is like saying to a physicist:

> ""your evidence about the sun only has to do with light being emitted from the solar periphery, and I don't call that evidence about 'reality'. For me, evidence about 'reality' is limited to experiments in a laboratory placed inside the sun where you actually observe hydrogen becoming helium, and so on.""

This, for Chomsky, is "obviously absurd". In saying this, he is quite right

but the analogy he wishes to the case of formal linguistics is not appropriate. The attitude in the above quote is not appropriate to the case as this would render incomprehensible many other branches of theory to which solar physics is related. Formal semantics has no such detailed and rich links which could prevent exactly the response above. The "crisis" in addressing the problem of its biological plausibility that was brought on with the advent of connectionism is exactly a manifestation of this. A new area of research is rightly criticised for a too distanced relation with its fundamental subjects when it has little other supporting and related theory that could aid in a justification of its proximal theorising.

Up to now, formal semantics has been in the position of attempting to provide leverage for its "explanations" by pushing against itself. One justifies ones theory by pointing to another formal theory that says similar things. However, when the assumptions are the same, this is simply not good practice. One cannot claim that formal semantics and linguistics are comparatively new subjects – a common line in defence to criticisms of lack of progress – and also maintain that it is has the necessary embedding in general theoretical history to support, say, its quantifications over questionable entities as an exercise in something more than mere modelling.

A further point against the analogy with the natural sciences is the rather specious argument regarding the support given by evidence. This is the classic view of the superiority of the natural sciences among disciplines concerned with knowledge: that it has a bedrock of "evidence". Chomsky has always been keen to cleave to natural science methodology and this aspect is obviously attractive. The practice of formal syntax and semantics has been concerned with using grammatical judgements and native speaker intuitions as the foundations of evidence for the resulting theories. Thus, the claim is that there is abundant "evidence" for the formalist approach, again utilising an analogy with the natural sciences. Again though, this is misleading. An obvious difference is that this "evidence" for formal-

ist enterprises fails to have one of the central features of evidence in the natural sciences: stability. Observations carried out in natural sciences tend to be stable. They tend not to change much when you alter certain things within quite wide limits. Iron here weighs much the same as iron in the next room. Native speaker intuitions famously vary considerably and it is often hard to have *any* intuitions regarding some of the contrived sentences used to attempt to provide "evidence" for particular theories of, for example, pronoun resolution.

It is lucky for natural sciences that there are reasonable mechanisms in place that help prevent the dictates of theory colour the evidence that supports them. Nobody, since the time of Mary Hesse's famous paper on the observation/theory language distinction[15], would want to suggest that theory and observation are independent but re-testing, blind testing and measuring tools all aid in attempting to keep clear the goal from the route to it. Intuitions about grammar and semantics are often, as anyone who has done enough work on a particular theory will know, eventually coloured by theory. It is reasonably common to have seasoned semanticists with little or no intuitions left regarding certain constructions, so battered have these been by the onslaught of theory. In this case, there simply is nothing to help us sort out the "real" evidence form the spurious "damaged" cases where the theory has influenced the data. Intuitions are simply not the sort of thing one can be justified in or have "better" reasons for. One might convince a scientist whose theory predicted red litmus paper that it was not red by spectrally analysing the paper. What does one do in the case of a formal semanticist whose theory predicts cross-discourse anaphora and who says he can, as partial evidence, intuitively see such? The point is that in the case of formal semantics, there is a particularly incestuous relationship between theory and evidence that destroys the analogy many wish to use in support of its theories.

---

[15](Hesse 1970)

4.5 An Alternative to Compositionality?

The preceding argument has been set to establish the following: that there is no good reason to suppose that there is a case for holding compositionality and thereby systematicity as central features to be upheld by a cognitive theory. Further to the aims of this work, this is a concrete example of how the stratificational view causes problems for traditional semantics. There are two main reasons why the error has been maintained: firstly, by a historical influence of technical ideas from the logical pioneers explicitly regarded by them as unsuitable for the task of natural language analysis – the use of such technical tools has significant non-technical manifestations in biases in the general type of explanation that is acceptable to the current paradigm; secondly by a misleading analogy with the methodology and results of the natural sciences.

As noted above, one of the main objections to compositionality as a realistic explanation of semantics is that its recursive nature means that one must decompose down to basic level elements every time an act of inference or understanding occurs. This sort of model provides for an enormous amount of computation in real time and depends on the assumption that the only possible account must involve a small number of very productive, systematically related rules. If we accept the picture presented earlier in this chapter; that of "semantically isolated" units, then we do not have the basic atoms with which to compose and nothing like composition can take place.

In terms of the geometrical picture sketched previously, we can describe "semantic isolation" thus: A is semantically isolated from B if it typically inhabits a geometrically different space. Thus, analogously with physical activities, we might use a different space to describe the degrees of freedom employed in an arm movement that was part of a game of tennis as opposed to an identical movement employed in waving to a departing

train. There are very different criteria for use and muscle deployment, even though the actual movement through the air might be the same. If we regard conceptions as features of geometrical spaces, then isolated conceptions are geometrically isolated in this sense. This is a simple idea but is important for the following reason. A compositional model is, essentially linear, each step requiring the completion of the previous before it may occur. Thus, the amount of parallelism required for pragmatic plausibility is absent. This is rather like having one enormous geometrical space with dimensions enough to accommodate every possible combination of predicates as these are defined in terms of the same domain. Thus, as mentioned in chapter 3, something the size of the cartesian product of the set of possible predicates would need to be computed in real time. Imagine a space with one axis for every predicate; every conception being, say, a point in this space. This is simply a geometrical picture of the kind of impossible situation realised by Dreyfus. To determine the coordinates along every axis is exactly the same as having the combinatorial explosion involved in explicitly describing the conceptual dependencies of any real situation in a logical formalism. It is the relations between concepts that provides for problems as this grows faster than the number of available concepts. Relations between concepts are handled normally by the notion of inference or set-theoretical operations that can be exploited in a compositional treatment. One has rules that say "All xs are ys" or "if P then Q" which are steps in the decomposition of thoughts and utterances. Here is the implementational crux of the systematicity requirement; explicit delineation of rules and relations that embody explicit steps in formal methods, procured to account for some of the regularities of language and thought. Fodor takes it as given that systematicity is explained only by combinatorial semantics:

> "... that [the] systematicity of cognitive capacities implies the combinatorial structure of thoughts ... I get ... for free for

want of an alternative account."[16]

Thus a rejection of compositionality requires an account of what might allow conceptual relations and inferences; what allows the manifest regularities.

4.5.1 The Importance of Analyticity

Analytic statements are, in a sense, degenerate forms of inference. If "All objects A have property P" is analytic, then an inference from being A to having property P is trivial. Since Quine's "Two Dogmas" it has been argued by many that analyticity is not something fixed and stable; what is counted as analytic changes as the theory in which such statements are embedded changes. This provides for a dynamism of analyticity and necessitates the famous Neurath metaphor concerning at-sea boat repairs; stability is now provided, not by immutable certainties of definition, but by dogmatism in acting *as if* there were such things. Epistemologically, there is no difference. Since, when we consider the brain in a modern monistic fashion, we are essentially following Kant and are, as argued earlier, limited to epistemology, we are in the same position as regards the nature of analyticity. That which our brain renders as analytic because of the structure that is imposed on perceptions is indistinguishable from that which we might regard as "really" analytic. The difference is utterly inscrutable since the conditions that we might use to distinguish between these two cases are the conditions that ensure that an epistemological difference can never be found.

I suggest, then that analyticity is exactly the mechanism by which we perform what we now call "inference" and that no "real" inference in this sense ever takes place, it actually being a simple matter of analytic definition. This would seem to suggest that we could never change our minds

[16](Fodor 1987)

about certain inferences or relations between thoughts or that we would all agree on certain things. The freedom to deny this is afforded us by the picture given by Quine. That which is analytic changes depending on the theory changes around it. These sorts of theory changes are slow and gradual as they require re-workings of ontological commitments and sometimes fundamental metaphysics. The brain is more temporally dynamic in this respect; changes in chemical balance are frequent and sometimes quite discrete. I propose that these changes facilitate a rapid change of geometry of "conceptual spaces" such that putative inferences and conceptual relations are rendered possible at great speed in a particular manner.

Suppose that the chemical and structural elements in a part of the brain embody a network that defines a space in which a certain activation in turn embodies a manifestation of the thought that the people walking past the window are wearing heavy coats. Classically, we would need to do something like provide a rather implausible meaning postulate in order to syntactically achieve the link necessary to obtain the inference that it was cold outside. Even then, there would be a significant amount of processing in order to obtain the result. On the present theory, what happens rather is that the embodied network, in which activation manifests as the first thought, has a geometry whose axes are interdependent in a manner that makes likely activation manifested as the second thought. This is a result of the structure of the space and thereby is analytic by, on Quine's terms, the only standards available. The structure of the conceptual spaces involved creates analyticity in various places just as the structure of the network of theory does so in Quine's view. The difference lies in the fact that the structure of the embedding network in the brain is rapidly changing due to constant chemical and physical perturbations; experience is a constant flow. Thus, what is "analytic" for certain conceptual spaces changes constantly. This implies that "inference" as classically conceived, need never happen.

A point of clarification is needed here. It is not being suggested that the concept of "inference" is empty. Inference in terms of the traditional rules such as Modus Ponens certainly exists as a reconstructive, normative science. The introspective arguments designed to show that such rules of inference play a part in our actual cognition are misconceived exactly in that they confuse being an approximate model with causing. The history of the Philosophy of Logic demonstrates the complications and confusions one finds when attempting to fit normatively conceived models to "real inference"; nobody believes in the adequacy of the simple material conditional anymore; simple quantification is often deemed inadequate; two truth values do not seem to be enough etc. and there is little sign of this continual tweaking of logical inferential patterns reaching an end. This is, I think, simply because it has nothing to do with actual human cognition in which "inference", as commonly conceived, is a myth.

A reconstruction of an "inference" is a process of giving a synthetic description of an analytic connection. The reasons for this are as above in Fodor's bias in giving a certain sort of "level of explanation" for systematicity; a synthetic description describes at the level at which the results of a putative "inference" become manifest. If an "inference" links linguistic element A with linguistic element B then our bias is in favour of an explanation of the link that employs nothing but other linguistic elements. Again, the tack is to "minimise the accidents" that we have, as argued above, no good methodological reasons to avoid. Also, analyticity is commonly defined in terms of linguistics units. A statement is analytic if, for example, one can count the subject as part of the predicate or the truth of the statement is a matter of "definition of words". It is clear that even though we might be able to *state* analyticity as matters of definitions of words, this need not necessitate the phenomenon also *being a consequence of them*. Again, this is the modelling/causing confusion. Wittgenstein had exactly this trouble in attempting to show that "object A is red and green

all over at the same time" is a logical contradiction. Whereas the problem is manifest in language, its cause is not simply a matter of words (see Chapter 7). It is clear, however, that something might be analytic as a result of sub-linguistic informational dependencies in the brain. The activation of certain networks manifesting as the thought that A might be intrinsically related – due to the geometry of the spaces involved – to activations manifesting the thought that B. If analyticity is simply a sharing of informational ingredients in this way, then the linguistic version with which we are most familiar turns out to be an aspect of sub-linguistic relations and thus any talk of compositional models *causing* and thereby really explaining language and thought becomes redundant.

George Lakoff seems to have adapted his view as to the fundamental appropriate ways of talking about semantics. In 1972, he was concerned with extending the notion of generative grammars[17] while more recently, he has turned his attention to a "Cognitive Linguistics"[18] This is akin to Gärdenfors' approach (see Chapter 7) in that it emphasises a more proximal vision of the data to be accounted for. The evidence is now largely cognitive rather than concerned with words; Gärdenfors expresses this as explicitly as "Meanings are in the head"[19]. Lakoff notes that metaphor indicates close connections between ostensibly different areas of language and thought. In particular, he notes an interesting regularity between terms used to express personal relationship problems and those used to express journeys[20] Lakoff accounts for this by talking of "cognitive topology". There is a commonality in structure between different domains of discourse that allows the metaphorical relationship. He characterises this in terms of more and more abstract mappings. For example, in the journey metaphor embodied in phrases such as "We're drifting apart" and "We'll

[17](Lakoff 1972)
[18]His main statement of this is to be found in (Lakoff 1987).
[19](Gärdenfors 1993*a*)
[20](Lakoff 1988) p. 302

have to go our separate ways", he identifies the mapping "TRAVELLERS correspond to LOVERS" and "PHYSICAL CLOSENESS corresponds to INTIMACY"[21]. There are, he thinks, more general such mappings such as the "event structure metaphor" where "LOCATIONS correspond to STATES" and "MOVEMENTS TO NEW LOCATIONS correspond to CHANGES TO NEW STATES". The crucial thing for Lakoff is that metaphors can map from spatial to non-spatial domains. This is useful as non-spatial domains have what he terms "container-schemas"; areas of distinct categories that are amenable to logic since the boundaries of categories define rigid units. So far, this account is almost identical to that given by Gärdenfors and detailed in Chapter 7 but of importance here is the following conclusion. Lakoff argues that metaphors can map into schemas with built-in topology and thus can have virtual logical relations for free. This is exactly as I have described in the example of the egg-box in Chapter 3. In Lakoff's account, inference in the sense of deductive formal manipulation of symbols need never happen; one merely employs a metaphor that maps into a scheme that has the topology that makes the conclusions of inferences "obvious". This is all rather vague and Lakoff talks of merely having to "shift focus" from one domain to another. This is, I think, a result of Lakoff's assumption that this model can be supported by an account located in the conceptual level of explanation. He asks for a reason why the mapping from visual scenes to concepts displays a tendency to lump certain regions of the continuous data together:

> "What kind of concepts permit such an infinite categorisation of visual scenes?"

This is an interesting question but contains an assumption to the effect that *concepts* are responsible for this phenomenon. If the phenomenon of categorisation is explicable in terms of properties of the concepts so categorised, this conflicts with Lakoff's model of metaphor in the following way.

[21](Lakoff 1988) p. 303

The structure of a concept is, on Lakoff's view, responsible for the categorisation of the continuous data of experience. Now, the essence of metaphor is to relate conceptual schemes to others and thus reflect the structure of one scheme in another. Now, if *concepts* are the originators of structure, the real work in this account is done by the notion of "metaphor" as it is this that relates conceptual schemes to others. If a conceptual scheme has a certain type of structure A and certain metaphors require this structure being embedded in another conceptual scheme, metaphor is that which performs this task. But since this whole account was, in part, intended to explain the notion of metaphor, it has, it appears, ended up relying on it for the lion's share of the explanation. Essentially, in what does this "relating" or "mapping" of conceptual schemes consist? A mapping or relating at the explanatory level of concepts is constrained by the sorts of operations available at that level. These are formal relations since concepts are supposed to be, on Lakoff's model, emergent from low-level brain processes and describable in the usual linguistic terms. Thus, this later phase of Lakoff's thinking is effected to the same degree as his earlier work as detailed in Chapter 3; there is no benefit in having *concepts* determine the patterns we see in language as this leaves us having to explain how the patterns interrelate at the conceptual level. This does not sit well with his notion that no deductive inference is necessary on his model. The relating of conceptual schemes at the conceptual level requires relations between concepts and these relations require formal systems of some variety. The problem is that Lakoff wants to retain the importance of the usual linguistic concerns and patterns, while showing that there is a basis for these in low-level brain function, more specifically, through connectionism. Others have also desired to do this[22]. This is an attempt to present a stratified model in which the "higher" level of explanation provides evidence to be accounted for by characterisation in, in this case, neural nets. Lakoff says that

[22]Notably (Gärdenfors 1993*b*)

> "Such an explicit mapping would also provide characterisations of such notions as "basic-level concept" . . . '[23]

This is a top-down approach where we take the evidence from our manifest language and thought patterns and hope to "ground" them in the neural substrate. However, the manifest patterns, such as "basic-level concept" are *output* patterns. They are what we see as results after input has taken place and after the brain has processed this input and combined information into output. We cannot ground manifest patterns in a neural substrate unless we are sure that these patterns have any meaning or applicability at that level. It would be like trying to model water movement by "mapping" the phenomena of waves onto molecule interaction. At that level, the concept of "wave" means nothing. It is merely something applicable as a rough description when talking of large collections of water molecules. Lakoff's problem is something we see time and time again in Cognitive Science. It is what I have called the "stratification" problem and its causes are to do with the methodological issues involved in attempting to relate, yet keep separate, different "levels of explanation". Compositionality is one of the main features of stratified views of language and thought and, as discussed above, this is confused with notions of explanation tied up with the desire of semantics and linguistics to be seen as sciences. These are themes that will recur in the exploration of the issues involved in adoption of a stratified view.

In order to provide a basis for remarks to follow in the more general approach to the problem I take in Chapter 6, the next chapter examines some crucial results and implications of some work in neuroscience.

---

[23](Lakoff 1988) p. 307

Chapter 5

Semantics and Neuroscience

Commonly, it is lamented that the gap between the sorts of evidence supplied by low-level brain science and that required by cognitive science and its work in natural language semantics is so large as to be currently inhibitive of relations between the two. This is regrettable since the excesses of the formalist program have been exactly those warned against by behaviourism in its fear of "mental models" and it might be argued that any sort of empirical evidence that might bear on semantics would be most welcome. Behaviourism in psychology warned of the dangers of constructing models which had no determinate basis. Postulations of mechanisms of the "mind" or of "cognition" were thought to be worrisome as no direct evidence was available in respect of them. The fear was that psychological ontologies would come to be accepted on the basis of nothing more than the attractiveness, coherence and implementability of models designed to fit data. Not to disparage the use of such criteria, it was not thought however that they should be the primary determinants of theory. Behaviourism in this field was largely a movement to preserve the basis of empiricism. Cognitive Psychology and the formalist approach to mind and language embodied a subtle shift in what was taken to be an appropriate basis for empirical research. Behaviourism was a denial of the possibility of such a shift. It was a shift from a basis of relevant observables to one of questionably relevant

observable manifestations of relevant non-observables; a shift from regularities in the subject matter to regularities in something you suppose to give you clues about your subject matter. Behaviourism worried that the lack of a relevant empirical basis for certain subjects caused them to create spurious ones by redefining the notion of "empirical base" in a way actually not really constrained by empirical requirements. So, in attempting to motivate a picture of semantics by data from the brain sciences in opposition to a stratified view, certain assumptions are necessary and broad implications of the type of data available need first to be examined in order to establish the basis of the evidence which is to be discussed.

Additionally, there is a need to elucidate some of the implications of the geometrical models favoured by empirical brain science as these have a direct bearing on the philosophical issues surrounding the relation between types of explanation in Cognitive Science. It is apparent that work in the semantics of human cognition cannot fail to be affected by issues bearing on the influence of perception on cognition and thus by post-Kantian thought in general. It is important then to note that work in neuroscience in the last few decades has found it germane to compare some of the results obtained with general post-Kantian concepts. This is interesting as it implies an empirically supported basis on which to draw the broad outlines of an account of language and thought. The general Kantian flavour of neuroscience research that has touched upon broadly semantic issues gives us reason to reject the traditional formal approach and to embrace a quite different theory. However, the theory suggested is so fundamentally different in terms of the sorts of explanations allowed that it is necessary to understand the motivations behind the neurological theory and also the broad implications that it has. In formulating these ideas, we must be careful to sidestep (and explain this sidestep of) the explanatory desiderata of the formalist lest we burden ourselves with explanatory criteria of no relevance to the actualities of brain behaviour.

5.1 The Import of Kantian Transcendentalism

Kant's famous work in the "Critique of Pure Reason" is held as a milestone in philosophy. It is a work that bears heavily on the relationship between mind and world and thus on issues of realism, idealism and scepticism. The implications for semantics and Cognitive Science are also immense but underappreciated. Part of the reason for this is that the philosophical concerns of Kant are not obviously related to traditional work in formal semantics; partly also because there have been many complex criticisms of Kantian philosophy since which are not easily couched in terms relevant to modern cognitive science. This latter problem is due mainly to the language and concepts available to Kant at the time; he talked of the "mind" rather than the brain and Cognitive Science, as is well known, is still concerned with the incessant mind/body debate and thus arguments that depend on this area are best left alone. A step towards rendering Kant's work as more obviously relevant to Cognitive Science is to facilitate an approach that is couched in the modern monistic idiom and that has support from brain sciences.

The fundamental message of the Critique has to do with a certain relation between mind and world where the mind is seen as an *active* element. Kant rejected the realist model where mind was a passive model of reality, taking the evidence from recalcitrant illusion further and arguing that the impositions of the mind are ubiquitous, all-pervasive and thus *constitutive* of that which we call "reality". The effect of this sort of theory is, as Kant undertook, to investigate "reality" by investigating the structures of the mind since, *ex hypothesi*, the latter is constitutive of the former. This forces a rather different paradigm onto any subject area that accepts these basic principles.

Kant's two great "impositions" of the mind were space and time.

> "It is, therefore, solely from the human standpoint that we

> can speak of space . . . If we depart from the subjective condition, the representation of space stands for nothing whatsoever."[1]

> "I can indeed say that my representations follow one another; but this is only to say that we are conscious of them as in a time-sequence, that is, in conformity with the form of inner sense. Time is not, therefore, something in itself, nor is it an objective determination inherent in things."[2]

The surface manifestations of these categorical impositions are not signs by which we could hope to separate the "real" from the "ideal" as this separation supposes a freedom from the very categories that lead to the signs. John Hospers[3] likens this situation to a village of fisherman who only have nets with holes 1′ square. They never catch any fish under 1′ long and conclude that there are no such fish in the sea. They regard this as a fact about the sea while not realising that it is a fact about their nets. The ubiquity of the effect of this on their epistemological life is taken by them to be a fact about ontology. Thus, any explanation of a part of what we take to be real is fatally flawed if it takes the evidence as supposing that the real and the ideal are epistemologically separable. That this is an epistemological problem, Kant was quite clear:

> "This ideality of space and time leaves, however, the certainty of empirical knowledge unaffected, for we are equally sure of it, whether these forms necessarily inhere in things in themselves or only in our intuition of them."[4]

What, it might be asked, about the certainty of more theoretical knowledge? This can hardly be unaffected as the ontological assumptions of a theory are rather dependent on views of idealism. A process of compositionality in formal semantics, for example, requires a base to its recursion.

---

[1](Kant 1787) p.71
[2](Kant 1787) p.79
[3](Hospers 1956)
[4](Kant 1787) p.80

This base is in danger if there is an assumption that it is possible by "direct" association with objects in the world or if the notion of "object" that is used to build up extensions involving sets for predicates etc. are motivated by a naïve realism. The emphasis on research, in the light of such Kantian ideas is to be on an examination of how our manifest language and thoughts are to be understood in terms of the categories of the mind. What is done in modern formal semantics is rather that the nature of language and thought is to be understood in terms of the aforementioned manifestations, its regularities and apparent structure. This is, on Kantian terms, a mistake as the apparent manifest structures arising from interaction of the categories need not be obvious and indeed one effect has been such as to lead us into the errors of naïve realism. Kant presents the archetype of the argument that how things appear to be on the surface gives no real clue as to how they are formed, caused or performed. The reason for this is simply the lack of an external point of reference that enables us to distinguish reality from appearance.

As an example, languages generally contain mechanisms for marking tenses. We can distinguish between past, present and future. As a result, we have logics and formalisms utilising time-coordinates in order to capture the difference in "meaning" between "I saw John" and "I will see John". We might have, t = now & at(t,see(me,John)) and t = now & at(t',see(me,John)) where $t' > t$ etc. The sub-formula concerning times is well-formed and thus can stand alone, be combined with other formulae and dropped etc. Thus we render the implicit assumption about the orthogonality of time. We take in factors of the language we use as being a guide to the models of its underlying nature, of its "real" source and origin. In Kantian terminology, this is the impossible inference from phenomena to noumena. Another good example is the work in "propositional attitudes". We note that we often have syntactic propositions occurring after the word, say, "believes". So, we have a "propositional" syntactic theory

of sentences containing the word "belief" which gives us certain syntactic entries for such words. However, we then are tempted to infer that belief itself is "propositional". We "believe propositions". Language structure has led us to a theory of a very different sort of structure; a "mental" structure and a structure of "meanings". If there really is a sort of division between something like Kant's phenomena and noumena – and I think there are good reasons why something like this can be upheld for reasons given below – then the evidence of language is very poor as it inextricably mixes up the way the world is with the way we force it to be. Using language regularities to explain semantics is then, problematic as the structure of language need have nothing in common with the structure of that which gives rise to it. Language manifests patterns certainly but the lesson from Kant is that the cause of a pattern is necessarily inscrutable from simply looking at the pattern. On the Kantian picture then, I shall argue that the methodology of traditional approaches to language and thought are completely misconceived in their use of patterns in explaining the cause of those patterns.

Naturally, what is required is an argument that the Kantian criteria are desirable. One could simply reject Kantian transcendentalism and leave formal semantics and theories of language and thought in Cognitive Science in place. However, there are good empirical reasons, I think, why this cannot be done. The support comes from research performed in the last decade and and a half in neuroscience and psychology. Obviously, whereas Kant was wont to talk of "the mind", modern brain sciences prefers to talk of the brain. This is of little consequence I think given that neo-Kantians, in particular Schopenhauer, rephrased Kantian theory explicitly in terms of the brain; where Kant talked of categories of the mind, Schopenhauer talks of the categories imposed by the workings of the brain. Worries about dualism had reared themselves in the meantime. The important point of the Kantian theory is that much of what we take to be reality is contrived *internally*. Two particular areas of modern research, those of "Categor-

ical Perception" and research into motor coordination in the cerebellum, display these Kantian tendencies clearly. I think it an essential step in approaching a reasonable methodology in Cognitive Science as a whole that this connection be made clear. It is essentially the task of this whole work to show that the implications of such a connection are overwhelming and involve even the most concrete instances of work in the field.

5.2 Categorical Perception

A feature of our perceptions of continuous data is that we tend to "quantise" into discrete sections. The categories that we tend to group things into are perceived as having a greater similarity between their elements than is the case between inter-category elements even though the difference can be shown to be the same. For example, it can be shown that we tend to see all red colours as more similar to each other than to any yellow colour, even though, for certain choices, the difference in wavelengths is identical. This is a general phenomena that can be used to provide, for example, a naturalised basis for the perception of natural kinds. These are, according to this model, merely types of things that the structures in our brain are so organised to see as having more in common with each other than with other things, even though other measurements - sometimes taken to be more objective – give different results. It is exactly that our categories are a result of the structure of our brain. It seems to me that this conclusion is very much in the spirit of Kant where the imposition of space as a condition of sensible appearances results in categories such as "unity", "plurality" and "totality". Of interest is exactly how this process is thought to occur. In modern times, the thought has been that there is some sort of virtual "conceptual space" which the neural activity of the brain embodies. This space has certain topologies and metrics which make the quantising of continuous perceptual data necessary. This is, I think, the modern analog

of Kantian categories; the sensible condition of space imposes categories on thought and perception.

An example is the *Voronoi Tessellation*[5] of a space of points. If we have a metric space in which points represent, for example, perceptions of colours, then as long as we have decided on certain archetypal colours – we need to be able to decide upon one example of each colour that we are sure about – the space will automatically be divided into convex regions corresponding to each colour. This happens by the following method: for each point p in the space, p is counted as being an example of the nearest of the archetypal points mentioned above. "Nearest" here requires a distance measure defined on the space; hence, the space needs to be metric. A convex region of a space is defined in the usual mathematical way: a region is convex iff for any pair of points s_n and s_m in the region, all points between s_n and s_m are also in the region. Such regions are also said to be "connected". This then is a possible definition of a categorical perception; one which falls within a convex region of a metric space.

It is then on this sort of picture, the structure of the brain that results in the manifest properties of the resulting thoughts. CP effects are, I think, a good analogue of the Kantian imposition of categories on sensible experience. An important thing to note is the *type* of explanation offered for their origins today. The theory couched in terms of metric spaces and points is geometrical and is something we shall come across again. The lesson from this however is rather that the existence of CP effects lends credence to the Kantian picture and helps to set it in a modern idiom. We might even hypothesise, as a more specific example, that the brain imposes the Kantian category of "plurality" by individuation based on Voronoi Tessellation of conceptual spaces. We conceive of many things by the space in which they are conceived, embodied in the brain, having a metric which defines in the way described, convex spaces. These convex spaces correspond to

[5]This example is briefly covered in (Gärdenfors 1993*a*).

our categories and figure in our perceptions and cognition. There are even more explicitly Kantian analogies to be drawn when we consider research in the empirical brain sciences.

Importantly, is is obvious in the case of CP that the categories resulting from the underlying geometry are not indicative of an essential level of explanation of our category use involving just these categories. They are manifestations of certain informational dependencies underpinning the surface features of out conceptual structure. Thus, the level of categorical perception is structurally misleading as to the causal factors in the influence on our thought and language. I note that this begins to address the concerns I expressed in Chapter 4 regarding the evidential status of the features of out language. This subject will be shown to be a central point of balance of my whole argument.

5.3 The Implications of Theories of Coordination

Investigation of the structure of the brain has grown considerably in recent years. If one were to expect corroboration of the form of idealism initiated by Kant, brain research is exactly where one would expect to find it. Starting around fifteen years ago, two neuroscientists, Andras Pellionisz and Rodolfo Llinas, began to publish work concerned with coordination in the cerebellum. Their concern was a post-Newtonian model of space-time representation in the brain. In an important sense, this work is contrary to Kant due to the well-known criticism to do with Kant's insistence on the *a priori* nature of Euclidean space representation in the mind. The development of non-Euclidean geometries tended to cast doubt on Kant's rather extreme insistence on the existence of certain synthetic *a priori* truths. Thus, when we deliver a model of integrated space-time representation, we go beyond Kant given his pre-Einsteinian idea that the two are separate and give rise to distinct categories. However, the general Kantian similar-

module for resolving time references, one for spatial position etc. in our theory and suggest this as a model of how we actually *do* manifest our semantic regularities. However, the seeming separation we manifest is not worth anything as a guide to how the brain manages to facilitate language and its semantic regularities. Separability in Einstein's physics is a special case illusion that is stable only within certain limits. Likewise, we cannot be sure that the separability of space and time suggested by language is not an illusion rendered by certain limits on the *causes* of the phenomena. Indeed, we have the rather uncomfortable evidence from neuroscience that a division between space and time representation in the brain is a myth. Time coordinates are blurred and space-coordinates are relative to a time-frame. Pellionisz and Llinas give an attractive metaphor for the situation in likening the situation to a group of high-speed battle tanks being coordinated by horse-cavalry messengers; the messengers are slower than that which they carry messages about.

Now, phenomenologically, in normal experience, we have a rather unitary sensation of time and space which can be used to perform many very complex tasks of coordination etc. This implies that the brain must perform some very complex intermediate tricks in order to present things in this way. Our experiences are of separable space and time positions but we know from neuroscience that this is an illusion not adhered to in the brain. So, the very fundamentals of our view of reality must be in some way engendered by the machinations of the brain. One could hardly ask for more Kantian evidence regarding the very foundations – space and time – of the transcendental framework. This is of pivotal importance in my comments and assessments to come. It is an empirical support of the simple and rather obvious observation that inspection of manifest features is of no use whatever in determining fundamental causes of such features. This is a more general form of the familiar skepticism about introspection and its evidential role. Chapter 8 considers the overall form of the assumptions in-

volved in traditional formalism with respect to the significance of manifest features.

More evidence is provided when we examine the nature of motor coordination. Take the example of a man catching a ball. The task of coordination is to manipulate the body in such a way so that the hand coincides with the ball at some point and in a state suitable for catching it. The ball, for us, exists in a four-dimensional space, three space and one time which appears orthogonal; that is, it appears that these coordinates are independent and that change in one is completely separate to change in another. However, the required position of the body is not described by a space so simple. The degrees of freedom are large, there are many muscles at work and they depend for their position on each other. Thus the space necessitated by them is many dimensional and complex. The task for the brain then is to map between these spaces in order to define the place of the ball in the motor-space, thus enabling the man to catch it. Unfortunately, the motor-space is what is known as *overcomplete*; there are many, many ways of embedding a lower dimensional space into a higher; there are many ways of catching a ball. However, we manage to perform the task very quickly and in many cases, in an expert fashion. Given that the task is overdetermined, again, the brain must impose some restrictions in order for us to be able to act in real time. There is not time to process all of the possible ways of catching a ball: all of the possible permutations of muscle tension and limb position, and then to choose the most economical.

It is interesting, and important I think, here to digress slightly and draw a parallel with work in formal semantics, particularly within the AI community during the 1960s and 1970s. Much work was done in order to address the problem of computational explosion in search paradigms. In a sense, we can recast their problems in these terms; the task of deciding between logically valid inferences regarding a situation and coming to choose the most appropriate one is exactly the problem of attempting to overcome

overcompleteness. There are many ways to, as it were, put blocks on top of other blocks and to infer how to remove a bomb from the room but we act so rapidly, search until one strategy is decided upon is simply not pragmatically feasible. Thus much effort was spent on attempting to develop formal systems with restrictions built into them. Non-monotonic logics, circumscription and restricted quantification were all used at one time to attempt to address this problem. Their aim was to restrict the search space and thus time taken to act; it was, in effect, to render a task less overcomplete. The problem with this sort of approach is that the time taken to recognise and process restrictions balances the time saved in following unpromising paths. The reason for this is the formalisms chosen. As mentioned in detail in chapter 3, essentially the illusion of the meta/object language distinction when it comes to real world implementation of certain formalisms means that overcoming overcompleteness by more rules, sentences or whatever of a similar formal system means no difference in real world plausibility. The mistake is to suppose that overcompleteness is, in itself, the problem. It is actually that overcompleteness ensures that you take too long unless something is done about it. Thus, one can solve overcompleteness within the constraints of traditional formalism at the expense of creating more processing elsewhere that ensures you have made no progress in addressed real time action. The problem is simply speed and not the merely technical consideration of overcompleteness which is only a particular method of hindering the former.

Again, to maintain the phenomenology of our usual perceptions whilst dealing with these sorts of considerations, the brain must, in some way, impose some structure on our experience. It seems as though space and time are indeed fundamental conditions of our sensible experience, seen in this way. The categories we might predicate of our phenomenal perception would, on this picture, be a product of these fundamental conditions, in the required Kantian sense. An important matter now is to describe how the

modern neurological theory accounts for the brain's action in these matters. This will give us a picture of the mechanisms of category imposition and also give us some information that will be relevant, as promised, to doubts about the foundation of traditional programmes in formal semantics.

5.4 Tensors and Invariant Relationships

A "space" in general, is merely a set of possible combinations of qualities. We might have a space of possible colours, or types of car for example. Each point in a space will have a value for a specified number of quality dimensions. For example, a car might have values for top speed, engine type; a colour vaules for hue and saturation. These quality dimensions are said to be "axes" of the space and the number of axes determines the "dimensionality" of a space. For example, a colour space typically has three dimensions of red, green and blue values. Each point in this space defines a unique colour and our general categories of "green, "yellow" etc. correspond to regions in the space. In order to map from red, green and blue value triplets in such a space to another specification of colour space using dimensions, for example, of cyan, magenta and yellow, we would simply take the colour specified by the red, green and blue components and decompose it into its cyan, magenta and yellow components. The colour remains invariant, the values along the dimensions in the two spaces are different. This is simply a case of *redescribing* the same thing.

If we map between arbitrary spaces with arbitrary dimensionality for example, we would not expect certain things of points in such spaces. For example, if we had a point in a four dimensional space (three space and one time) that represented a ball in flight, the actual values along each axis we would not expect to remain the same when the point was represented in terms of a fifteen dimensional space, each dimension of which was the tension of a certain muscle or the angle of a certain joint in an arm

in the act of catching the ball. However, it is a matter of mathematics that certain transformations between such different reference frames maintain certain relationships. If the hand coincides with the ball in one space then there are certain sorts of transformations that will guarantee that they coincide in the other space. It is not determined *which* point since the axes and scale of the axes are so different, but the fact that they coincide will be preserved. This is a simple case of a *transformational invariant*; a relationship that is preserved in character between changes of reference frame. The example above of the colour remaining the same through the change of reference frame (from red, green, blue values to cyan, magenta, yellow values) is an example of a transformational invariant. The mathematics of transformational invariants is tensor analysis. A tensorial relationship is one which remains for all coordinate systems, regardless of their degrees of freedom (dimensionality), scale of axes or other complications which we shall discuss below. This generality of relationship has a distinctly Kantian ring to it and I think it is a particularly good paraphrase of Kant's "forms of intuition". The universal subjectivity of the conditions of space and time in the *Critique* is mirrored by the universal invariance of certain relations that the brain preserves. The actual neuronal implementation of such invariances is not important:

> "This approach implies that while the neuronal networks of a particular brain are individual, there exists an invariant geometrical property ... that is common for all networks."[8]

Tensors are technically a sub-class of "geometrical entities"[9]. Such an entity is supposed to be something that has an existence independent of the reference frames in which it might come to be represented. Tensors are

[8](Pellionisz & Llinas 1980)

[9]A phrase from (Kron 1939); a classic reference for tensor analysis of networks, even though his treatment is largely of electrical problems. It is known that Kron's terminology is rather idiosyncratic but I use it here for its clarity in expository purposes.

such entities with certain restrictions on their mathematical definition[10]. The general point of a tensorial treatment is that it specifies relations between points in arbitrary reference frames. An example will help to illustrate this. Suppose we have one-dimensional space representing the absolute lengths of pieces of a particular wood in metres. Then if we chose to represent the wood in another one-dimensional space whose axis represented weight in pounds, we would have different numerical values for the value for each piece of wood. However, the relation that one piece of wood was two-thirds as long as another would be mirrored by it being two-thirds as light compared with the other in the new space; the relation between the values for these two pieces of wood as being two-thirds is *invariant* with respect to conversion between these particular spaces. This is the heart of tensor analysis; it is a method of determining invariants for transformations of reference frame. The invariants of a transformation define a certain geometrical entity which allows one to map between the different reference frames. Now, the invariants of a system define this in a unique way, allowing a unique mapping from one reference frame to another. Thus, overcompleteness is avoided as there are many *possible* ways of representing lower dimensional spaces in higher but only one *actual* way; the way defined by the invariants of the system. The invariants *constrain* the mapping to a unique solution in the way that the ratio of lengths of the pieces of wood in the space above constrained the representation in the space using weights to a certain proportion. This way of treating overcompleteness will define a unique mapping that obviates the need for search through many possible mappings. Thus, relationships are determined by the geometry of the brain and we have grounds for saying that Kant's "conditions of sensible experience" are accounted for by possessing a common physically embodied tensorial relationship between all of the manifestations of

[10]Technically, tensors do not cover relationships that depend on certain functions of the related elements.

fundamental brain-imposed structures.

It is this notion of "invariant", defined above, that suggests a certain approach to understanding language and thought in humans. If the invariant features of our language and thoughts are, in essence, constancies in the mappings between references frames that the brain performs in its everyday operation, then we have a naturalised basis for our common ontologies. We might define what we phenomenologically take to be an "object" thus: take all reference frame mappings. An invariance of a certain relation between our input information is what we call "objects". A "property" might turn out to be an invariance of a different sort etc. This is a particular instance of the general way in which systems can be defined in terms of the invariances they give rise to. Kron's work on electrical systems demonstrates in detail how we may characterise a system by the invariances that hold between the different frames of reference that model it throughout its operation[11]. It also demonstrates that the invariants of a system are usually extremely abstract. Typically, for electrical networks, the invariances are complex relations between currents, voltages, fields and the like. This means that there is no intuitive and obvious way of classifying a particular system, hence the need for a mathematical framework.

5.5 Holism and Dependence

As mentioned above, one further matter implied by the approach we have been discussing needs to be made clear. This matter is, I think, of supreme importance for Cognitive Science and its attitude to logical formalism. It seems very likely that the sorts of reference frames that the brain employs to perform motor coordination are not *orthogonal*. This means that the degrees of freedom a system has are not independent of each other. Pictorially, this corresponds to a space having axes that are at more or less than

[11](Kron 1939)

a 90 degree angle to other axes. This involves then, informational dependency. This means that the components representing an invariant will not be "separable". This, in essence, is to suggest that the elements that occur in such spaces are not decomposable into separate contributions from entities in the world. If each contribution to a geometrical entity (point, line, area etc.) in a space is dependent in complex ways on all others, then there is no way to single out independent "qualities" that are constitutive of the building blocks of the points in the system[12]. Pellionisz and Llinas point this out for the case of motor coordination:

> "Thus, while a goal of the CNS is to establish an external coincidence of events, this goal has to be achieved by using space-coordinates inside the CNS such that each of them refer to a different external time-point."[13]

That is to say, the space and time coordinates are not separable. Now, if this sort of complexity is necessitated by motor coordination, I argue that it is reasonable to suppose that much more complex and evolutionary posterior phenomena – language and thought – will certainly involve at least this level of complexity and will thus necessitate non-orthogonal (sometimes called "oblique") reference frames.

It is worthwhile, then, to explain exactly what this means in terms of language and thought. Firstly, it is an empirical matter to determine what relationships are preserved in transformations between reference frames. One needs to look at the operation of a system and investigate what sorts of things remain constant. This in itself means that the fundamental defining features of a system are not really open to *a priori* theorising. The manifest features of a system do not give any good guide to the actual causal processes that give rise to it since the characteristics of it will belong to a system whose constitutive components are inseparable, for mathematical

[12]This is because the information is *contravariantly* composed; see below
[13](Pellionisz & Llinas 1982)

reasons. If this is true, then attempts at "decomposition" of meanings and the like are simply inappropriate. The surface manifestations of the activity of brain systems is opaque to this sort of investigation. Investigation has to be of low-level systems, their reference frames and the invariants that they define. Secondly, the types of invariants we would expect to find are nothing like the normal "explanatory" features of a formal system. The sorts of things that remain invariant between reference frame transformations are generally very abstract; the case above of simple coincidence is an atypical example. For example, in geometrical systems, invariants tend to be things like complex relations between lengths, angles etc. These relations generally are not intuitive, have no correspondence to obvious features we recognise and can only be determined by examination of the system. This is a facet of that which was mentioned at the beginning of this chapter; that the normal explanatory desiderata of the formalist will not be met but rather side-stepped. The formalist tends to like a theory where each element has some describable correspondence with elements of the area to be explained. For example, sentences might correspond to "propositions" in the model, sub-elements of sentences like adjectives have a corresponding model element like "property", nouns are "objects" and so on. Each lesser element of the model has a corresponding and often intuitive correspondent. Indeed, one of the oft-cited reasons for moving on from Montague Grammar is that its theoretical constructs become "unintuitive". Language is "intuitive" and thus we are fooled into thinking that its causes must be. This desire for a one:one mapping between domain and range of a theoretical system is, I think, quite obviously a desire caused by features of the formal systems adopted to describe the theory. This is discussed more fully in Chapter 4.

If I am correct about this, then there simply is no way of describing in language, contrary to the practise of logical formalism, the basic features that contribute to our use of language and thought. The basic features are

extremely abstract properties of informational relation, combined in an inseparable manner. Formalists are wont to criticise "network" approaches to language and thought for failing to provide sets of basic building blocks and modes of combination. Apart from this criticism begging the explanatory question, the tensor approach has a good argument that this is simply not relevant. It is necessitated by the type of systems required by the brain and implied by empirical research that the sorts of explanatory material traditionally desired are inappropriate and impossible.

5.6 A Key to the Historical Problem

A "metric" tensor defines the notion of "distance" between points in a space and thus allows one to translate between different reference frames by translating the geometrical notion of distance from one point to another into terms applicable to a new reference frame. In our common encounters with geometry, and in the simple systems designed to suggest "geometrical" semantic theories (see Chapter 7), the metric tensor is merely the identity matrix as transformations from one Euclidean space to another of the same dimensionality and same axes require no special mapping. In spaces where the metric tensor is the identity matrix (these are spaces where the general rule for the distance between two points reduces to the Pythagorean square law), we lose, however, a distinction crucial to a general tensor theory and more importantly, crucial to an understanding of the nature of semantics as perceived as manifestation of brain processing. Indeed, it seems to me that an explanation of why the formalist programme was led into error can be constructed on the basis described above in association with two concepts typically lost in such simple models but which are central to general tensor theory. The components of points in a reference frame can be of two types. They can be *covariant* or *contravariant*. Technically, this means that the components in one reference frame are determined in another ref-

erence frame by one of two different methods. However, the feature that interests us here is rather the nature of the separability of components. A point with covariant components has components that are decomposable into their separate values independently but which do not combine, by the simple parallelogram rule, in the usual geometrical manner, to create the point. A point with contravariant components is not decomposable into its unique separate values but its components do combine to give the point. Figure 5.1 summarises this difference. As mentioned above, in simple geometry, this distinction does not exist. The covariant components of a point in space are exactly the same as the contravariant components since the axes are orthogonal (at 90 degrees to each other) and thus the metric tensor is the identity matrix. The disappearance of this crucial distinction in simple reference frames will be of importance when examining recent "geometrical" approaches to semantics in chapter 7. In terms of semantics, I see

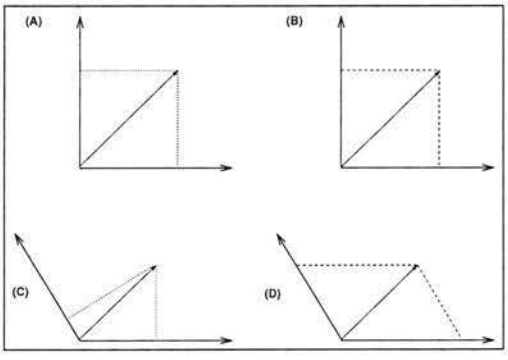


Figure 5.1: Summary of covariant and contravariant vector expressions in orthogonal and oblique systems. Dashed lines **(B)** and **(D)** represent contravariant components which combine to create the vector by the common parallelogram rule but which are dependent in oblique reference frames. Dotted lines **(A)** and **(C)** represent covariant vector components which do not combine to create the vector but which are independent even in oblique reference frames.

the implication of this as the following; certain thoughts or phrases can be represented in two ways. Either they are a homogeneous whole that cannot be broken down into contributing parts but which the brain knows how

to construct – they are contravariant information. Or they are separable into parts which do not combine to give that which they are parts of – they are covariant. The "decomposition" in the case of covariant information is nothing like the sort of thing commonly so-called within formal semantics. The decomposition is simply a way of chopping things up so that the information may be reused in some way to build behaviour – contravariant information. Pellionisz and Llinas' work supports this as they view sensory perception as taking in covariant information and resulting behaviour as outputting contravariant. So, behaviour is not decomposable into that which composes it as the information is contravariantly conceived. What sort of examples do we have of this in nature? A good example is the movement of points on the face of a cube. If the cube is made of a rigid material, there will be no movement along certain axes of points on faces orthogonal to the direction of movement. However, if the cube is of an elastic material, the general inertial stress of movement in one direction will have an effect on the movement of points on orthogonal faces due to distortion caused by acceleration. Put another way, in an elastic medium, the position of a point is dependent not only on forces in the direction of planes in which the point lies, but also on forces that would be irrelevant in a rigid material. Charting the movement of points on a cube of jelly is more complex than charting the movement of points on a cube of iron.

Sensorially, we chop covariant information up along the axes of the reference frames that the brain embodies and then need a way of eventually constructing a contravariant representation that can actually combine components into a behaviour. Dysmetric lack of motor coordination can be modelled as failing to do just this. It transpires that it corresponds to the mistakes that one would expect if attempting to construct behaviour using covariant components directly without transforming them into contravariant counterparts[14]. That which converts between covariant and con-

[14]See (Pellionisz & Llinas 1980, Pellionisz & Llinas 1982)

travariant information is the metric tensor. This is a geometrical entity that essentially defines a metric on reference frames. Lack of motor coordination then corresponds to bypassing the metric tensor. The situation is like measuring the size of an angle in radians and then attempting to reconstruct it by taking the radian measurement to be normal degrees. I argue that the trouble formal semantics has had in the history of AI and Cognitive Science is an aspect of this. Whereas motor coordination problems are due to taking covariant information as contravariant, formal semantics has taken contravariant information as covariant. This obviously needs some explaining.

Our behaviour needs constructing. It is universally agreed that language and thought are phenomena possessing components; it is the actual constitution and nature of these components that is at issue. If we accept the above theory, then it is clear that our language utterances and examples of our thoughts are essentially *outputs* of certain brain functions. Outputs have to be constructed and thus must be contravariant in nature. Only contravariant components can actually combine to give an output due to the simple geometry of the well-known parallelogram rule. It is also not particularly controversial that our *inputs* must be analysed in some way by the brain. Information must be extracted or gleaned. If this is the case, sensory input must be seen as covariant; covariant components can be independently determined from different sources which are embodied in the modalities of sense experience. Now, without knowledge of the exact composition of the metric tensor mediating between covariant and contravariant representations, there is no way of decomposing contravariant information nor of constructing correct outputs from covariant. However, this is exactly what formal semantics attempts to do. It attempts to break down our linguistic output into contributing and orthogonal components and then to build back up to language. Simply put, there is a suspicion that we are in danger of taking what we do to be data in the investigation of how

we do it. Here we have, in the form of a tensorial model of general brain function, good reasons to suppose this cannot be done. If what we do is the result of a very complex and interdependent system of information combination, then taking this as the basic data in an attempt to explain how we do it simply will not work. It means that the concepts that we may use to explain how we do what we do are completely different to those involved in describing what we do. The key to relating these descriptions would be, in this theory, the metric tensor relating covariant and contravariant information. Determining this is not a matter for formal semantics however; it is a matter for low-level brain research of the appropriate sort. The essential interdependence of the information combined in the manifestations of our brain's processing means that a theory that attempts to take its cue from these manifestations is fundamentally mistaken. The key to the patterns in the manifestations of the system are rather in the embodied metric that allows information to be analysed and reconstructed. This key is not open to *a priori* speculation, rational reconstructions, formal redescriptions or any of the common strategies for analysis in formal semantics. It is a purely empirical matter to determine the invariants defined by such a key and thus a matter for neuroscience and brain research. It is, because of this very fact, impossible to give concrete examples of the invariants involved in the geometries contributing to our characteristic uses of language and our common patterns of semantic inference. The invariants would be couched in purely mathematical terms best paraphrased by giving general features of patterns manifested as a result. The biases of the formal systems are against this sort of explanation since it does in no way mirror the manifestations of language. One of the main purposes of this work is to remove the basis of this prejudice as it rests on the built-in assumptions of a certain formalism.

The ramifications of the fact that the outputs of a system conceived of in this manner are not a guide to their causes are, I think, enormous.

It is, indeed, the fundamental theme of this work to explore the implications of this phenomenon. When the manifestations of a system are only understandable with reference to the structures mediating between its inputs and outputs, the outputs lose their attraction as a starting point for theorising. Also, the necessity of postulating *sui generis* levels of explanation with regard to such a system are seen to be unnecessary and simply a product of the mystification involved in wondering what on earth to do otherwise to satisfy the requirements of what we have come to see as desirable sorts of "explanation". These themes are to take up the majority of what follows. The next chapter is a critique of a more philosophical and abstract expression of the stratified viewpoint. It demonstrates the underlying assumptions clearly and I relate them to work in Cognitive Science in order to provide an important conclusion to this thesis; that much work in Cognitive Science is infected with the assumptions of a general philosophical position which prevents it from achieving many of its aims.

In this chapter, I have discussed tensors as analogs of clasical vectors. This is because vectors can be considered as tensors of "rank one". This is a common convenience employed since Einstein (see (Sokolnikoff 1951) p. 61 and particularly p.62) which I shall relax in Chapter 8 where I draw out my views of the consequences of a tensorial approach.

Chapter 6

The Problems of Stratified Theory

It is important to set the problems with the space of solutions for traditional formalist concerns within a wider context in order to obtain a better view of the central issues. Within the more philosophical community, there are examples of a practise, not obviously, but I maintain, fundamentally related to the practices of the formalist enterprise in Cognitive Science and its concerns for natural language semantics in particular. The attitude we are concerned with involves a certain approach towards aspects of the mind and their explanation in terms of reductionism. It is the task of this chapter to explore this other side of the coin and to show that the mistakes here are merely different but more fundamental aspects of the same problems that formalism in natural language semantics suffers from in more technical and specialised instances.

We might term the general philosophical approach underlying the formalists' strategies as a *stratified* one. There is a basic belief that the mind, language, thought or whatever cannot be properly explained without postulating different "levels" or "strata" of processing, conception or whatever. A typical suggestion might be that we can only understand language use if we do so on a model involving different "levels of representation" or "realms".

A rather basic view of this form is that of Pylyshyn's "levels of explanation" where systematicities in our data supposedly necessitate a level that exhibits the required systematicity in order to correctly account for the phenomenon. This is too simplistic, obviously. A systematicity in our data may well be an artifact of how we choose to look at it; everyone working in semantics has had the experience of playing with attractive formal systems and then encountering the daunting face of real language and wondering where all the systematicity of the model domains disappeared to. Also, to merely adopt a level of explanation that contains exactly the putative systematicity we need is to rule out a huge space of possible explanations that do not "explain" systematic features by simply mirroring them. For example, if we notice that "A" always follows "B" in spoken language, we can certainly "explain" this by generating a level of explanation using a formalism that ensures this. However, this is a rather degenerate case of "explaining" and is surely not something *necessary* for explanation of a phenomenon. Mathematics uses very abstract building blocks to give explanations of phenomena without merely incorporating the regularities into the description of its formalism. To mention a theme of other chapters, adopting a compositional formalism to explain putatively compositional language semantics is like inventing a new mathematics to explain a new phenomenon when combinations and abstractions of existing mathematics will do perfectly well. This latter strategy also facilitates connections with existing explanations. So, the penalty for inventing other levels of explanation is quite high; we cannot obviously link the resulting theories in with existing knowledge – exactly the problem formal semantics has in addressing the brain sciences for example.

However, in addition to this sort of stratified view, there is another that seems on the surface to draw from a different concern. There is a long-standing problem since Kant and the significant post-Kantians regarding the relation of the mind or brain to the world. What is the relation and

how do we explain it? How much of our picture of the world can idealism claim? The tension is between allowing complete idealism and thus no role for a reality-in-itself to play and complete realism which allows the possibility of an unpalatable reductionism. Since Kant, philosophers have been desperately looking for a middle ground between these two positions, mainly because Kant failed to provide convincing arguments for noumena and thus leaving us drifting towards complete idealism. It is important to see, in general at first, how this fits in with the stratificational theories mentioned above. The point of those who wish to hold a middle ground in this debate is that complete realism is denied because there is a level or a "realm" which is irreducible to physical reality and thus which necessitates a different sort of explanation. This is not really an advocation of dualism, Cartesian or otherwise, merely a weaker claim that a certain sort of explanation is necessitated by a certain sort of phenomenon. The sort of explanation is familiar; it is one couched in terms of formalisms dealing with "reasons" and concepts. Thus, there is a commonality between this sort of philosophical predilection for a style of explanation of the inherently reasoning part of man and the formalists' argument for a formal level of explanation. At the end of the next Chapter, I shall discuss an approach to a closely related problem, the Symbol Grounding Problem, that mirrors in Cognitive Science the difficulties that are present in the more abstract treatment of this chapter.

What is the philosophical root of this sort of bias? In an elucidation of this topic, I shall draw on the work of John McDowell, particularly his *Mind and World* as it gives a clear exposition of the philosophical basis of the desire for the orthogonality of explanations in this area. McDowell's arguments are typical of many anti-reductionist arguments by the way in which the central motivations of the position are rather coyly mentioned in passing. This weakness shows up in the internal contradictions which are forced on the theory as it tries to reconcile a link between different

levels of explanation without having the levels collapse into one another. The task of the philosopher and the formalist here is to render the levels of explanation close enough for their relation to be possible but far enough from each other for them to remain distinct as explanatory levels. In this, I shall argue that McDowell variously fails on both counts and shall point out what parallels with formal semantics this failure exhibits.

6.1 Realms and Explanation

The task McDowell sets himself is to accommodate Kantian idealism into a picture that avoids the lack of "friction" with the world necessitated by complete idealism. He also wants to avoid the opposite where the connection between the mind and the world is so close that both realms are resolved into one. He terms this "bald naturalism". McDowell espouses a dichotomy between the "realm of reasons" (ROR) and the "realm of law" (ROL). The former is that which is characterised by Kant's notion of "spontaneity" – the possibility of combinations and syntheses of concepts in order to create new concepts, independent of the sensory impacts of the external world. The quality is "spontaneous" as it does not, as does perception, rely on an external impingement from without; judgement is seen as an internal affair, capable of regulation and manifestation from within its own sphere. When the causes of manifestations originate from two distinct places, we may, it is argued, consider them separate and *sui generis*. There is a certain realm of explanation that makes reference to the "representations the mind produces from itself" as opposed to those it makes as a result of sensory experiences. It is the former that constitute the ROR and the latter, the ROL. The ROL is that of external reality. The realm in which experience occurs and where the mind/world interface exists. This distinction would be the same as the real/ideal if it were not for the import of the notion of "justification"; there is no easy demarcation between the ROR and

the ROL since Kant, as we cannot now see justification for beliefs and action as coming from two sources, the world and the mind. Since Kant, the epistemological difference has ceased to exist and the ontological difference come to be questionable. So, the difference between the ROR and the ROL is that between levels of explanation in the justification of belief and knowledge. We must admit post-Kant that the space of possible reasons for something is no larger than the space of concepts. The "Myth of the Given" is exploded and we may now no longer hold that some non-rational access of "given", direct contact with the world justifies and explains certain states of mind. So, the distinction between mind and world is blurred at the very least. What McDowell aims to do is to prevent the distinction becoming too blurred; in becoming a non-distinction. This, he thinks, would be to collapse the ROR into the ROL and give in to bald naturalism along with its concomitant reductionism or physicalism.

We can see here a similar strategy to that adopted by the formal semanticists. There is a desire to prevent the lion's share of explanation going to a "low-level" theory. This is motivated either by noting features that are putatively only explicable at "higher" levels of description or by more generally holding that the higher level of description is ineliminable for other reasons. Two responses are generally possible to this sort of position: firstly, to show that maintaining the stratified theory leads to contradictions and secondly; attempting to remove the basis that supports the necessity of certain strata for proper explanation. We proceed along these lines in turn.

6.2 McDowell's Approach

In order to achieve the results he wishes, McDowell has to reject Davidson's position that the world exerts a merely causal influence on the mind. This, he thinks, would collapse the whole picture into the ROL. If there is

nothing in the relationship between mind and world that is essential to the ROR, then there is no good reason to hold a stratified account involving it. So, for McDowell, the world must assert a *rational* influence on thought. This is a general way of expressing the formal semanticists' belief in the necessity of logical formalism in the explanation of language. Something more than a merely causal influence implies more structure; McDowell accounts for this by postulating the ROR and formal semanticists by the structural expedients of formal systems. A rational relationship is inherently more *structured* than a causal relationship since reasons are, by their very nature, *composed*. A reason is something we present as an argument, evaluated by its clarity and the components of which it is made, thus, a rational relationship is one that can be broken down according to the canons of rationality: validity, clarity, soundness and the like. If the influence of the world on our minds or brains is not a rational one, then it follows that it lacks the canons of rationality and thus appears to us to be inscrutable. This is, to many, intolerable as it suggests that man's place in the world is not determinable as his links with it are forever obscured. The belief in a rational influence on thought, a rational connection with the world is thus a desire for a picture of the activity of mind that shares these features. The activity of mind includes language and thought and thus it is natural that we would, if following this motivation, come to warmly regard a technical system that embodies the characteristics of rationality. This is exactly the desire for systems of formal semantics. McDowell's tactic is to remove the clear boundary between the conceptual and non-conceptual, as indicated by Kant, while maintaining that the ROR – the conceptual – is *sui generis*. Thus, the boundary must go but the distinction must not wholly collapse. Again, for formal semantics, the distinction must not be too hard otherwise there is a risk of independence from facts about the brain and this threatens dualism of various sorts. However, the necessity for formal systems must remain, even though they must relate to the low-level work-

ings of the physical organism somehow. It is exactly the same tension in both cases and the way in which McDowell attempts to approach this is, I think, very illuminating as to the difficulties for formal natural language semantics and to an understanding of the essential tension in traditional Cognitive Science.

Now, how to remove the distinction between the conceptual and non-conceptual? This is, in different terms, the age old question in Cognitive Science of "how does a reduction of language to brain states work?". Well, at first, it would seem that there is a fundamental difficulty as our concepts are relatively coarse as compared with experience. This is part of the famous rejection of standard AI by Dreyfus; that much knowledge is not conceptual and thus not amenable to formal analysis. There is, it is thought, "extra information" that concepts do not capture as they are essentially abstractions from the real world and thus less sensitive to detail; indeed, this is part of the attraction of the notion of a "concept" as it means it can be applied in many situations. The extra information involved in experience would seem then to be, *ex hypothesi*, non-conceptual. This would, in turn, imply that there is a necessary and fundamental distinction between the conceptual and the non-conceptual, due to the possible informational detail of the conceptual, of the ROR. In McDowell's terms, the ROL would then be necessarily non-conceptual and thus we would need to rely on a version of the Myth of the Given; would need direct access to certain information. This is something that Kant and any following realistic epistemology forbids.

If conceptual granularity of information is fundamentally different from non-conceptual granularity, then we have a difference that blocks the move that would reconcile the ROR and the ROL and thus we are led to idealism as McDowell despairs of the only alternative: a merely causal connection. The suggested solution is to not suppose that conceptual granularity and experience are independent. We become conceptually adequate "on the fly"

and in real time, principally by ostension. We say "that colour" or "this" and thus make our concepts finer tuned as we act[1]. Thus, in this way the supposedly fundamental difference between the ROR and ROL is bridged by allowing experience to inform conceptual discrimination. This is how formal semanticists avoid having to specify the entire set of predicates in their systems; they assume that predicates are employed as they are found and thus an exhaustive lexicon is not a prerequisite of a theory. This seems obvious until one examines, in McDowell's case, the assumptions this requires and the contradiction that it implies.

In order for this to have the desired effect, ostensive content must be, in some way, *rationally* related to conceptual content. Any lesser link will not do as it would not, as shown above, suffice to prevent a collapse of the ROR/ROL distinction. Now, conceptual content is conceived, in an explicitly Quinean manner[2], as being conceptual due to its place in a network of capacities joined by their possible employment in spontaneous judgement. McDowell asks the rhetorical question "why should short-lived ostensive content not be rationally integrated into spontaneity?". The best that McDowell can do in answer to this is to argue that unless is it indeed so integrated, we will not have a rational link. This is transparently question begging.

> "If those impingements [of ostensives] are conceived as outside the scope of spontaneity ... then the best they can yield is that we cannot be blamed for believing whatever they lead us to believe, not that we are justified in believing it."[3]

Here, McDowell is appealing to a notion of responsibility in order to support an argument purporting to establish a conclusion having much wider reaching implications. It is an argument for the rational linkage of the ROL with the ROR based on a completely orthogonal desire to correspond to an

[1] See (McDowell 1994) pp. 57–58
[2] See footnote on p. 13 in (McDowell 1994)
[3] (McDowell 1994) p. 13

idea of responsibility and justification. This is like the formalist replying to the question "but what about the link of these formalisms to the physical brain?" with the retort "there must be a link otherwise the formalism is vacuous". The reply in both cases is question-begging and ill-motivated. Furthermore, in direct reply to this issue, there are good reasons why ostension *cannot* furnish us with the required rational link with the ROR.

Repetition of experience, familiarity and history of usage are features that result in content becoming embedded in a conceptual network. Conceptual content is that which is so semantically worn that it constitutes a Neurathian boat in which less worn content floats. It is precisely because this sort of content is *not* short-lived and transitory that it is conceptual. Ostension therefore cannot possibly have the desired *rational* links with the conceptual since rational links are in the remit of the ROR which, at the very least, excludes this aspect of the ROL. Ostension is a way of *overcoming* the inertia that conceptually embedded information has. It is not something that provides a necessary means of rational relation to such information. If ostension could provide the link that McDowell thinks it can, then it would not be part of the ROL, would not be essential to the non-conceptual and thus would not relate the ROR to anything. Idealism would then result.

This problem can be emphasised from a different direction. McDowell's theory requires that experience is in a sense passive. By this, he means to ensure that there is some friction between the conceptual and the world, thus avoiding idealism; if conceptual capacities are thoroughly active all the way out to the point where, in the Quinean idiom, they touch the world, then there is no constraint on their application from without the conceptual. But, to have spontaneity restricted so that it does not extend all the way out to experience is to fall victim to the Myth of the Given. So, we require that some, but not all, aspects of the non-conceptual are common to spontaneity, thus making a link but not collapsing the distinction. The

short-lived experiential capacities, typified by ostensives, that are required to link the separate realms thus are passive but are required to share some recognisable feature of the spontaneous in order to perform their theoretical duty. This content needs to have some real conceptual characteristic in order for it to be correctly dubbed, in part, "conceptual"; this must not be mere "word-play"[4]. What, then, is that which allows experience a flavour of the conceptual, of the spontaneous? It is simply being involved, elsewhere, in spontaneous judgements and thus having conceptual relations.

To recap then, the fine-grained experiential capacities that seemed to demonstrate an irreconcilable difference between the conceptual and the non-conceptual are not insurmountable as the gap is bridged by recalcitrant experience making the conceptual more sensitive in real-time. It remains a passive act but connects with the conceptual by having rational relations forged in spontaneous judgement elsewhere. So, such experiential content is conceptually linked parasitically upon its conceptual use in other places. Now, presumably this "other" conceptual usage is mostly in the future. If it were not, a history of previous spontaneous usage would mean that we would indeed be indulging in mere "word play" in dubbing it "non-conceptual" at all. Such a history of participation in spontaneous judgement would mean that there was no gap in conceptual and experiential granularity that it would be needed to bridge. If I ostend many times a new colour that I am conceptually unable to accommodate, then it becomes part of my conceptual apparatus and thus this is modified and the ostension looses its non-conceptual character. Indeed, it is not too radical to suggest that this is how many concepts are formed. This leaves McDowell forced to say that the necessary spontaneous-like feature of experiential content is only such due to its hypothetical use *in the future*. This is indeed mere word-play. One cannot relate A and B by something that has the necessary property only hypothetically. That is like saying that x relates A

[4]p. 13

and B because if it were used in a certain way in the future, it would have the property that would enable it to do so. The passivity McDowell requires to prevent idealism seems damaging to the project of providing more than causal links between the ROR and ROL for the simple reason that passive content is, *ex hypothesi*, not rationally integrated and thus not conceptual. Thus the link cannot be rational. It seems to be clutching at straws to attempt to render experiential content conceptual by appealing to a future active use of the content. If such content is passive, it seems barred from becoming conceptual in addition to maintaining its passive nature. If it is not passive, the ROR collapses into the ROL and one has bald naturalism.

After this argument, we need to step back and see what has been said regarding the fundamental issue about stratified systems and their approach to semantics and language. We have seen that the attempt to maintain a difference – one that does not lead to idealism – between a level of explanation in terms of low-level laws and one in terms of rationality and spontaneity is doomed to collapse. This is due to the tensions between the requirements of friction with the world and fear of too much friction that causes, so to speak, any "higher" level of explanation to melt away. In modern semantics and linguistics, we have analogues of the requirement of spontaneity in the observation that language seems to be very productive: we can create a supposedly infinite variety of sentences. This is exactly the modern formalist echo of McDowell's more philosophical concern for a realm of ineliminable spontaneity. McDowell's desire for this Realm of Reason is exactly the formalists desire for, say, a recursive formalism. The formalist has a typical set of features that are supposed to be definitive of the higher level of explanation. Compositionality is a favourite. Their task is, like McDowell, to keep this while not alienating themselves from the attachments to the low-level data which would render them irrelevant to a study of human cognition. There is one level, the "low" level of explanation, usually in terms of physics, the brain or whatever which must be accommo-

dated. Additionally, there is supposed to be a "higher" level of explanation that accounts for features of the manifestations of the low-level operations. A significant characteristic of stratified views is that the higher-level explanations are taken to be ineliminable and are *sui generis*. This is the most important point and indicates that this particular level of explanation demarcates a particularly necessary part whose omission would render the theory useless. It is mainly the *characteristics* of the higher level of explanation in which the *sui generis* nature of the system lies. It is an essentialism with respect to some of the patterns manifest in our language and thought. This essentialism serves a twofold purpose. Firstly, it provides – the question of whether it does so honestly can be left until later – a virtual basis of data to be accounted for. We reify the patterns and hold them to be data to be explained; inputs to our formal systems. The second and more obvious reason why this essentialism of manifest patterns is so important is often mentioned; if one's theory is reducible to another, it is not as interesting, a whole "depressing" ontology of physical facts is the only "real" description if the higher levels are eliminable etc. McDowell, indeed, lets slip the latter concern in his comments about the unromantic overtones of a collapse of the ROR into the ROL. The formalists must hold that there is something that necessarily has characteristics of the formal level of explanation which must be invoked in an explanation of the relation between mind and the world. This is simply a manifestation of the underlying philosophical worry that the central apparent features of mind are not to be "explained away" or theoretically reduced but are to be be guaranteed by holding a theory that has such apparent properties as being *sui generis*. The properties are enshrined in a level of explanation said to be irreducible. The common problem is then to relate these *sui generis* properties to the real, empirical world, thus avoiding what McDowell calls "rampant Platonism". We have seen that McDowell pushes his ROR so far away from the ROL that it is impossible to provide the sort of link that is

required in order to make sense of our animal nature; the fact that we are embodied and subject to physical law. The ROL is certainly relevant, the ROR only tenuously so. In order to strengthen his case, McDowell paints a *sui generis* picture of the ROR so as to contrast it with the ROL but this provides for a lacuna between them that he cannot fill and thus he is led exactly to a "rampant Platonism" where the ROR stands alone and aloof from the world. Formal natural language semantics follows exactly this pattern in its insistence on the formal level of explanation. The strictures regarding its explanations are such that it has famously been noted that it fails in real-world application; the lacuna necessitated in a formal stratified theory is, again, unfillable because the distance required to have two, separate parts of the theory is too great to be crossed by something having characteristics of both parts without them collapsing into one. Famously, recursive formalisms, knowledge-base searching, production system rule-bases etc. are all at odds with the speed of real human action. One of the levels of explanation, it is clear, must go. The ROL is certain. We are more confident about the fact that we are subject to the laws of the physical world than we are about there being a special realm of reasons, a level of logical explanation for our mental cogitation. Therefore, it is only reasonable to drop the ROR, allow it to collapse into the ROL and settle for a "bald" naturalism, no-matter the irrelevant concerns about the unromance of it.

6.3 Second Nature

McDowell proceeds to expand on his solution by introducing the notion of "second nature" which he claims to derive from Aristotle. This is simply that our spontaneous judgement is so embedded within our environment that it plays as significant a part in our experience as that played by nature, which sits firmly in the ROL. We can, I think, see this as a reverse

of the strategy espoused previously. Before, McDowell was attempting to illustrate the rational links afforded by aspects of experience; he was trying to show that experience has features that link it to the ROR. Here, he is arguing that the realm of the conceptual has features that link it to the ROL. Our conceptual capacities have the quality of being second nature and thus link us to the primary nature of the empirical world. McDowell thus sees language as a medium for transmitting shared conceptions that become part of our mental furniture; it is primarily a cultural transmission device rather than a means of communication. The link between the ROR and ROL is thus forged in two ways, once by law-bound experience sharing some characteristics of the conceptual and once by the conceptual possessing law-like characteristics of nature.

Once again however, this argument only allows for a collapse of the ROR into the ROL. This second nature, derived from the ROR, is "naturalised" in experience and thus contributes to the content of experience. This, I think, is actually an attractive picture in one respect; it reflects the Kantian doctrine that our experience of the world involves more than the simple experience of an external reality. The conceptual indistinguishability permeates everything. However, it does not reflect the necessary concomitant which is that a decomposition into the contribution by the world and a contribution by the mind is something that can only be approached through a method, highly suspect to modern thinkers, of transcendental critique. McDowell aims to demarcate the two while allowing them to be so similar as to be mistaken for each other in experience. If the ROR can really furnish us with a second nature, then how can we establish that there is indeed a ROR? Introspection could not help for if it could, this second nature, this surrogate empirical reality would not have succeeded in disguising itself enough in order to carry off the link required to firmly embed it within our experience. This second nature is certainly not an *argument for* the ROR/ROL distinction. It is something that could not be perceived

for that would contradict its very existence. Second nature must always be perceived as nature or else it is perceived as artificial. So, the argument from second nature is attractive but has little to do with establishing the *sui generis* nature of the ROR. I argue for a different conclusion using the concept below. No, the ROR requires the notion of second nature in order to provide a link with the ROL otherwise the result is "rampant Platonism".

So, McDowell is forced into such a position as the following illustrates:

> "Given the notion of second nature, we can say that the way our lives are shaped by reason is natural, even while we deny that the structure of the space of reasons can be integrated into the layout of the realm of law."[5]

It appears, then that the ROR is non-natural but acts upon the ROL in a natural manner. It is *sui generis* but its action is fully naturalised. The paving of the road to dualism here is more than merely apparent. The motivation for this view is clear where on p. 87 of *Mind and World* we find

> "we cannot capture what it is to possess and employ the understanding, a faculty of spontaneity, in terms of concepts that place things in the realm of law."

This is one of the central motivations for the entire picture that McDowell espouses and it echoes the formalists concern that we cannot capture the central features of human languages without a formal system. However, the thrust of Kantian idealism is such that an argument from the features of the *explanandum* can only ever establish a much weaker claim such as

> "we cannot capture what it *seems like* to possess and employ the understanding, a faculty of spontaneity, in terms of concepts that place things in the realm of law."

This is uncontroversial and follows simply from the epistemological opacity of the origins of our perceptions. Quite simply, an agent's account of

[5](McDowell 1994) pp. 87 – 88

his perceptions is no evidence for anything other than how things seems to be and *that* is no evidence at all for a theory that wishes to establish the existence of *sui generis* but necessarily experientially indistinguishable realms of the mind. Formal semantics echoes this exactly in taking language to be evidence for the mechanisms giving rise to language. We seem to be explaining X when we perceive features in X and then design a level of explanation that is sensitive to those features. Really, we are merely redescribing the features that we perceive and bolstering up the level of explanation achieved by reification of the features that are manifest in the output.

Second nature is a useful concept as it embodies what Quine terms the "pollution" of the stream of empirical experience by every preceding experience. It embodies Kant's "conditions of sensible experience". It contributes to an epistemological opacity of causal determinations but can indicate nothing about ontology. The mistake is to concentrate too much on the "second" rather than "nature". The former seems to suggest something masquerading as the real thing. However, nothing is as clear as this. We have terrible difficulty in ordering things according to their "naturalness". What are the natural kinds? Some people think that chairs and tables are more "really part of nature" than the legal system, plastic bags or justice but no-one has a good idea of what is really natural and what is not. Since Kant, we hardly know what to make of the question. So, to have a second nature that is clearly "second", is to slip a distinctly incongruous element into a theory which claims epistemological homogeneity. One cannot have a truly "second" nature without some way of telling fakes from the real object and we cannot do this by taking our cue from the perceptions that result from the indistinguishable causes. Perceptions have patterns, language has patterns but these need be no guide at all to the patterns of the causes of these patterns. The main reasons that formal semanticists and many philosophers have been fooled into using language or the structure

of thoughts as a guide is simply because they are not aware of the applicability of ways of generating patterns from unlike patterns. Formal semanticists utilise systems designed around the notion of a canonical language, motivated by shortcomings with natural languages. Not very surprising then that the patterns their formalisms show are patterns mirrored in the language structure they claim to explain. McDowell sees patterns in the spontaneous actions of judgement; he sees a structured conceptual realm. He does not see this structure in the ROL and thus requires a *sui generis* level of explanation to accommodate these seemingly unique patterns.

"... movements of limbs without concepts are mere happenings, not expressions of agency"[6] according to McDowell. This is a tacit and common overstatement of the rejection of behaviourism. It is more a definition than a declarative statement. Movements of limbs are taken to be expressions of agency regardless of whether they are "really" without concepts millions of times every day by children watching actors at the cinema, children playing with toys and pretending to react to the "actions" of other toys, people fooled by shadows which, presumably, move "without concepts". If these cases are "mistakes" and a misapplication of the notion of "agency", then it is a mistake that is generally inscrutable in the manifest actions. The idea that actions without concepts are accidents comes from Kant and was meant to indicate the all-pervasive nature of the categories of sensible intuition. Since all of our actions are permeated necessarily with the categories of the mind, we do not, by exhaustive enumeration of cases, know any other way of talking about action that does not involve concepts. However, this is not to say that such an argument can establish that the concepts constitute a *sui generis* level of explanation. This would be a reification that went beyond the implied weaker claim that the concept of "action" involved the concept of "concept". I may not be able to explain the basics of thermodynamics to beginning physics students without the concept of indivisible

[6] p. 89

elastic atoms. This does not, of course, mean that such things exist or even that they are thereafter *necessary* for anything other than this pedagogical task. McDowell's position is more relevant to what to *call* something when we err, it does not establish a *sui generis* realm over and above "mere happenings" since it fails to establish a rational as opposed to a causal link between the ROR and the ROL and thus that there really are two different realms at all. Given the epistemological opacity of the causes of perceptions, we might *stipulate* the difference between components of actions or thoughts; those involved in the ROR and those involved in the ROL. However, these can never become apparent to us epistemically since, if they did, the connections that prevent rampant Platonism or bald naturalism would disappear. We avoid these two only at the price of having no possible empirical basis for a theory holding a division between the two realms. The only basis for a theory wishing to hold a stratified view is to hope that patterns resulting from the interaction between the putative realms are indicative of their essentially different natures. I turn now to an account of how an attention to modern brain theory, placed firmly in McDowell's Realm of Law, can allow an explanation of manifest patterns that does not necessitate a higher *sui generis* level of explanation.

6.4 Patterns from Invariants

It is a simple idea that the patterns that we perceive in something need not be causally efficacious. Patterns and regularity can fit data but do not thereby imply anything about the causes of the data. Quine famously champions this piece of common sense with respect to Chomskian linguistics[7] but it is a common problem within the areas under discussion. We are tempted to ground the patterns in our data we take to be important in something we understand and if we cannot understand a certain way

[7]See especially (Quine 1972)

of building these patterns, we invent a level of explanation that enshrines the patterns *sui generis*. The geometrical picture of brain operation given in chapter 5 relies on a general mathematical model, tensor analysis, that requires that the invariants that define the systems are *real* invariants. That is, an invariant across the abstract spaces employed to process information in the brain must have a corollary in the real world; there must be some invariant in the environment. This is expressed more intuitively by saying that the central informational features that characterise our brain's activity must have an anchor as the mathematics presupposes this. Thus, such a model is committed to a minimal ontology of "invariants". It is committed to no more than this as the only thing common, by definition, to the spaces employed by the brain in its operation are unchanging relations. It is only these that are required to be real for the mathematics to make any sense. This is ontology by reduction. We strip away those concepts we find evidence to suppose are imposed by the geometries imposed by the brain and we embrace what is left. This is an empirical enterprise. If we trust the work of Pellionisz and Llinas, then we already have an account of how space and time are imposed by the geometrical structures and processes that the brain employs in order to implement sensory-motor coordination. Given the considerations involved in labelling the axes of a geometrical brain-space (see Chapter 7) we should be extremely sceptical about being able to give convenient words to describe the basic feature contributing to our language and thoughts.

Now, this is strongly Kantian as abstract invariant relations are not the sort of thing easily classifiable. The sorts of invariants in complex geometries are, as explained in chapter 5, highly unintuitive and certainly without names in the usual medium-sized object-ontology of common sense. The invariants of the transformations between different geometrical spaces give rise to geometries. These geometries are, I argue, exactly that which shape our experiences and perceptions; they are Kant's conditions of sen-

sible experience. Now, the patterns we see in our experience are, on this model, the result of the underlying invariants of the system. It may be a constant ratio of one informational ingredient to another, it may be a very complex constraint on the possible information from a particular source, as restricted by other sources. The famous lateral inhibition effect of the eyes is a simple example of this and there are many more that are common currency in basic psychology and biology textbooks; our speech is not partitioned phonetically into the neat word divisions we phenomenally perceive, our skin does not register temperature in a linear manner commensurate with the usual scales etc. To give an overview of the lateral inhibition case, our eyes exaggerate the light/dark boundaries in our visual field considerably; the world is "actually" more blurry than we see it. Thus a ratio of light/dark is kept constant (invariant) on boundaries by a relation between the activities of the cells bordering the boundary reflection in the eye. So, this invariant defines a geometry with, in this case, a non-standard metric which increases the perceived "distance" between the light intensities of areas on different sides of the light/dark boundary. Note that this invariant could in no way be determined merely from the phenomenology of the case. Holding things invariant is an important thing for animals to be able to do as it ensures a regularity in the environment one can exploit. It matters little if the invariant is "really" there as this cannot be determined by the animal. The survival value of being able to rely on a *constant*, however caused, is enormous; maybe an animal's eyes filter everything so that it appears uniformly dull at all light intensities. This is not what happens "in the real world" but affords a huge advantage when hunting or escaping in tree-covered environments where one might burst from cover and into sunlight many times. Invariants *define* phenomenology and thus are transparent to it. The discovery of lateral inhibition and all of these other examples of distorting invariants were empirically afforded by studying, for example, the cells of the eye or the transduction of the skin. They could

hardly have been discovered any other way. This is a classic case of the necessity of empirical study of the "low-level" of a system in determining effects that are the result of the physical make-up of the system. This is a form of "embodied analyticity" where the composition of the system ubiquitously determines its outputs. It is, so to speak, a matter of definition of the physical structure that perceptions are as they are. If one is to revert to the Kantian notion of analyticity, the subject of a particular instance of an animal is contained in the predicate of "being so and so type of creature".

The first moral of this picture is that patterns may be determined in extremely complex ways by nonintuitive elements such as abstract relational invariants. This means that a simple *sui generis* treatment of language, semantics and thought is certainly not exhaustive of the possibilities for explanation in this area even though the radically unintuitive nature of the explanational elements in the sort of theory here advocated may be construed, irrelevantly, as counting against it. As detailed in chapter 5, certain types of information, so called "contravariantly" composed information, cannot be broken into its constituents due to their interdependence. The particular interdependence is a result of the geometry of the particular space which is, in turn, a result of the invariants underlying it. For example, the original position of a particular coloured ball in a bag of fifty similar balls after vigorous shaking is not determinable from the final configuration, even if we know all of the laws that govern elastic collisions. The relation between final and original position is so dependent on those of all other balls, we do not have the information to determine this. So it is with language and thought; the relation between the patterns of our language and thought and their realisation in the brain are necessarily inscrutable from the patterns themselves since the elements of the pattern are so interdependent. It is my view that interdependence is captured in the approach here advocated by the notion of contravariant information which means that the sources of the patterns manifest in our language

and thought are not determinable from the surface data. Any account of the origins of regularities in language and thought must, therefore, be restricted to an empirical study of the low-level geometries and invariants that the human brain is sensitive to, just as in the case of the discovery of lateral inhibition.

6.5 Rethinking Second Nature

What is left of the desirable features of McDowell's Aristotelean notion of "second nature" in a picture such as this? It gave us a way of expressing Kant's observation that the origins of the objects of our perceptions were not dividable into "world" and "non-world". The point is that nature, while possibly divided ontologically, is not divided epistemically. This is an important point. A general conception of the brain as implementing complex geometries requires real invariants in the environment. The crucial thing about the notion of an invariant is that is it essentially a *relational* concept. Invariants are *relations* between things like angles, length ratios, areas, volumes etc. Now, an invariant relation can hold between, so to speak, anything. It is a constant relation between information. Information, in turn, is absolutely neutral with respect to our everyday ontologies. Information, geometrically expressed, is the same whether it is, to us, "of" a table, a chair, a convention, a fashion, a contemporary intellectual issue and so on. The patterns we see in our mentation are numerous but do not necessarily have anything to do with the organisation of the brain processes that give rise to them. Information and the invariants of its multifarious relations in a given system are neutral to our classifications. This is simply a more precise way of expressing the common empiricist doctrine of "raw data"; it is simply all we can be committed to when we divorce from our own conditions of experience. So, the notion of second nature is a useful mnemonic to remind us that invariants in the environment are as much a

part of "nature" as anything else we might take to be. Second nature is only "second" in the classifications that result from this indeterminacy of the origins of our experiences.

This moves us onto the central point of these observations. A common tactic in the face of Kantian considerations or the inadequacy of formal structures is to hold on to a stratified theory and to allow that the relation between the parts of the theory is very complex. The parts are *interdependent*, as in McDowell's supposed relation between ROR and ROL. It is the same in traditional semantics. The relation between "meaning" and context or "pragmatics" is glossed as complex in order to provide some reason why the theory faces problems in accounting for certain data. This view is completely backwards. The interdependence, to have any plausibility in the light of Kant and modern brain theory, must be so close, that there is no non-empirical[8] way of determining *what* the elements said to be interdependent *are*. An interdependence of information complex enough to support semantic life *prevents* a casual or *a priori* specification of the underlying parts and sources of the information. It is not something than can *augment* an ailing stratified system in order to give a name to its failures. In the case of McDowell, we cannot say that the ROR and the ROL are separate but closely linked as a link close enough to account for a reasonable interpretation of second nature blocks any philosophical attempts to conclude their very existence and separateness. If the brain is information neutral with regard to its inputs, as it must be at a low-level – discrimination of source of impulse is not something that neurons etc. do – and its manifestations are not thereby sortable into those originating in the ROR but naturalised to appear to come from the ROL, and those directly emanating from the ROL, what hope is there for a theory that *starts* from such

[8]This point must be carefully understood in the light of the comments about lateral inhibition above. Empirical work may tell us about the mechanisms contributing to our manifestations and this can provide the only route to a meaningful analysis of them. This is discussed further below.

a distinction? The crux of the problem is that there is no way to give an *argument* for a stratified theory based on the phenomenology or manifest structure of perception and experience. Empirical investigation into the invariances defined by the processing are more promising but the power of the geometrical and mathematical systems commonly employed is such that stratified theories such as McDowell's never arise. Since the structure of the manifestations of a systems are not taken as *sui generis* features to be explained, there is no need to pay lip service to such features and thus no need to maintain an irrelevant concern for intuitive models, theories which embody the features to be explained and *a priori* constructions of features.

6.6 Analytic Truth and Noumenal Meanings

Quine's landmark "Two Dogmas of Empiricism" claims to demonstrate that stratified theories in semantics require an analytic/synthetic distinction in order to demarcate between contributions to meaning from the system on the one hand and the world on the other. He claims that this distinction is essentially unclear and thus stratified theories of the type he is concerned with[9] are without basis. This is akin to that which I claim: the contribution of the world and the mind are essentially indivisible due to the way that information is processed in the brain. This division is utterly inscrutable in the manifestations of the brain and thus language regularities are of no use whatever in properly explaining human semantics. In natural science, we have a matter to be right or wrong about, according to Quine. We have such a rich set of controls and background assumptions that we can hold certain things constant in order to examine the independent effects of certain phenomena etc. In semantics, we have nothing to be wrong or right

[9]Quine talks of intensional linguistic semantics which is a paradigm of the stratified system in Cognitive Science with its level of "propositional" representation. This is, for example, a direct philosophical offshoot of the tenor of McDowell's ROR.

about. We are looking for "meanings" but need them, in turn, in order for the entire enquiry to have any constraints. The point is that semantics has no stable information, the source of which we can be sure, while sciences have such a rich background, much can be held as stable in order to lever new experience into place. The useful "facts" in enquiry are those which allow us to determine whether our results come from our methods or the world. According to Quine, despite underdetermination, we have enough coherency to be able to make this distinction do some work. In semantics, however, we do not have the standards, because of the famous indeterminacy of translation which guarantees an infinity of possible ways of slicing up the world/word boundary, and thus have no "facts" to help the enquiry. I have expressed this by arguing for the impossibility of decomposing information embodied in a contravariant manner. Indeed, the ways of breaking up such information into components is vastly underdetermined unless one possesses the key to the invariants of a system. As mention in an earlier chapter, these are not obvious at the level of their manifestation and thus I claim that the geometrical picture presented supports fully Quine's claims in respect of translation.

It is well here to examine Katz's argument that linguistics, of all of the approaches that claim to be able to distinguish analytic and synthetic, succeeds where others fail. If Katz is correct, there is indeed something to be gleaned from manifest patterns and therefore a stratified theory may be based on such. Unfortunately, Katz's argument is completely circular. He argues that modern decompositional semantics provides the only sensible manner in which to define analyticity as it gives a systematic meaning to the Kantian formulation of this in terms of the subject being contained in the predicate.

> "Relative to an assignment of decompositional representations to sentences, we can define analyticity in terms of the semantic representations of the full predicate and each of its terms but one being formally contained in the semantic representation

of the remaining term."[10]

The notable first phrase gives the entire game away. Need it be pointed out that a stratified theory cannot be justified on the basis of a successful demarcation of analyticity, and thereby a possible account of the import of manifest features of language, if it rests upon "an assignment of decompositional representations"? This, after all, *is* just a stratified theory; the very one which assumes a level of propositional representation akin to McDowell's ROR. Katz is rather casual in remarking that "The first thing to note about such a definition of analyticity is that it makes no reference to thought processes ...". Not explicitly, granted, but the kind of representational system he envisages is exactly that, supposedly captured in a formal system designed to picture language-neutral content as a spoke in the wheel of a general grammar. Katz is keen on the generally *systematic* nature of the formalism he envisages. It is this which is supposed to capture the division between the contributions from the world and that from the language itself. However, the systematicity of the formalism he advocates is not something that is derived independently from that which it is applied to, as in the case of the physicists application of mathematics. It is desired because of and indeed derived from the imperfections of natural languages for scientific pursuits. The systematism of the formal is guaranteed to be able to make distinctions that fit language as it is distilled from that very language. There is no surprise that we can draw an analytic/synthetic distinction in a formalism designed to do violence to natural language precisely because the latter cannot make such a distinction.

It may be objected that the implications of a general geometrical theory of brain processing imply nothing stronger than the conclusion that the stratification inherent in the relation between mind and world, thought and language is inscrutable, but real nonetheless. We may not be satis-

[10](Katz 1990)p. 190

fied that the current epistemic is the boundary of our possible knowledge in this matter. Maybe the issue regarding the stratification language is merely *underdetermined* rather than intractable. This is Katz's view and means that, as he puts it, semantics and the philosophical positions underpinning it are in no worse position than underdetermined theories in the natural sciences. He asks why we might not allow ourselves noumenal meanings that we cannot get to but that are still real in some sense. The difference is simply that underdetermination requires something to be undetermined about. This something must be a central high-inertia element of our Quinean web-like ontology. We have physical reality in the case of natural science but we have nothing in the case of semantics. As detailed in chapter 4, formal semantics simply is not richly embedded enough in our conceptual framework in order for it to have points of leverage that would enable it to claim a subject matter fundamental to our understanding of human cognition. We have no evidence for the existence of "meanings" other than the dictates of the formal systems that sometimes like to claim that they have scientifically suggested the existence of. In turn, these systems require the existence of meanings in order to justify themselves. In semantic formalism, as for McDowell, there is only one possible outcome; a huge *petitio principii*. The stratified view that we would like to conclude from the manifestations of our thought and language has nothing to motivate it other than patterns based ultimately on a singular explanation that obviates stratification. A reasonable motivation for a stratified view would require a division between informational contribution from the world and that from the brain but a distinction needs something solid to push against. The rejection of the analytic/synthetic distinction removes this possibility and formalists such as Katz are left begging the question by arguing that formalism saves the distinction which is then used to save formalism's central feature: the stratified view.

Chapter 7

Geometry and Semantics

Since the advent of connectionism, there has been growing interest in distributed models of information processing. That this interest has spread throughout the AI and Cognitive Science communities is in part due to a desire to compete with the dominant symbolic paradigm but is also borne of a recognition of the power of such models. The surge of interest in this direction is, perhaps more modestly, a resurgence: as long ago as the 1960s, there were attempts to develop geometrical models of certain semantic phenomena[1] but these were more concerned with the task of increasing the usefulness of artificial languages for use in the philosophy of science.

In recent years, there has been more and more interest in exploiting broadly geometrical systems in order to regiment order into the modelling of human language semantics[2]. The main theme of this chapter is that most recent attempts of semanticists to embrace aspects of geometrical representation fall foul of some of the most undesirable features of the symbolic paradigm they seek to compete with. These failings are, as I shall show, examples of stratificational assumptions fundamentally the same as those in traditional formal natural language semantics. It is interesting to

[1](van Fraassen 1966, van Fraassen 1967)

[2]Recent accounts have included (Gärdenfors 1990, Gärdenfors 1993*b*, Gärdenfors 1993*a*, Churchland 1986*a*, Churchland 1986*b*, Crangle & Suppes 1989, Jackendoff 1988)

note that these failings are less apparent in related research in psychology. For instance, Roger Shepard has been pursuing research into concept formation and use for many years using quite sophisticated geometrical models. See e.g. (Shepard 1980). However, the concerns of the psychologist are different to the concerns of the semanticist: the latter more with the lessons to be learned from the native speaker rather than the acquisition by the inchoate. If a "geometrical semantics" is to constitute a worthwhile direction of study in contrast to the traditional symbolic program, it is important to identify and divest of the elements that are objectionable in such models.

The task of developing a "geometrical" account of semantics is a difficult one but one that needs to be tackled if we are to fully appreciate the basic assumptions involved in explaining human language and cognition. As an example of the advantages of this approach, let us outline some early steps by Peter Gärdenfors (Gärdenfors 1990) who has shown how a geometrical notion of "concept" can interestingly account for the tendency of humans to choose, to use Nelson Goodman's terminology, projectable over non-projectable predicates. This is a problem for a symbolic semantics as the extensional definition of predicates that lies at the heart of such accounts is neutral with respect to the actual contents of a given extension. If predicates are defined by extensions then the criteria for choosing between possible predicates to describe something is the task of choosing between extensions. The projectability problem is exactly the problem of deciding between extensions. The problem can be formulated in a simplified version simply in the following manner.

Suppose we have a green emerald in front of us. What predicates apply to it? "Green" is obvious but extensionally adequate is also "grue" which is true of something that is green now and blue tomorrow. Which predicate should apply? Green or grue? Since predicates are sets of objects in the classical tradition, does this emerald belong to the green or the grue set?

Green is said to be a projectable predicate as it is projectable over all times. Humans, as a matter of fact, rarely choose predicates like "grue" that involve a time-coordinate. Why? Traditional semantics has no answer to this as all extensions are theoretically equivalent. By construing concepts as *convex* regions of conceptual space, Gärdenfors avoids the projectability problem by making projectability, in effect, a *structural* feature of a conceptual space rather than a conclusion of a symbolic inference. If one conceives of a conceptual space describing the colours and having a time-axis, it can be seen that "grue" occupies two unconnected areas of the space whilst "green" does not.

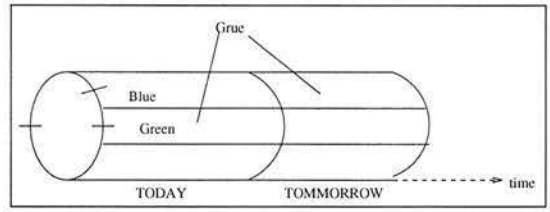


Figure 7.1: *Projectable predicates are convex regions of a conceptual space.*

An initial important point to note here is that we do not need to *infer* that grue is not a sound predicate: it follows from the structure of the geometrical spaces that form the basis of the conceptual processing. I believe that this sort of approach comes very close to addressing the problems so forcefully expressed by Hubert Dreyfus in his classic (Dreyfus 1992). Dreyfus was one of the first to bring to the attention of AI workers, the fact that the storage and processing needed to model human cognition in a symbolic paradigm is utterly implausible as an explanation of how humans manage the task. Dreyfus draws from Heidegger, Husserl and Merleau-Ponty to argue for a primacy of context that is impossible to store as explicit facts and is resistant to traditional approaches to inference because

- The facts required for human inference are seldom explicit.

- The facts are never statically meaningful. They never have an invariant meaning between situations.

- It is therefore never possible to enumerate all of the facts necessary to perform inference in any formal symbolic system.

The sort of model proposed by Gärdenfors demonstrates that the "facts" can be seen as the result of the structural arrangement of aspects of the brain. The main problem with the traditional treatment of inference and predicates is one of plausibility. The number of facts one would need to store in order to be able to make reasonable inferences about e.g. the objects in a small room is more than even the most liberal sub-neuronal model of information storage would sanction. Work on trying to implement "contexts" and "presuppositions" have sought to overcome this with notions of "default reasoning", "contextual hierarchies" and the like. These strategies traditionally appear as lists of static information used to restrict inference in one way or another. If Dreyfus is right, these lists are nothing to do with real contextual information and are necessarily incomplete. The grue problem highlights this well as it shows that the resolution of a problem to do with choice between extensional equivalents cannot be solved by appealing to fundamentally extensional theories. In addition, one encounters an infinite regress of contexts as one has to appeal to higher and higher level contexts to decide what information is relevant at a given time. This is a fundamental point and I shall expand upon it further.

A choice between alternative statements of facts might depend upon other facts. Possible alternatives between these further facts might, in turn depend on other facts. In the end, however, a decision requires grounding if it is ever to manifest. If we learn anything from the grue paradox, we learn that facts require something of a different character to determine them. Essentially, we need some *structure* in the framework in which facts become facts for the indeterminacy to halt. Something must force us to

stop at accepting one fact out of a possible variety, of accepting green as opposed to the infinity of non-projectable predicates. For Dreyfus, it is the "context". One might say: a context is something that gives facts specific meanings and is not something created out of facts. Context precedes facts; context creates facts. To model contextual information as lists of facts or presuppositions is precisely preceding the horse with the cart. The problem here is that the symbolic paradigm sees "context" in a very different light to the sense in which Dreyfus means it. Context for the symbolist is more symbols whereas, for Dreyfus, the whole thing that makes context tractable and meaningful is that it is not. This is another, more philosophical expression of the point raised in chapter 5, that we must act in real time. We must choose, the choice process must end and it is not clear how this can happen if a choice depends upon something about which there is as much choice as the first thing.

The difficulty with this and one of the main reasons why Dreyfus is so vehemently attacked is that we simply have little idea how to approach the issues after we have accepted that the symbolic paradigm's notion of context is not what he was looking for. The whole goal of computational implementation and modelling in the traditional paradigm is made possible by the explicit delineation of things. One cannot use the context to compute if one does not have a computer-readable form of the context. There is an alternative to having to explicitly store facts if we countenance the idea of a structured conceptual space. In a geometrical model of information processing, we have at our disposal a very powerful set of ways of expressing information dependencies and relations. Information dependencies can be a result of the structure of the space and thus part of the very basis of our conceptions; this is much closer to the sense of "context" we need as it is not expressible in terms of the facts it will define. In a sense, this means that the starting point of semantics is not "basic" units – semantic values of syntactically simple parts, but rather the structures in which the puta-

tively simple parts are embedded. This casts doubt on the notion of compositionality since, on this picture, semantics of a "part" is possessed only derivatively from surrounding contexts and compositionality is exactly the reverse of this.

This has important ramifications for understanding the notion of analyticity. Given a structure imposed by a certain space, relations between elements in such a space are such by *definition* of the space. Indeed Gärdenfors says

> "... it follows from the topological structure of different quality dimensions that certain statements will become *analytically* true (in the sense that they are independent of empirical considerations)."[3]

This is encouraging as the computational explosion of performing traditional inference on machines much faster than human processing speeds is well-known evidence that people are not likely performing classical inferences when they cogitate. If the conclusions they reach are simply a result of informational dependencies enshrined in the intricacies of the spaces implemented by the brain to process information, nothing like inference needs to take place. One can see why Dreyfus talks of humans just "seeing" answers and "focusing" on things. Thus, I think that the key to understanding human information processing is more a case of understanding the way in which the information is stored and used rather than concentrating on static interpretations of "meanings" and how they can be combined. The hypothesis is that the central things about our semantic usage must have a basis in the *structure* of the system employed to process the information otherwise supposing that we are explaining how humans perform such tasks is at odds with at least the computational requirements of our model. I say "at least" as I think that without this, it is at odds with the *philosophical* requirements too.

---

[3](Gärdenfors 1993*a*) p.9

7.1 Some Geometrical Theories

There has been little work in semantics examining what may be deemed to be "geometrical" theories. A geometrical theory is one in which the system used to model semantic relations, combinations and values is represented as some kind of space of possible valued n-tuples. Each point and possibly area or volume would correspond to some semantic feature. For example, we may have three dimensions labelled "Number of Legs", "Fur density" and "Height". Maybe "dog" might be a region for which the values of these dimensions falls within certain limits. Taking a cue from physics, one might suppose a much more general space with more abstract labels for the axes involving what you take to be the primary qualities shared by all things and hope to delineate everything in that space. Perhaps you might choose "atomic mass", "space-time coordinates" and "velocity" as your dimensions. This is unlikely to distinguish, semantically speaking, a book from a chair however. It is clear then, that a large part of the evaluation of a geometrical approach will depend on what it chooses to label its axes. It is, in effect, the same problem as what decomposition to give to a semantically complex element in the traditional approach and this is an area we shall return to examine shortly.

Paul Churchland espouses a view in which multi-dimensional spaces (hyperspaces) might contain hypersurfaces on which, points are all of the grammatical sentences of a language. Further,

> "... the logical relations between them [would be] reflected as spatial relations of some kind."[4]

Here we have a geometrical model that seeks to be able to account for the usual concerns of semantics in a new way. It is a way that Churchland hopes might provide a reduction of the familiar Chomskian picture. The relations between semantic units would be seen as geometrical relations.

[4](Churchland 1986*a*)

We are considerably more informed regarding the brain's implementation of these than its supposed implementation of any of the formalists traditional theoretical entities and processes. Of course, this would only be a reduction of the familiar picture if the basic elements of this picture were not part of the system being reduced. So, Churchland is careful to choose "objective" and blatantly brain-level qualities for the dimensions of the spaces that he proposes. For example, his canonical example is Land's colour cube wherein the three dimensions are Hz scales corresponding to the amount of red, green or blue in a particular colour. Similarity of colour in our perceptions is explained as metrical distance in the colour space. The first problem that results from choosing such labels for the axes of a space is that the common problems regarding qualia arise. Basically, to take such empirically fundamental attributes as frequencies as the labels of the axes of a semantic space, is to suggest a fairly bland sort of reductionism that seems to reduce things to features so inessential, they fail to demarcate between experiences. This is just the usual qualia inversion problem. If my experience of green is your experience of red, then knowing that green is a certain triple of frequencies and red another different set helps not at all to distinguish between these two experiences. This is because the relation between the Hz and the phenomenal descriptions seems to be completely contingent. It is hard to see how the phenomenal result of the combination of an n-tuple of basic, physical dimensions could be linked to those dimensions in any way that would make sense of semantic similarity. In the same way, the notion of metrical distance is of no use in explaining our phenomenal perceptions of similarity as this relation is also contingent. The point is forcefully expressed by Fodor and Lepore in their critique of Churchland. They argue that the dimensions must be related, in some non-contingent way, to the *content* of the phenomenal perception otherwise the link between the *explanans* and the *explanandum* is too weak to avoid even basic concerns about qualia.

There is a point of value here from Fodor and Lepore that will be be discussed more below. It is to do with the types of label that will occur on the dimensions of the spaces. Their point about the contingency of the values of non-semantic labels is ill-founded however. If the relation between the phenomenal qualities and the dimensions of the underlying space is *not* contingent, then one explains nothing – one merely defines. Fodor and Lepore intend to have semantic relations accounted for in terms of other semantic terms. They would rather see the dimensions labelled with semantic labels that could contribute to the *content* of the concepts defined by the space. However, the "content", as they see it, is the usual propositional attitude notion of content and thus the relation between the dimensions and the points in the space must be necessary, given a suitable compositional treatment. It is clear that Fodor and Lepore are failing to make the conceptual shift necessary to accommodate Churchland's geometrical theory. This is fuelled, however, by Churchland's desire to accommodate some of the desiderata of the formalist approach; he would like to account for language in terms of "grammatical sets of utterances" and "logical relations". However, Churchland employs basic features simply not suited to the task of doing this. The tools of the formalist have built into them certain assumptions regarding language (compositionality, truth functionality etc.). In addition, they have certain goals in mind (adequate grammars etc.). Churchland is making the mistake of desiring the latter without wanting to use the former. In their critique, Fodor and Lepore are correct in chastising Churchland for attempting to build the formalist goal out of parts not suited to the task. However, their own conception of his mistakes goes too far in arguing that the problem is the *contingency* of the relation between the parts and the whole. Fodor and Lepore seem to require that we must have an answer to the question of "why" do we think, say, one word means something like another word or one colour looks more like green than red. Thus, the connection between the basic theoretical posits of a system and

their linguistic or cognitive manifestation must be more than contingent. It is not clear why we need to answer this question. It is simply just part of the way our brain is built that certain things are true of us and thus the "why" question needs either no answer or one that merely describes the structures that lead to the manifestations. A "why" question to a formalist is really only answered by showing a decomposition of "meanings" in a formal system and thus there is no surprise that Fodor and Lepore think that Churchland has not answered the question. He does not, in fact, think it is important. That is the whole thrust of his Eliminativism.

Nevertheless, Fodor and Lepore have a good point when they say that Churchland's geometrical theory results merely in "semantics by stipulation". If the dimensions of a space are low-level physical features and are thus non-semantic, then how do we determine what is represented by theoretical components of the theory? For example, why does a particular semantic construct become represented in the theory as a region rather than a point or a line? We cannot explain this by reference to the semantic features as these are traditionally determined by their *contents* and the dimensions give no clue as to this, as mentioned above. The dimensions might *cause* a certain mental state but only their *content* can classify them properly. Thus, we must merely stipulate what geometrical features correspond to what semantic features and this is, according to Fodor and Lepore, ludicrous and unempirical. This is a point well taken since there is no way of mapping geometry to semantics in this way that is not governed simply by the desire to make it all work. Since this is exactly the main part of the basis on which I criticised formal semantics in Chapter 3, I can hardly endorse Churchland's position here.

The way out of this is to realise that a move to a geometrical picture is one that requires rather more than a shift in terminology. It requires a fundamental change in the conception of methodology in Cognitive Science. Fodor's approach is typically "top-down" in that he takes for granted

what is to be explained and attempts to do so. He has started with a problem and seeks to solve it. The problem is to find out what the basis of the manifest patterns of language and thought are. This is done by taking the patterns as data and attempting to account for them. Churchland is in an unpleasant position as he alludes to similar top-down goals but uses rather bottom-up tools. These tools are designed to show how patterns are constructed; the patterns are the conclusion, not the premise[5]. A move towards removing this problem is to drop the requirements that the formalist program forces on us. Compositionality is one which drives Fodor's theorising and one which I think can be relaxed. This is discussed in Chapter 4. However, there is a more important point in relation to Fodor's biases that it is germane to discuss.

Mental states are only classified by their contents if we believe that being able to understand the classification is important. It is important if you would like to have the usual solution to the qualia inversion problem. You show how the states are actually differentiated by showing the steps by which the states are constructed; by analysing their "content". But the notion of "content" is hardly something obvious. "Content" is something that we abstract from the surface manifestations of language and thus the assumption is that such manifest patterns are a guide to the actual make-up of a mental state. But what evidence do we use in order to decide upon an analysis of the content of a state? We are guided by the formalism we are using and the way in which it is designed to split up the patterns. A propositional calculus chops into larger pieces than a predicate; modern logics make still finer distinctions (see my remarks on Katz's attitude towards this in Chapter 3). The formalism decides what is to count as a pattern – predicate/argument pairs are not a pattern in propositional logic

[5]See p. 419 of Patricia Churchland's *Neurophilosophy* where this consideration of "top-down" vs "bottom-up" methodology is briefly noted. She too does not see the problems, however, of combining the top-down aspirations with the bottom-up tools.

for example – and thus decides on the evidence. If the formalism decides on the evidence, then the formalism decides on the "content" of a mental state. Classification of mental states is therefore dependent on the formalism used to perform the classification. This is much worse than the underdetermination of the natural sciences as there is no observation, or empirical base of any sort. It is entirely faked by the choice of formalism: see Chapter 8. So, the top-down approach in semantics is a methodological bugbear that requires no loyalty on Churchland's part as it conflicts with the empirical aspirations of his theory. More importantly, the axes of a conceptual space evidently need not be semantic, as Fodor suggests, as this solves nothing at all. It merely reflects the structure of the formalism used to analyse conceptual content and, for lack of empirical relevance, reduces the specification of "meanings" and thus semantics to an exercise in definition rather than explanation. One explains by showing how one thing connects to another, not how one might impose a structure on something and then claiming how revealing of structure the formalism that imposed it in the first place, really is.

Churchland's picture was well-motivated since it attempted to address the unempirical nature of the formalists view by embodying a connection between the non-semantic and the semantic. One could hardly "explain" semantics otherwise. There is still, however, the problem of what to label one's dimensions in a geometrical picture. If the dimensions are semantic features, we merely ape Fodor's formalist view. If they are simple physical attributes, we are at a loss as to how to combine them to produce semantics, a point well taken from Fodor. Remember that Fodor and Lepore questioned the role and labelling of the dimensions in Churchland's geometrical model

> "We're claiming, in effect, that Churchland has confused himself by taking the *labels* on the semantic dimensions for granted."[6]

[6](Fodor & Lepore 1992) p. 199

Let us look at the work of Peter Gärdenfors again briefly in order to see this problem in relief once more. Gärdenfors has a theory of conceptual spaces that has a similar form to Churchland's. He has similar problems when attempting to define the dimensions of his spaces. For example, Gärdenfors argues that the dimensions of a given conceptual space are things like

> "color, spatial position, weight, temperature, etc."[7]

Again, we have the problem of knowing how these combine to give us anything recognisably semantic. As with Churchland, where does Gärdenfors obtain the labels for these axes? If they are broken down from the patterns in our semantic manifestations, we have the element of top-down methodology that infected Churchland and which makes Fodor's claims to explanation, as opposed to mere definition, so questionable. Also, why are geometrical theories vague about the exact dimensions? As is the case for traditional approaches to this problem, the "etc" in the quote above is a promise to fill in all the determinate qualities of a static model. Giving a complete specification of all of the axes of a space is just like trying to lexically decompose it into a unique and complete description. If we take Dreyfus' criticisms seriously, this is not possible and no-one, since work began on this in the 1950's, has succeeded in filling in even a tiny portion of this dangerous promissory abbreviation or plausibly suggesting how we might. This weakness in current geometrical approaches does not go unnoticed by its opponents, as we have seen. Indeed, concepts of colour, weight etc. are not constitutive of a conceptual space. They are "facts" we perceive as a result of the information dependencies enshrined in the structure of the space. If such predicates were the labels of the dimensions of a space, it would be hard to see just how then a geometrical approach would differ from a traditional symbolic one. Both would be using atomic and basic facts or concepts as the building blocks to higher-level structures. This is a

[7](Gärdenfors 1993*a*) pp. 8–9

mistake as the putative basic elements, the labels of the dimensions of our spaces, only exist by virtue of being embedded in a context; a context that results from the dependencies necessitated by a highly structured space. The labels are not determinable from the patterns that the space gives rise to since this would be exactly the same mistake as supposing that the formal building blocks of semantics are discoverable from the manifestations of language. We must assume, in order to believe that the top-down analysis of language patterns is of use in explaining anything in semantics, that the putative parts of the patterns are independent. We must be able to determine them separately otherwise the dependencies involved in the composition of manifest language forms would make the task completely indeterminate. In fact, this is exactly the case, as argued in Chapter 5. The processing involved in information manifesting as what we call "semantic" must involve complexity at least as great as phenomena such as limb coordination and thus must involve geometrical models where the dependencies between information in output are so strong as to make it impossible, without knowing the "key" (metric tensor) that generates it, to work down to the basic building blocks in a top-down fashion. The only way to even make it *appear* as if one is doing this is to employ a formalism that imposes a structure on the patterns of the language data. The choice of formalism determines everything in a system inherently immune to analysis at the level of its manifest patterns. What sort of dependencies are we considering? Typically, the paradigm of the traditional approach to semantics in Cognitive Science and AI is the logical atomism of Wittgenstein's Tractatus. Here, all atomic facts are independent, hence the need for a huge mass of inference rules to link them. The facts are *orthogonal* to each other. This would correspond to a simple geometrical model where, in Euclidean space, the axes were all at 90 degrees to each other. In this model, the metric tensor is the identity matrix. This is the picture corresponding to the basic formal models. Each determinable feature of a

semantic unit is separate and can be ascertained independently of the others. In geometrical terms, change of value along one dimension involves no change at all along the other. Many of the geometrical approaches involved in neural network research never move beyond this simple model; Churchland and Gärdenfors, due to their mistakes in embracing the usual top-down approach, also share this sort of geometrical picture. The reason is, I think, because any form of informational dependency would prevent one from determining the axes for a particular natural language example and thus from giving expository clarification to the theory. This is a temptation that should be resisted. If a geometrical system really implies that manifest patterns are no use as evidence, then one should be honest and simply not give examples based on such evidence. It is misleading as it suggests that the system is suggesting basic units of analysis – the labels of the dimensions – but it is not. These are being determined intuitively by appealing to some vague notion of "basic" or "abstract" features, taken as obviously fundamental in some sense. In a way, this is worse than the formalists' tactic of taking the basic units from the way that the formalism breaks the manifest patterns up since this, at least, has some independent motivation from formalism, albeit misguided.

In a geometrical picture where the axes are not orthogonal to each other, as neurological research strongly suggests for more primitive tasks such as motor coordination, fundamental changes in our conception of information processing are necessitated. Here, a change in value along one dimension necessitates a change in value along another: points in this type of space have a context of informational dependency underlying them due to the space they inhabit. In addition, the notion of "distance" between points and hence the notions of area and volume of regions are dependent on where in the space they lie. These sort of complications make the task of deciding on the geometrical representations of manifest features completely indeterminate and thus Fodor and Lepore's criticism of the "seman-

tics by stipulation" of Churchland is well taken. There are countless ways of deciding on which geometrical features correspond to which semantic features; this is rife underdeterminism. Gärdenfors seems to think that there is an obvious way of dividing up the concepts in order to map semantics onto geometry.

> "... *individual names* are assigned vectors (i.e., points in the conceptual space) ... *predicates* of the language that denote primary properties are assigned regions in the conceptual space."[8]

Certainly this is an intuitive way to proceed but methodologically, it fares no better than the imposition of structure by formalism choice in the traditional approach. If we acknowledge a more sophisticated geometrical view, we have available *invariants*; abstract features embodying constant geometrical relations. The richness of the systems that these define is the subject of tensor analysis. To assume that we can make a geometrical model of semantics based on a simple, orthogonal Euclidean geometry is analogous to supposing that we might fully capture natural language in a propositional calculus. Unfortunately, once one embraces a more abstract geometrical theory, certain implications render the top-down methodology irrelevant. This is explicable in the following manner. Our cognition displays, certainly, *regularities.* It is natural, upon seeing regularities to think of rules that underly them. Regularities are described in a certain vocabulary; that of "nouns", "verbs" etc. is a familiar one. It is, however, a mistake to suppose that the regularities of a certain vocabulary can be explained only by reference to that vocabulary. A regularity described in the vocabulary of, say contemporary linguistics, need not be *caused* by anything so describable. The reason why it is commonly attempted in this manner is because an explanation in another vocabulary is much less easy to understand *as* an explanation. The stages of the explanation do not use the

[8](Gärdenfors 1993*a*) pp. 8–9. Italics his.

vocabulary of the explained and thus it is difficult to follow it as an explanation. We tend to call, instead, such things *reductions* and this has come to have rather unpleasant connotations since it seems to be *eliminating* something. However, if we realise that there really was nothing to eliminate in the first place; we only had vague regularities in our common vocabulary after all, it is obvious that a reduction, in this sense, is indeed an explanation. In fact, it is more of an explanation than one cast in terms of the vocabulary that the regularities are phrased in since otherwise it is difficult to see how such an explanation connects to other parts of science and thus it is hard to understand just what such an explanation has to do with human cognition at all. This is a shadow of the problem that McDowell has, which is discussed in Chapter 6; that of distancing ourselves from pointlessly feared reductions and its resultant rendering of the whole enterprise as irrelevant to human cognition. Overall, the problem is well captured by the epithet that formalism and contemporary cognitive science as a whole mistakes regularities for hints of systems of rules. It is common in the natural sciences that regularities on the scale our perceptions can detect are the result of features of micro-level systems. Patterns in snowflakes are explained by crystallography, not by reference to snowflakes. Linguists and formal semanticists still, however cling to a methodology that explains semantic patterns by reference to semantic features. Famously, we have failed to find hard rules for any semantic phenomenon of the sort we have, for example, in chemical bonding. This is because the patterns we find in semantics and language are explicable only in terms of informational dependencies involving a completely different vocabulary. It is the hypothesis of this work that that vocabulary must be a geometrical or topological one if we are to be realistic about the evolutionary and neuroscientific evidence.

7.2 The Rethinking of Examples

As discussed in Chapter 5, the necessity of more complex geometrical structures involves a certain amount of new distinctions in the theories that result. There are two ways, the distinction between them obscured in simple systems, of determining information in a geometrical space: contravariant and covariant. Contravariant vectors are, on this model, constructed for the outputs of a system. Our execution of acts is contravariantly conceived. The feature of this notion most important here is that because the components of a contravariant vector can be combined easily in any sort of space, they cannot be broken into their constituents without knowledge of the way in which they were combined. This knowledge is only possible through knowing the tensor employed in the creation. This, in effect, amounts to knowing the topological structure of the spaces involved. So, an account of the origins and components of the outputs of a complex system, such as that involved in inferring, uttering a sentence or thinking, requires knowledge of the fundamental structure of the brain-implemented geometry physically embodying the system. It is clear that this sort of knowledge is quite out of the realm of logic systems, linguistic grammars and most of Cognitive Science. It is a strictly empirical neuroscientific enterprise involved with identifying and examining networks in the brain and the possible geometrical systems they can plausibly be seen to embody.

There is a serious implication in this for Cognitive Science, quite apart from the implication that it would render much of it obsolete and mean "massive unemployment in Cognitive Science"[9] – something that is almost presented as an *argument* against geometrical systems by Fodor and Lepore. The common tactic in linguistics and formal semantics is to illustrate the supposed explanations by actual decompositions, constructions or models etc. If the information we have, the manifestations of language are

[9](Fodor & Lepore 1992), p. 187

contravariantly constructed, then we simply can give no such examples. Without the empirical research required to demonstrate the structurally imposed conditions of the operating of a system, our examples are constructed "blind". We have no reason to suppose them to have anything to do with how the regularities are manifested by the brain. Additionally, the only possible "basic" elements that determine the manifest regularities on our geometrical picture are the invariants described by the geometry of the systems involved. These invariants will be, as I have mentioned, extremely abstract relations between geometrical features and thus will not correspond to concepts we recognise as "semantics". We are used to this in the natural sciences where we have many concepts that are mathematically defined but which have little correlation with our phenomenal categories. "Electron spin", "torque", "rest mass" and the like are powerful explanatory concepts whose bases are solely mathematical and sometimes blatantly geometrical invariants. Incredibly, formal semantics and linguistics seem to expect that human cognition, surely one of the most complex of all systems we know of, to be explicable in terms of concepts we are very familiar with. Indeed, in the very early formalist empiricist traditions, the basis was supposed to be *the* most simple and obvious observables. It would be simply incredible if this were the case; a *sui generis* level of explanation in terms of semantic features and propositional attitudes would be completely at odds with the rest of science that, for example, the Chomskian tradition desperately wants to unite with.

So, what of our earlier question; what of the axes of the geometrical systems the brain embodies in its role as manifesting agent of the realm of the "semantic"? Clearly, we simply cannot tell from looking at the manifest patterns of language and thought. These are *output* and output in the systems under examination is not a good guide to the causes of the regularities in output. There is indeed no guarantee that the output is even slightly revealing of its causes. This means that proponents of "geometrical seman-

tics" should avoid at all costs attempting to specify *a priori* the dimensions along which we can determine the possible points in a space. If this is attempted, we must undertake the top-down labours of the traditional theories and fall foul of the same problem in different guises. We have, at least, to thank Fodor and Lepore for pointing this out. The methodological urge to "give examples" so prevalent in natural language semantics today is to be resisted as I regard the implication of the above that this simply cannot be done and, when it is, it is extremely misleading. Language cannot be built out of language since this would not explain how it is caused. There has been, however, a certain amount of research conducted into how the interface between formal systems and the accepted brain-like models can be described. This is known as the "symbol-grounding" problem and is an attempt to block exactly my criticism here. It is an attempt to "ground" the formal systems in naturalised theory, preventing the *sui generis* level of formal description being too distanced from explanations of real human cognition. In this sense it is exactly that which, in Chapter 6 I describe as concerning McDowell. It is an attempt to connect with the real issues of the physical implementation of semantics without loosing the formal level of description. In the terminology of Chapter 6, it is exactly an attempt to connect the Realm of Reason and the Realm of Law. What is particularly interesting about the Symbol Grounding Problem, is that it is a much less abstract attempt than McDowell's, well set in the traditional of Cognitive Science. As such, it is essential to examine it as an attempt to bypass a thoroughly geometrical picture and thus the implications it has.

7.3 The Symbol Grounding Problem

Since a formal system manipulates, by design, symbols, there is a naturally a wonder how and why the symbols connect to that which they are taken to be symbols of. If this question cannot be answered, then we are

left with the conclusion that the symbols in a formal analysis have no relation to that which they are putatively about. Basically, a system of formal symbols is nothing, in Cognitive Science, if it has no connection to what is presumed to be an important factor in the embodiment of such systems, the brain. This is just the same problem as McDowell's worry of the Realm of Reasons "spinning in a void" with no connection to the physical world. Traditionally, formalist Cognitive Science tends to regard this particular problem with disdain. There is an assumption that we need not concern ourselves with the actual grounding of symbolic manipulation of logical structures; all that is taken care of by someone else who is looking at input/output issues. Ever since the advent of cognitive psychology, we know this to be a mistake. There is no clean division between perception and cognition. The latter infects the former, thus explaining illusions and the like. "Perception is active" is the motto today and we cannot suppose that the manipulation of formal systems can be looked at without regard for how the symbols get their significance. Stevan Harnad has examined this problem and he points out that when one strips away the sensory-motor areas of the brain, one is not left with very much at all and certainly not:

> "... some homuncular computational core-in-a-vat that all this transduction [is] input TO. No, to a great extend we ARE our sensorimotor transducers and their activities, rather than being their ghostly computational executives."[10]

Harnad is absolutely right to attack this as a fundamental flaw in traditional symbolist thinking. It is conceptually naïve to suppose we might just "bolt-on" I/O modules to symbol systems. In fact, it seems surprising that anyone in AI ever really thought that it would be possible given the general reaction against passive perception that arose in tandem with cognitive psychology. A great deal of cognition indeed goes on during what used to be considered simple, mechanical I/O and this is a fact that Har-

[10](Harnad 1993)

nad uses to launch into his exposition of the Symbol Grounding Problem. The problem, as Harnad formulates it, is that of determining how the basic symbols in a formal system get their "meaning". Harnad asks how the arbitrary, meaningless symbols of such a system can be grounded in anything other than other arbitrary, meaningless symbols. He calls this the symbolist "merry-go-round" and likens the problem to trying to learn Chinese as a first language from a Chinese/Chinese dictionary. It would just be, according to Harnad, a progression from one meaningless symbol to another with meaning never entering the picture at any point. The symbols are never "grounded": they never link to the world of experience and thus can never say anything about it. The system is "hanging from a sky-hook" and never touches the ground of reality. Therefore, a pure symbol system can never be adequate to human cognition; the Realm of Reason is not enough in itself. Symbol systems are in need of grounding in a non-symbolic way in order to prevent the merry-go-round and to link them to the real world and thereby endow them with implicit meaning. The problem can be expressed more clearly thus: the beauty of the symbolist approach is that it takes all of the trouble out of dealing with messy domains by providing a mapping from the domain into a formal system. If one can make this mapping in a systematic way, one can then manipulate symbols that can be said to stand for objects in the domain in a way much more suited to the apparatus with which you must work. More concretely, if the brain can provide a mapping from the world to arbitrary symbols, efficient computational methods for manipulating the symbols will preserve the relations of things in the world. The real benefit comes as the computational methods are purely syntactic, they only need to know the general shape of the symbols rather than the complicated issue of their references, meanings etc. Of course, as Harnad points out, this is only of any use when one can interpret the symbols and their manipulations in a systematic way in the real world. This is not a problem for the observer of such a system as he can impose such

an interpretation. However, for the system itself, the only way that this can happen is if it has a link with the real world at some point in order to give the mapping a reference point, a grounding. Consider figure 7.2. This

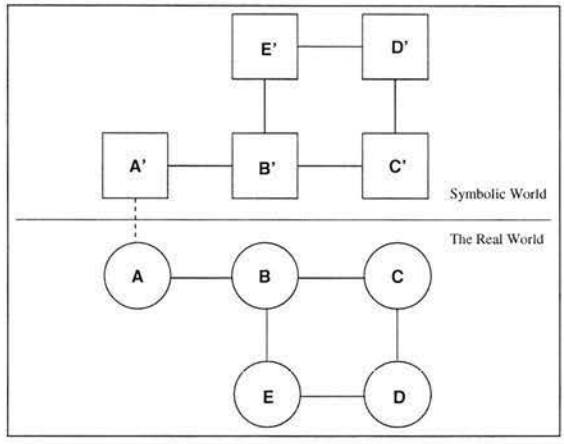


Figure 7.2: *Symbol/World Mapping*

rather simplistic diagram serves to demonstrate an important point. The topology of the symbolic world can only be a useful mapping of the topology of the real world if there is a link (the dotted line in the diagram) at some point that sets up an initial interpretation for some part of the symbol system[11]. In the diagram, we can see that the structure of the symbol system is topologically isomorphic to the structure of the real world. The crucial point is that this is only possible through the existence of the dotted link. This provides a *non-arbitrary, non-symbolic* grounding for the entire symbol system. It *must* exist for the symbol system to have any claim to real meaning[12]. McDowell realises this problem in a more abstract sense; he expresses it in the requirement that the Platonism of the Realm of Reasons must be *naturalised.* It is instructive then, to relate how the way in

[11]Harnad makes it clear that the systematicity requirement for symbol systems is strict enough to ensure that the chances of a symbol system being systematically interpretable in itself as referring to the real world without a link to it is negligible. The symbol system has to be grounded in the world somehow.

[12]Note that there does not have to be only one link. It is just there there must be at least one such link in order to ground the system.

which McDowell fails to support this sort of stratified view translates into a similar pattern in Harnad's failure to support a hybrid formal/non-formal system. This is important as it demonstrates the fundamental nature of the difficulties underlying a large portion of Cognitive Science as a result of the considerations given in Chapter 5.

7.3.1 Problems With "Grounding Symbols"

Why would one want to ground symbols? Why not simply have a thoroughly non-formal system and do away with the troublesome formal stratum altogether? The reason is that Harnad is convinced of the necessity of Fodor and Pylyshyn's criteria of explanation in terms of systematicity and compositionality. He is convinced that the formal strata is the only one possible that could account for these features and thus wishes to retain it. We ground a symbol as follows. Our perception performs fundamentally two tasks: identification and discrimination. Discrimination is merely the ability to tell one thing from another but identification allows us to classify things as one of a *type* previously experienced. Identification is allowed by the *iconic* nature of perceptions. This is in direct contrast to the formal stratum – as it must be in order to be a solution to the formal's impotence in grounding – as an iconic representation maintains features of the actually raw sensory data; the iconic representation of a cow might be a vaguely cow-shaped blob on the retina for example. Identification,

> "... somehow reduces iconic representations to the invariant sensory features that will subserve successful categorization."[13]

and categorisation imposes a structure on perceptions: so-called "categorical perception" (see Chapter 5). This means that they are "discontinuous" with the physical data: these categorical representations are distortions

[13](Harnad 1992)

or "warpings" of the analog sensory data and identification is the process of warping sensory properties in order to keep a useable set of categories for future use. Notice the strongly Kantian element here which is directly comparable to McDowell's desire to adhere to this sort of requirement. This categorisation is a natural process as the analog world is so continuous that if we did not categorise our input, we would be at pains to develop the most important concept of all: regularity. This phenomenon makes possible the crucial step in the solution of the Symbol Grounding Problem. The categories we contrive from our data can be given unique, arbitrary names and the names can be our "elementary symbols". These names are grounded as they follow directly and non-arbitrarily from our iconic representations. Harnad considers the groundedness property a transmittable (or perhaps "transitive") one and so any symbolic representations created using these grounded symbols will also be grounded. The symbols thus constructed can be considered grounded in their relation to a unique determination of category membership. The following example is given: Suppose a robot could distinguish objects and categories Turing-indistinguishably from a human. In particular, suppose it could identify a horse on the basis of its iconic representation and "stripes" on the basis of its iconic and categorical representations. When presented with the symbol string:

zebra = horse & stripes

it could "decode and use" this symbol string even though:

> "... "zebra" was a previously undefined symbol and unencountered object."[14]

This is because "zebra = horse & stripes" is grounded in the non-arbitrary categorical representations given by its experience of horses and stripes. Note it is not claimed that "horse" or "stripes" are fundamental categories.

[14](Harnad 1992)

Harnad *is* committed to the existence of ground-level fundamental categories but he does not suggest what they might be. In any case, these categories are always revisable as he subscribes to a neo-Popperian fallibilism dubbed "approximationism". This is the doctrine that our categories, once created, are only approximate and converge to relatively stable states. The equilibrium can always be upset by a recalcitrant identification: our categories are revisable. Harnad assumes that there are certain "invariant" features of perception that constitute the basic categories in this picture. This is necessary to get the whole thing started; we must "ground the grounding procedure", so to speak. Now, I agree that there is a necessity for postulating, as a minimum for an explanation of human cognition, invariants in the environment[15] but this is completely at odds with the retention of the formal stratum in the explanation since Harnad thinks

> "... introspection is unlikely to reveal the mechanisms underlying our robotic and cognitive capacities ..."[16]

Harnad is motivated in his insistence on the inscrutability of such invariants by the problem of vanishing intersections. This is the age-old AI and philosophical problem of determining what properties are essential to all instances of a concept. What, we may ask, are the invariant features that pick out dogs, chairs, or more famously, games? Obviously, if we hold that the invariants are inscrutable, then this problem is, if not explained, sidestepped. So, Harnad believes we cannot discover these invariants by introspection. This is rather strange as they combine, in apparently rather simple ways if the example above is representative, to form the formal level. Formal systems are, as Harnad knows, designed to be compositional so that they can be broken down and examined in exactly the manner required to get to their basic elements. Thus, inscrutable in-

[15]There is also a *mathematical* necessity if one is to employ tensor calculus. See Chapter 5

[16](Harnad 1992)

variants are not compatible with a stratum issuing from them with the explicit feature of being scrutable. Environmental invariants, I argue, are indeed inscrutable, but not because of the problem of the vanishing intersections. That particular problem is a result of the way that our manifest concepts are constructed by the brain. Classification and thus conceptualisation is an *output* of the system and thus is a contravariant execution vector according to a tensorial model. This means that the components of our concepts will not be obtainable from the concepts themselves, simply by virtue of the way the information is combined to give them. The unhappy marriage of inscrutable invariants and a formal level stratum is Harnad's attempt to keep the Realm of Reason and the Realm of Law connected to each other. It is clear, once the implications of the relevant brain research are taken into consideration, that this results in a contradiction between the necessary inscrutability of invariants and the scrutability of the formal system they result in. At least Harnad wants to avoid the mistake of the philosophers who he thinks have,

> "... from the vantage point of their armchairs, based on introspecting about the definitions and sensory properties of abstract categories."[17]

In this he is correct but it is not possible to really avoid the problems if one has a view that allows a formal system that, by design, is bound to want to let you split up the "data" in a perspicuous manner. The naturalist streak advocating robotics research, neurobiology or whatever, cannot tolerate a formal system that seems to make it easy to find "atoms" in your data; it spoils the whole point of doing this sort of naturalised investigation. The problem is not, as Harnad seems to think it is, armchair speculation as such. It is the stratified view that it goes hand in hand with. The strata have to be very different but then they contradict each other as we have seen with McDowell.

---

[17](Harnad 1992)

7.4 Conceptual Semantics

The late 1980s saw quite an explosion in various theories designed to move away from the traditional externalist and extensional semantics. We have seen Gärdenfors and his semantics in terms of "conceptual spaces" and Lakoff's "conceptual linguistics". I would like to dwell a moment, as a concluding note to this chapter, on Ray Jackendoff's "conceptual semantics"[18]. The motivations behind this are in accordance with the issues I have raised so far:

> "... since one's construal of the world is heavily mediated by complex computational processes which have little if anything to do with language, reference in natural language is likely to reflect the internal representation of the world as least as much as it does the external world *per se*."[19]

Jackendoff wants, essentially, to dig a little deeper into our manifest patterns of output and to take seriously certain of them that traditional semantics ignores. He, for example, wants to take the "pragmatic anaphora" involved in sentences like "I can do *that*", which accompanied by a pointing action, refers to, say, an action by another, as serious evidence as to the ontology we must impute to natural language. We must look at patterns like these and conclude, Jackendoff argues, that things like "events" and "places" are an essential ingredient in the ontology of a natural language semantics. This, I think, while having a reasonable motivation, is incorrect. *Any* patterns in our outputs are suspect as *causal* components in our cognition. While it is true that traditional semantics has biases built in as regards ontology, this does not mean that extending the patterns we take seriously in our language output will give us a better understanding of the processes involved in generating that output. It may, and as Jackendoff demonstrates *does*, provide technical and formal advantages if we take cer-

[18](Jackendoff 1983, Jackendoff 1985, Jackendoff 1988)
[19](Jackendoff 1988) p. 83

tain patterns as evidence of basic-level ontological categories, but this has no implications for the way in which we *actually* construct and employ language. Jackendoff assumes that there is a level of *mental representation* located somewhere between the level of neurons and the formal systems that provides a grounding for the patterns we see in the latter. This is basically a stratified view with three levels, the third intermediary has the task of making the link from the formal to the neuronal more plausible. This cannot work since the intermediary level that Jackendoff endorses is simply a more liberal analysis of our manifest output patterns. "Cognitive semantics" has failed to realise that it is not the *type* of patterns we find in our output that is misleading; it is the use of *any* of these patterns in forming causal stories about the process of cognition and language in humans. There is a general point to be made about the possibility of removing objections to stratified theory by adding more strata and this will be brought out in the following.

In this chapter, I have shown that the philosophical assumptions of a stratified approach runs deep, even into supposed radical alternatives that aim more towards the type of view I advise. I shall now conclude by drawing together these assumptions to see what is left and what philosophical implications there are for Cognitive Science and its branch of natural language semantics.

Chapter 8

Objections and Conclusions

8.1 Objections

Here I shall tackle the main objections to the view described here that might come from the community of those engaged in the practises of formal semantics. It is important to first draw the following distinction which has been addressed elsewhere in this thesis but not quite in the way I now approach it. We must divide the field of "formal semantics", broad as it is, into certain practises. I do this because I am wary of trying to address pseudo-objections since I think that much of this large community would actually be completely apathetic to my view because they are engaged in *reconstructive* enterprises rather than developing explanatory accounts. Viewing formal symbol systems simply as useful models of "high-level" cognitive processing is not something with which I have any quarrel. Such a view is extremely useful for implementing many Information Technology tools and language engineering products. My concerns are about the *causal* role of such systems and the explanatory use made of them and so I have little to say about and little to fear from such an attitude. A lot of the information-theoretical approaches seem to me to be like this. The emphasis seems to be on developing a view based on some broad notion of "information" that might include not only the usual views found, for example in mathematics

and computer science, but also some idea of linguistic "meaning" and more colloquial versions. The impetus is one of integration and regimentation of similar concepts in order to provide a cohesive framework. An example of this is Channel Theory ((Barwise & Seligman 1992, Barwise & Seligman 1993)) which attempts to model "information flow" by use of a theoretical construct called an informational "channel". This is influenced by ideas from Dretske[1] and is explained as a model of errors in information flow that enable better implementation of, for example, fault-tolerant software.

The real objections will come from those who argue an explanatory and causal role for their symbolic formalisms. A prevalent objection, as evidenced by certain arguments often given to support symbolic semantics[2] is that the notion of context of which I am making so much should be addressed at the level of symbolic representation since, if we do not do this, there seems no way to do it and Cognitive Science misses something essential out. I reiterate that if this is an argument at all, it is certainly not one in favour of a symbolic treatment of context but rather, on my view, a plea to be allowed to err consistently. Situation Semantics is a modern semantic theory which advocates just this and I examined objections coming from this school in Chapter 4. I have addressed objections stemming from fundamental tenants of the formalist position in my discussion of Fodor and Chomsky previously but now I wish to turn to more recently occurring possible objections from working semanticists.

An objection might be raised that I am collecting an awful lot of different theories under the heading of "formal" and dismissing them all with a broad sweep. There are, it might be argued, a lot of differences between theories broadly termed "formal" or "symbolic" that are relevant to my objections to that whole area, thus rendering my views only partially effective. One such argument would probably invoke the crucial differences

[1]Especially (Dretske 1981)

[2]I noted this in specific reference to vonEckardt in Chapter 2 for example.

between traditional systems of formal logic as applied to language - the "Davidsonian" programme and those systems following Montague which considerably enrich the formalism and the motivation for its use. For example, (Lappin 1997) holds the view that the Montagovian tradition is a more successful approach to natural language semantics as it does not have the strict logicality requirements of the essentially first-order Davidsonian model. This relaxes some of the formal constraints on the account and allows in a wider range of possible explanations. This is seen as a major difference and one which might be taken as an objection to my view as I do not consider this type of difference in my rejection of the general formalist level of explanation. In order to reply to this, I must begin by saying what I think is in contention here. It is the fact that a richer set of semantic types is more able to account for semantic distinctions than a more restricted set. Now, I cannot see why anyone would want to deny this. The further notion is that the crucial seemingly non-symbolic aspects I dwell upon can be addressed by developing the symbolic notions to encompass more and more subtle and complex aspects and interactions of more basic semantic terms. Montague was a key figure in being able to account for many more semantic phenomena by extending the allowable semantic types of the formal language and providing rigorous justifications for the manner in which this was undertaken. The natural reply to this is that the differences between such approaches, whilst differing radically in semantic typology, are differing with respect to a notion of "type" that they both share. I cannot allow that a richer set of semantic types will help to develop a notion of context, provide an interface between biological and formal levels of description and also aid in a full descriptive, explanatory and causal model of cognition since the whole point of these notions is that they require something necessarily disjoint with such a thing. If one builds up these desirable results through richer semantic types, in terms of explanation, there really only *are* semantic types and thus you declare that

there really weren't any of the problems you needed richer semantic types to solve! No semanticist takes seriously the idea that there is no explanatory worth in "context" and brain structure. If semantic types can *in a causal theory* be used to explain connections to such things, then one is *not* taking them seriously. A causal theory is something very serious and if one attributes causal powers within human cognition to semantic types then the price is that there is little work for anything else to do. Since work is certainly done by context and the brain as far as cognition goes, this is almost a *reductio ad absurdum* of a formalist position that claims to explain in the sense of providing a causal account. Conversely, as I have stressed all through this work, a non-causal account is not really very interesting to Cognitive Science as an explanatory discipline.

A possible major objection arises from the more recent formalist schools concerned with representations that have more radically modified or extended the traditional logic formalisms. A large concern has grown up around the notion of "structured representation", particularly in the area of discourse analysis. The main protagonist is Hans Kamp and his "Discourse Representation Theory"[3] has been widely accepted as a model of good research in formal semantics. Kamp's concerns are directly relevant here as he has quite radical claims for the cognitive status and indeed causal properties of the representation he espouses. The key here is that the representation is highly structured: the explicitly graphical mode of presentation of the theory is essential to its theoretical claims which are centred around the fact that semantic structure is attributable to syntactic units such that they combine in a combinatorial manner. There is a central construction algorithm which takes the structured representations of syntactic units (usually parametric as in Montague grammar: the model theory is compositional in a fairly standard way) and combines them to form larger and larger representational structures up to the level of multi-sentence dis-

[3]The canonical statement of this can be found in (Kamp & Reyle 1993)

courses. The whole power of this is in the structure assigned to syntactic units. In a sense, this is nothing new as the same is true of Montague grammar. The difference is that one of the motivations for the particular graphical representation chosen is that Montagovian models become horribly complex for large sentences and their structure becomes so complex, it is not perspicuous or appealing to methodological parsimony. Leaving aside the question of whether this is really a suitable motivation for a new explanatory paradigm, it at least is clear that the notion of structured mental representation has reached a certain plateau here with it being seen as so important, the whole theoretical impetus is based around its desiderata. Philosophically, the issue is the same as with Fodor who also insists on structured mental representations. I have arguably thrown this out since I have questioned compositionality which is so closely related to structured representation, it is inseparable. I would reply that there is no disagreement about the necessity for a notion of structure. There manifestly is such a thing in our cognitive life. But the question is, where does this structure come from? Kamp is clearly arguing that it is derived from the formal level he aims his theory at. Kamp and Reyle's whole notion of structure is predicated on the basic case of an "initial" sentence in a discourse which has no information structure (i.e. context) to which it relates. This of course is an impossible situation which they admit themselves:

> "Almost all interpretation relies on antecedent information, deriving from general knowledge and from earlier communications between speaker and the recipient. We will have more to say about this in Volume 2."[4]

The fact that this is relegated to a footnote is as interesting as are many other aspects of this admittance. Firstly, no mention is made of the role of cognitive categories in the contribution of antecedent information. These are far more fundamental and even constitutive of the effect that

[4]Footnote to p. 54, Ibid

"general knowledge" and "earlier communications" have on communication. Further, the elements that *are* mentioned are specified in terms of a formal account identical to that which they are supposed to ground. That is, it looks very much as if the "general knowledge" and "earlier communications" will be modelled in DRT too even though it is exactly this theory that requires this completely unrealistic notion of "initial sentence". I think it germane to point out that Volume 2 has not, at time of writing, appeared. It looks, in retrospect, an awful lot like Terry Winograd's famously never written book "Semantics" which was to follow on from "Syntax".

My argument, to recap, runs as follows:

- Structure that we posit to play an explanatory role in human cognition must have causal powers. That is, different structures must have different effects on our cognition. It must do something.

- Structure posited at the level of formal symbols cannot have these causal powers because if it did, it would not take seriously the idea of non-formal and non-symbolic elements relevant to cognition which is really beyond question.

- The reason it would not take these seriously is that its causal powers can only be phrased at a level which is, by definition, completely orthogonal to these other elements.

- Thus, there would be no connection between formal semantics and notions like "context" and obviously relevant neurological data.

- So, keeping its causal powers to itself renders formal semantics isolated from matters which are, by any reasonable standards, far more obviously relevant to cognition than formal symbols: no-one would doubt that brain's topology is relevant and no-one with any grasp of the history of philosophy would doubt that context is. Many doubt that formal symbol systems are. Allowing causal powers elsewhere

renders formal semantics explanatorially irrelevant. This is exactly the tension in Cognitive Science.

So, I must keep a notion of structure but reject a formal semantics approach to this idea. This is very disturbing to the whole practise of Cognitive Science since, as I have argued, with this goes an awful lot of methodology too. We cannot give detailed systems of semantic categories, types, examples of derivation, constructions etc. any more. I keep the notion of structure by talking of structured spaces which give rise to the structure we see in our language and thought. This, I have argued, is really the only possible way to see things if we take seriously a holistic and integrated view of cognition. Structure in our surface thought and language *is* present but only by virtue of much more abstract structure at the neurological level. Only this can be seen to be causal. It is upsetting to Cognitive Science as often practised as the kind of structure I am talking about is not specifiable in the usual logical terms. That this is a large problem demonstrates the methodological grip that formal semantics has on Cognitive Science. So, my reply to the advocates of modern structural formal semantics is that I have no quarrel about structure, just about where it originates and where its causal properties are operational.

I now tackle the most vital criticism that might be raised against my view. I have argued that the prescriptive nature of the logical machinery enables it to be able to cope with all possible things that it takes to be data; that is, I cannot see that there any limits that a formalist might reach with his work given that it is, for reasons I have described, guaranteed to be able to account for any linguistic data. Anticipating this problem, Edward Keenan[5] has attempted to reply to a form of my objection that comes from Putnam[6] and is given credence as an argument against formal semantics

[5](Keenan 1996). See also (Keenan 1976, Keenan & Faltz 1978, Keenan 1987)
[6](Putnam 1981)

by Lewis[7]. Putnam's point is that there are an unlimited number of semantic models which differ yet are extensionally equivalent. This can be seen to be a demonstration that there is no real (i.e. external to formal semantics) criteria for deciding between formal semantic models of language and cognition. I am in agreement with this conclusion although for different reasons given earlier. Keenan's goal is to show that this is not a problem and that there are actually good reasons, at the level of formal semantics, for ruling out certain approaches because he claims to be able to distinguish a rigorous notion of "logical objects". This hinges around the notion of "Invariance under Automorphisms" which establishes the class of things that are invariant with respect to logically important transformations. Now, if Keenan is right, he has established an invariance at the level of formal symbolism which might be used to claim that this is the empirical data of formal semantics, thus contradicting my argument in Chapter 4 that formal semantics really has no respectable notion of empirical data. After all, I put great weight on the notion of invariance myself and so, such a notion should at least be acceptable coming from the formal semantics community. Keenan makes a crucial point though when he says:

> "One might have hoped for a more "absolute" notion of logical object, one that did not depend on the choice of universe or type."[8]

He then goes on to say how the notion is relative to the ontology of the semantic model and the type of object being considered. This is very important as it shows that the standard for defining this concept of "logical object" depends crucially on the prior acceptance of the central concepts of traditional formal model theory: namely the notion of "domain of interpretation" (the universe or ontology) and the notion of "semantic type". If this notion, so important for providing an independence of formal semantics, is

[7](Lewis 1984)
[8](Keenan 1996) p. 19

definable - as Keenan admits - only in respect to the methodological fundaments of that discipline, then, I think it is clear that "logical objects" are in no way going to provide any real invariant that can be said to establish a realm of data for formal semantics. They are only an invariant of formal semantics and thus cannot provide a foundation for this very practise. You might classify the set of all unicorns on the basis of the "invariant" of "horn-owning" but that doesn't mean there are any. Only real invariants can provide such a basis, that is, invariants that are such in virtue of something other than the theory you are using them to establish. As a corollary to this argument, it is a feature of tensor theory that the invariants modelled must have a real existence in the world i.e independent of their description in the theory. It is clear that the idea of a "logical object" does not have this.

A less serious objection but one which will occur to all fair-minded readers sympathetic to formal semantics is this: formal semantics is surely useful in helping us to understand language and cognition. To reject it completely is going too far; it simply must be of *some* real interest. This is not clear to me at all. It is of *some* interest certainly, just not one that can play an explanatory role in Cognitive Science. But, the objection is stronger than that. It asks whether a model, describing causal mechanisms or not, might not help us understand a phenomenon. Normally, I would say yes. However, I think the case with formal semantics and Cognitive Science is so unique that the normal rules do not apply. Primarily, the difference is that the hindrances outweigh the helps: the methodological dogma that results from taking formal symbolic models even as helpful models is too high a price to pay. Secondly, I have argued that a model that is depicting the non-decomposable behaviour of the workings of the brain is really of no use to understanding how cognition is caused which I take to be the primary goal of anything called Cognitive Science. Describing surface structures *is* of use in helping cognitively distressed people since, as I have argued, these

surface structures of cognition re-enter into the pool of information about the world in a way which makes them indistinguishable from any other information. Thus, these structures can *effect* the way we think. This does not mean they have causal powers in the *genesis* of cognition, merely it is a recognition that we are interacting with our environment in a very deep way. Basically, our cognitive structures play the role of ingredients rather than as the recipe itself – a recipe-like model might help us whip some order into our cognitive life. Again, the effect is more therapeutic than explanatory. Models help us *do* things far more often than they help us *explain* things. Of extreme importance is that the ability to *do* implies nothing about the ability to *explain*. One can, for example, enable somebody to program in a high-level computer language by giving them a model involving little men (variables) running about between rooms (functions) or telephoning from room to room (passing by reference). One can imagine this model extended with all sorts of ingenuity to cover the usual programming idioms. It enables one to program. It does not explain anything at all about how the language really gets things done. One may say though, "but it is isomorphic" - that is, getting back to the real issue - formal semantic models are isomorphic to the "real way it is done". This is not taken seriously by anyone though since the opening part of any material on, for example, connectionist language processing, laments that we don't know what semantic concepts like "proposition", "predicate" etc. map onto at the neural processing level. That is, that sort of isomorphism is extremely contentious. When it is claimed, the formal symbolic level is usually claimed as "virtual" or "reducible" to the neural. Again, this is the danger that has permeated my thinking on this whole topic: as soon as the formal semantic level is close enough to the neural level (the limiting case is complete isomorphism), it collapses into this level and disappears as an explanatory element. If it stays far enough away, it becomes too isolated to bear the pressure of the canonically relevant empirical and philosophical evidence.

8.2 Conclusions

It is clear then, that there are certain assumptions that the formalist approach to semantics in Cognitive Science cannot be upheld in the face of empirically motivated brain research. The philosophical implications of the way the brain seems to perform processing are quite far-reaching and touch, as we have seen, not just on particular features of semantics but on the very idea of a stratified picture that allows for links between "levels" of a fundamentally different kind. In particular, there is a temptation to reify the manifest patterns in language and to suppose them to constitute an independent and *sui generis* level of explanation. I have argued that the reasons for this are as follows. Firstly, the formal systems employed in the traditional approach to language and semantics were designed as reformative tools to aid the coercion of natural languages into canonical artificial languages, in order to help the progress of a science free from the misunderstandings possible otherwise. The mistake has been to employ these systems as descriptive and explanatory, thus reifying the categories, concepts and distinctions inherent in them which were regarded as essential features of *desired* languages rather than existing ones. The effect of this has been to force language and thought into a mould where the only possible option is the postulation of an independent level of explanation corresponding to the coerced features. This resulted in the famous failures of symbolic AI earlier this century in its attempts to model knowledge in terms of symbols and logical manipulation. Chomsky's competence/performance distinction plus the advance of research into formal systems has enabled such systems to survive still, however, since anomalies can now be addressed by postulating another "deeper" level of symbolic processing involving more formalism specifically designed to solve particular problems. The competence/performance distinction allows modern linguistics to avoid altering fundamental assumptions regarding formalism when the obvious manifest

features do not conform.

Secondly, there is a traditional suspicion of the notion of a pattern that does not allow an explanation at the level the pattern is manifest. We are not comfortable in calling something an "explanation" unless it is composed out of units we understand as being related to the *explanandum*. This is explicit in, for example, Fodor's requirement that the contents of mental states be semantically specifiable; they are to be specified in the language of a formal system, specifically a compositional system. I have argued elsewhere that it is exactly the requirement of compositionality that prompts us to favour such explanations in the first place[9] and thus the inference from the formal level of explanation to certain models of language is exactly backwards in this respect; it is our formal systems that dictate our theory, particularly because they are designed to be reformative systems and not descriptive of natural languages. As such, they have, built in to them, assumptions about the patterns manifest in language and the ways in which the patterns are constructed. Chomsky has confused the whole issue with the above mentioned competence/performance distinction. This is a model where the actual performance of language – its manifest patterns – are not seen as a guide to its cause. This is something I have been urging. However, the solution Chomsky proposes renders the "real" basis of data on a hypothetical level of competence which is, naturally, more conducive to modelling with the formalisms he favours. The causes of language are, in this picture, even less related to anything obviously seen as "evidence" than with a theory based on performance. Thus, even though the manifest patterns in our output are no good guide to their causes, the patterns in a hypothetical level of description are certainly in no better state. Patterns at such a level are simply imposed by the description of the level. This clarifies the problem considerably. The reason that manifest patterns are no guide to their causes is due to the way that the patterns are indeed caused.

[9](Kime 1996)

I have argued generally that this involves informational dependency of a certain, complex sort where each element of the cause is dependent on some of the others for its effect. If this exact matrix of dependencies (the matrix describing a metric tensor) is not known, then nothing can be inferred from the patterns. A hypothetical stratum such as Chomsky's is designed to be amenable to the sort of formal analysis that "performance" denies. Thus, in its design, the perspicuity desired will remove the essential interdependent feature of the manifest output and thus will be, quite simply, an *a priori* stipulation about the components of the output. If one takes the manifest output as data, nothing can be inferred from this alone. If one attempts to create a level of description not having the feature of interdependency of causes, then one is in conflict with the implications of brain research and the essential characteristic of our output manifestations. One cannot bypass a problem by simply allowing oneself to re-describe it in more manageable terms but missing out the central feature that makes it so problematic.

It is important to see what is being suggested here. It is not that the regularities of language and thought are being modelled *badly* by formal systems. It is rather that the patterns themselves are not *evidence* of anything and thus the formal systems are merely describing regularities that we have no good reason to suppose indicate anything about the causal structures giving rise to them. In fact, if we believe two rather plausible claims: that the brain is reasonably seen as employing certain highly complex geometrical processing and that these are involved in the manipulation of information tentatively called "semantic", the mathematics involved ensure that the manifest patterns cannot possibly, by themselves, give us any indication as to their causes.

The patterns we take to be important in language and thought lack an essential characteristic of the patterns that a natural science aims at explaining. Such patterns are not, in an important sense, *pre-theoretic.*

Much has been made of the movement from common sense to science by, e.g. Quine and this is a vital ingredient in accounting for any variety of empirical content. In natural science, we encounter certain patterns without any contrived scientific terminology or distinctions and aim to investigate them by designing such. In linguistics, our data is much more dependent on having a theory to talk about them in. The regularities are in terms of certain chunks – "verbs", "modifiers", "clauses" which are part of the formalisms and theories designed to account for these patterns. In a natural science, we have general tools like mathematics that presuppose much less specifically, if at all in some cases, anything about the structure of the patterns in the data. For example, we might mathematically model patterns in plant distributions using concepts such as "average", "density" and the like. These are not nearly as explanatorially loaded with suggestions of causal mechanisms as terms used in linguistics such as "property", "proposition" and the like. This is not to deny that we may err in supposing certain patterns ourselves in the use of our tools through biases of certain sorts but the *tools* themselves are very abstract and do not force us into a position where we are forced to see certain patterns. This, however, is exactly the case in linguistics and formal semantics as the tools were not designed as general systems but as specific, normative moulds into which we pour our data. Thus, the "basis" of formal semantics and linguistics is not pretheoretical in the sense of being free from the impositions of the structures implied by the tools used. Thus, there is, as Quine has pointed out, no real "empirical base" of any sort for natural language semantics. This makes underdetermination much worse for such theories, see Chapter 7.

So, while one would not want to deny that the patterns we term as involving "nouns", "predicates", "quantifiers" etc. exist in our manifest language and thought, one *would* want to deny certain claims as to the sort of basis this constitutes for an explanation of human language and cognition. We would expect patterns in our output and we would expect this to bear

some relation to patterns in both our input and patterns in the structure of our processing mechanisms. However, since Kant, we are suspicious that the latter are considerably preponderant over the former and that their action is ubiquitous. Thus the task is to separate the different contributions to the output patterns. Since the patterns contributed by the processing itself will not be introspectively available and thus, neither will the input patterns, simply put, it is näive to put faith in analysis of the output which is essentially exactly what formal semantics and linguistics does. Chomsky's move of giving linguistics a more regular level of "competence" as data is, as I hope is obvious from the above, of no use as it sidesteps the whole issue involved in the troublesome feature of the manifest output and thus weakens its claims to relevance in the case of human language and thought.

A famous view, stemming most obviously from early Wittgenstein is the "picture" view of language where the structure of language and thought mirrors the world. From an evolutionary perspective, this is rather uneconomical as, if there is structure in the world, re-representing it again internally is a heavy overhead in processing and therefore time. Much better to use the information already in the environment to enable you to act. Formal semantics often finds itself in trying to be adequate to the structure in the world. We like to make our semantics square with the way the world is presented to us. Thus we worry about the temporal order of pronoun resolution, quantifier ordering, researchers like Harnad try to marry iconic representation with formal systems that preserve iconic features etc. This is a strange occupation since it effectively completely forgets and bypasses the structure imposed by the internal processing by the brain. Trying to map our output onto our input with no concern for the mediation by Kant's categories of perception is ignoring possibly the greatest contributor to order in our phenomenal experience. It is like attempting to sneak in all of the mediation of the brain into the structure of the outputs; we try to cram

everything into the formal system we see implied by patterns in the output and get ourselves into horrible complications as a result. There is no substitute for empirical work in examining the mechanisms for mediation of input/output in the brain. Attempting to do this by modelling the resulting patterns in an artificial system of formal symbols results in exactly the problems AI has experienced over the last fifty years; computational explosion, monotonicity problems, the messiness of real language and so on. This is because, although there are patterns we might model on a formal system, the *constraints* on the patterns are not even manifest at the level of the formalism. The constraints are, according to the geometrical model advocated, abstract informational invariants of the systems involved in the brain. Thus attempts to constrain, say, inference to "natural" patterns have been rather basic and the results rather disappointing ((Dreyfus 1992) has some good examples). Basically, all of the information we need to properly account for language and thought is simply not present in its manifestation, which is simply to say that Kant was correct in that much of the important material is in the mediation and not the output. Some have seen part of this problem and have, as a result attempted to bring into their theories certain topological ideas as an attempt to explain the ubiquity of certain patterns. Lakoff (see Chapter 4) is a good example of this although he fails to realise that the constraints are not specifiable at the level they actually have force. If they were, we could have a purely formal system that controlled the application of its rules and inferences with no regard for the role played by the brain in human cognition. Lakoff's failing is, I think, a direct result of his linguistics training that make him favour explanations whose components are of the level of formal systems; Lakoff still favours a system that can satisfy the desire for compositional models for example[10].

I have expressed the overall problem as one of spurious stratification of explanations in Cognitive Science. Part of this problem is of an abstract na-

[10]See (Lakoff 1988) p. 306

ture to do with the relationships between the strata. Part of it (see Chapter 3) is a more pragmatic issue of relevance to embodiment of the supposed levels of explanation in real systems. Since, as Quine argues in (Quine 1972), explanation is a matter of delineating *causes*, stratification is essentially a way of *generating manifest causes to aid explanation*. We do not feel that we have explained until we have shown a cause but in cases where causes are essentially inscrutable, we tend to fabricate levels at which they *are* scrutable. It is not felt to be sufficient to know that there *is* a cause; it is necessary to show it. This is the basis of Fodor's view that giving up propositional languages of thought is to be resisted on the grounds that, without it, we would not have a story to tell about cognition. This is the difference between those who think there might *be* a story to tell, as yet untold or practically untellable and those who actually require that a story be told now. Fodor is of the latter sort. He prefers a story, however fantastic, to none at all[11]. The unfortunate consequence is that a story, once told, becomes embedded and thus limits the telling of others[12]. The comfort of having a story comes to outweigh the importance of knowing that things are more complex than it appears; possibly more complex than it is possible to understand. I discuss this latter point below.

McDowell also prefers a story and his is more grandiose in style; it requires a stratification of two fundamentally different but linked aspects which are what we might picturesquely call the Platonic forms of the strata in Fodor's more technical distinction. The reasons for rejecting both of these views and thereby the basis of traditional formal semantics and Cognitive Science is slightly different in each case and it is illuminating to examine the differences in order to see what is essentially incorrect about the urge towards stratification in such endeavours. Interestingly, the criteria by which we rejected McDowell's stratified view in Chapter 6 is used,

[11]This is explicit in (Fodor 1987).
[12]This is addressed in (Kime 1996)

by McDowell himself in his critique of Quine[13]. This is based on Davidson's supposed identification in Quine of an untenable dualism. In a commentary on Quine[14], Davidson notes a section from *Word and Object* where it is written:

> "we can investigate the world, and man as a part of it, and thus find out what cues he could have of what goes on around him. Subtracting his cues from his world view, we get man's net contribution as the difference."[15]

Davidson sees this as establishing a "third dogma of empiricism", that of the dualism between scheme and content. This dualism is that of supposing that there is a a separation between that which is cognised and the *framework* in which it is cognised. We have a certain *way* of looking at things which are independent of the way of looking. This, for Quine, is the basis of empiricism and is not meant to imply that the relationship between scheme and content is simply, separable or scrutable. However, McDowell agrees with Davidson that this is an appeal to a neutral empirical basis which has little relation to the conceptual realm. Thus, the elements of this dualism being exactly McDowell's Realm of Reason and Realm of Law, there would be no rational interaction and thus this, for McDowell, constitutes a reductio of Quine's position. So, McDowell uses the conclusion that Quine's dualism of empirical content and conceptual scheme fails to maintain a connection between the two and thus casts doubt on the very idea, to borrow a phrase from Davidson, of there being such a dualism. Notice, that this is the form of argument we used with respect to ostensives in order to demonstrate that McDowell's own dualism of Realms of Reason and Law could not connect in the desired way and were therefore part of a spurious division. McDowell's problem is in the *connection* of the levels of the stratification. In trying to account for the *single* phenomenon of human

[13](McDowell 1994) pp. 156–157
[14](Davidson 1990)
[15](Quine 1960) p. 5

cognition and its relation to the world, he employs a *dual* of explanatory levels. In order for them to coexist as an explanation of the same thing, they must be linked but their postulated differences conflict with this to the extent of contradiction. Davidson himself echoes this mistake in the context of further critique of Quine's third dogma:

> "Different points of view make sense, but only if there is a common coordinate system on which to plot them; yet the existence of a common system belies the claim of dramatic incomparability."[16]

The mistake is essentially the latter clause. Common systems of "coordinates" say nothing about the status of claims to incomparability when the ways in which something might come to be a point in such a system can essentially be disguised. The way in which I "come to the conclusion" of a belief, not in the sense of a reasoned path but the sense involved in the brain's activity and processing leading to such a belief, will be the result of a complex combination of information, the components in which are not obvious from the final resulting belief. Now, this does not imply incomparability on the level of stories told as to how two people arrive at the same belief; our language may well enforce a particular way of telling stories and describing the telling of stories. However, in an examination of what *caused* the two beliefs, the incomparability is mandatory at the point where the beliefs are manifest, given the way in which outputs are constructed according to the geometrical picture of brain processing advocated. This is again, an expression of the maxim that the way things seem need not be of any use in determining the ways things are. We can have a common basis, in the invariants in the environment, but construct our beliefs founded on this basis in ways that are inscrutable to comparisons. This is not to say that we do not construct in the same way. It is likely, being of the same species, that we do, so even the *causes* are indeed the same. However, this

[16](Davidson 1985) p. 184

is not enough. The manner in which the causes combine to manifest in beliefs is sufficient to block any comparability. This demonstrates clearly the first reason why a stratified account of cognition might fail: because it cannot deal with the interface between the supposedly independent levels of explanation.

Now taking Fodor's problem; his is not so much the *connection* of the levels in his explanation but rather that which he takes as the *basis of evidence*. This is thought to be the patterns which are manifest in our language and thought. The systematicity requirement is exactly this – that such patterns are important to any attempted explanation and thus are a mandatory feature of such. This, as we have seen is a mistake as it tries to perform *a priori* what must be undertaken *a posteriori* by research into the contributing factors to the manifest patterns. So, there are two main problems with stratified views in Cognitive Science, the relations between the strata and the basis for postulating the strata. This merely reinforces the pragmatism crushed by the advance of formal AI and cognitive psychology; our theoretical models are basically ways of arranging what we take to be data. We can be wrong in two ways; in our arrangement of the data (McDowell) or what we take to be data (Fodor et al). I explain this further as it is the essential distilled conclusion of this entire work.

There is an assumption built into most theories of semantics, language and thought. There are two components involved in this assumption; firstly that there are *parts* involved in creating our manifest outputs and there are also *ways of combining parts*. This is indeed what is suggested by Quine and is necessary if one is to be an empiricist of any minimal sort. This is indeed also Davidson's "third dogma"; it is a different expression of the scheme and content dualism and, in removing it, one does indeed remove the last prop of empiricism. However, I think that this particular assumption of modern semantics shared even by its sometimes stalwart opponent Quine, can withstand Davidson's attack. He attempts to deny the

idea of a conceptual scheme, of a *way* of combining parts as opposed to the parts themselves, by maintaining that the idea cannot be made sense of, even with the idea of a "theory-neutral reality" to help[17]. Davidson's error, quite surprisingly mirrors one of the central misunderstandings involved in formal semantics. He supposes that a certain conceptual scheme, a way of combining parts, is intimately related to a certain language.

> "We may accept the doctrine that associates having a language with having a conceptual scheme. The relation may be supposed to be this: where conceptual schemes differ, so do languages."[18]

Indeed, the latter might be the case but that certainly does not mean that one can justify one's conclusions in a picture that talks about conceptual schemes by talking about language. This, Davidson does as his arguments against the notion of a scheme are based on an analysis of translation problems. The ubiquity of informational dependencies enshrined in the mediation of the brain imply that languages and conceptual schemes may well differ together, so to speak, but the reason is because they both depend on the same fundamental informational dependencies underlying our whole cognition. Different conceptual schemes would, therefore mean the necessity of intertranslatable languages. Davidson goes on to attack this latter point as he thinks the two components, language and conceptual structure, are equivalent. However, they are not. Conceptual schemes are complex ways of combining the parts that, in the Quinean idiom, impinge on us. These parts, thinks Davidson, are thus independent and allow for a common basis, contradicting the requirement of non-translatability and thus contradicting the idea of a conceptual scheme. The crucial step in the argument is the notion of a "common basis" which does not allow conceptual congruity, contrary to what Davidson thinks, if the manner in which the

[17](Davidson 1985) p. 195
[18](Davidson 1985) p. 184

basis is processed to give the manifest features of language and thought is necessarily inscrutable to language. This is exactly that which I have argued. Manifest language is not a good guide to the "common basis" and thus is a bad thing to rely on in attempts to prove anything about such a basis. So, conceptual schemes stand, regardless of the status of language and thus empiricism stands also. This is in agreement with the assumptions of a tensorial picture of brain processing as this requires a division between invariants and the geometry governing their interrelations. We are indeed now suggesting a "real"/"ideal" distinction but it is important to realise that the "real" here is basically a very different thing to the rich ontologies of the näive realist. So, after all of this, we are bound to say what the implications for the future of Cognitive Science are. This is dependent on what we take to be our data and thus what we take as our ontology. Cognitive Science requires a stand on realism. Without it, we have no idea what we are basing our theories on, we have no idea exactly what there is to cognise about and thus we cannot design models to relate cognition to anything. Again, to overcome the two fundamental problems of stratified reification, we need to know what there is and how this relates to humans. The latter obviously depends on the former.

As promised at the end of Chapter 7, a note is in order regarding the possibility of adding extra levels to a stratified account in an attempt to fix the problem of interrelation of levels. The thought seems to be that one might render different levels of explanation more plausibly connected if there is an intermediate level connecting them. The "distance" from one to the other is thus a combination of two, more easily connected distances. Jackendoff's strategy is like this, as is the work of many of those who desire a "cognitive" element such as Gärdenfors and Lakoff. This cognitive level of representation is meant to provide a sort of buffer between the world and the brain which can explain the relation of the two without having to worry about how to reduce one to the other or interrelate very different

vocabularies. A reasonable idea but one which falls foul of the consideration that any "level of representation" meant to play a necessary role in our cognition cannot employ concepts that are both manifest in our output *and* thought to be causal. This should be obvious from the preceding argument. Now, since a cognitive level of representation is thought to be a more careful examination of exactly which output patterns we manifest are of relevance, such theories are in no better position that full-blooded formalist systems. Maybe they do not take the world for granted but they still take manifest patterns as evidence which is a more fundamental mistake and which subsumes the former.

8.3 Realism

Realism is often feared because of that which it gives us for free. We are given rich sets of existing things and the ontologies are often very liberal. Post-Kant, we are concerned about the sort of things that might populate our ontology and this leads some to argue that the ground of empiricism, which seems to require some sort of basic ontology, is missing. This is, I think often based on a rather limited notion of the possibilities of empiricism. It is generally thought that the main meat of a theory of mind and language must rest in the basic parts that we are able to sense and thus build into our more formidable "mental" structures and phenomena. Little attention is given to the *way* in which we might combine, process and interrelate the basics of out inputs in order to achieve the ordered output we see. Thus, for example, does Davidson deride the idea of the "scheme" involved in this task. We might, for example balance the load carried by the actual basic ontology of a theory with that carried by the manner in which the basics are combined in order to produce the output we perceive. The emphasis is different in such a theory but it is still empiricism, Davidson's third dogma and all. The usual realist tactic is too simple, driven by

the consideration that it is easier and more intuitive to put the structure into the world as it can be more easily mapped onto the structure of our manifest outputs; we feel more comfortable in understanding the relation between input and output. However, as we have seen, this will not do. Firstly because the outputs do not wear their causes on their sleeves and secondly because putting structure into the world ignores the necessary Kantian lesson. The empirical brain sciences are instructive in demonstrating that the manner of processing in the brain can provide us with a quite startling insight into the possibilities of combining information to produce manifestations completely unexpected and unintuitive compared with our usual manner of constructing explanations within formal systems. Since we know we have brains in our heads but we do not know we have formal systems in there, any conflict should probably err on the side of the neuroscience.

There is something which I will call the "principle of semantic indistinguishability" (POSI) which makes ontology in the neuroscientific age a thorny area. An account of this will serve also as a summary of the model I propose for viewing semantics within Cognitive Science. If we agree that, as I have repeatedly urged, our inputs are opaque to the patterns in our outputs, then we have no reason to suppose that our language and thought give any guide at all to ontology. If it so happened that it did, we would have to suppose that the brain does little to information reaching it through perceptual receptors. This is simply not the case. So, what does the brain take as input? Quine and Davidson have argued at length, as have many, about that which we should take as the basis of a theory of meaning. Davidson prefers a distal theory which starts with the causes of intersubjective perception. Quine prefers a more proximal theory where we start with an impingement on our receptive surfaces[19]. This is just a question regarding

[19]Although, later, Quine has moved towards a more distal theory, he remains convinced that it must be as proximal as possible. This is not really an issue relevant to our discus-

one of the fundamental points of trouble in semantic theory I have identified: the basis of the levels in an explanation. If we take seriously the idea that the brain is an information processor, we must realise that it can be of no consequence to processing strategy whether input information has a certain "content" or not. The brain cannot determine whether certain impressions come from a dog, a bus or the feeling of the presence of abstract beauty – information is not like that. "Content" is something we abstract out from our *outputs*; our talk, our beliefs and the like. From the point of view of inputs, information is all the same. In terms of meaning, all inputs are indistinguishable. This is the principle referred to above. Now, if our ontology divided things into the "real" things: dogs, cats etc. and, say, the "abstract" things: truth, justice etc., then we might be fooled into thinking that the brain might, through evolution become sensitive to the difference and thus we might employ a formal theory with this difference built-in. However, the clues to our ontology are from our output only and we know that our perceptions are "output" too; the output of active processing in the brain and perceptual systems. This is an echo of the empiricist doctrine stressed by Quine that our evidence for ontology is exhausted by our sensory receptions while our sensory receptions are not "simple" and do not "give" us anything untainted.

> ". . . to represent cognition as a discernment of regularities in an unadulterated stream of experience. Better to conceive of the stream itself as polluted, at each succeeding point of it's course, by every prior cognition."[20]

This means that there is no way from an examination of the "stream of experience" to determine what the basis of a theory of meaning is. This is exactly why Quine himself took the naturalist turn and asked:

> "But why all this creative reconstruction, all this make-believe?

---

sion but the factors involved can be found in (Quine 1990).

[20](Quine 1953)

> ... Why not just see how this construction [of knowledge] really proceeds? Why not settle for psychology?"[21]

Quine's idea was to look at the way we *really do* construct our manifest patterns of language and thought since this is the only possibility in the situation we find ourselves in. I would argue that it is not quite naturalistic enough to "settle for psychology" these days since modern cognitive psychology is so dependent on models using formal systems that it becomes subject to the criticisms of the usefulness of such systems as have already been made above. At the time Quine was writing, this was not as prevalent: formal linguistics and psychology have grown since, mainly as a result of Chomsky. We should now "settle for neuroscience". If we do this, we see that we must obey the dictates of the Principle of Semantic Indistinguishability. Information is not differentiated by its content as "content" is a notion only applicable at the level of our manifest outputs. Information is not differentiated at all. Thus it is meaningless to talk of different sorts of inputs and all we are left with as an ontology is the blanket term "information". The brain takes this information and extracts certain aspects, noting the invariant features. Invariant features will be highly abstract relations between information and will not correspond to words or concepts as these sorts of things are manifest patterns in our *output*. This is the foundation of the mistake in most putative explanations of semantics and that which Quine sees in *a priori* epistemology; that the elements involved in the basis of our perceptions – the very basis of empiricism – need not correspond to any concepts or words that manifest as a result. This is why I question "settling for psychology" since the dominant cognitive paradigm today tends to classify supposedly "basic" things, as does Lakoff, in terms of concepts only applicable at a level these basics are meant to contribute towards.

The invariant features that the human brain picks out in the envi-

[21](Quine 1969)

ronment are then processed through networks embodying very different geometries. They might be motor-spaces as described by the work of Pellionisz and Llinas, or they might be "semantic spaces" roughly described by, say, the Churchlands. The invariant relations will however, on a tensorial model, be preserved through changes in representational space. Our output consists in combining the abstract information involved in a manner such that our outputs can be constructed, regardless of the complexity of the spaces involved. This, as mentioned, must necessarily be a contravariant combination. Once the output is manifest however, there is no way in knowing what the components were unless one knows the way in which the information was combined. This is, in my view, not an issue for Cognitive Science and semantics but one for neuroscience. Thus, this renders the possibility of an explanation of the workings of human semantics and language impossible from within an *a priori* discipline or indeed, since Chomsky has taken misguided pains to extend the notion of evidence to try to circumvent the particular problem of being seen as *a priori* (see above), a discipline not concerned with the actual neural embodiment of processes. The POSI means that semantic "content" cannot play a role in any of this and thus it is causally impotent. This means a collapse of stratified views dependent on levels of explanation and representation with *sui generis* characteristics. There is no independent "Realm of Reasons" or level of compositional propositional content simply because there is no way that this could be causally efficacious without being part of our input and *this* is not possible since these supposedly higher levels of explanation are completely dependent on concepts derived from our *output*. If one's theory is based on phenomena manifest in the output of a system and one does no empirical work in addressing *how* the outputs are generated, then there is no guarantee – in fact quite the opposite in the present case due to the technical consideration of abstract geometry – that a resultant theory has anything to do with the causal factors playing a role in what one is attempt-

ing to explain. Put another way, if explanation requires us to understand causes, then a formalist picture requires that we know how things at their advocated *sui generis* level of explanation come to have their causal powers. It is maintained, for example, that believing propositions causes us to act. We cannot account for this suggested causal power by associating such features with causally efficacious elements in the Realm of Law since this would mean the brain does nothing between input and output; if properties are passed through untouched from input to output, then we contradict the obviously true statement that the brain *does something* to inputs to manifest our outputs. This much Kant made us suspicious of and now, through modern research in neuroscience and biology, we are certain of it.

So, what of realism? It is the post-Kantian dream that we might factor out the contribution of our categories in order to perceive the noumenal world. This is obviously impossible in the sense in which Kant expressed the problem since the very conditions of perception would thereby be removed and thus to talk of "perceiving the noumenal" is nonsense. With this idealism being couched in the modern neuroscientific idiom, things may be different however. If the patterns imposed on our outputs are the result of specifiable networks embodied in the brain, then the possibility arises that we might isolate the contribution of certain sorts of regularities, as in the case of the example of lateral inhibition mentioned in Chapter 6. If it turned out, as a matter of evolutionary contingency, that the operation of different networks were isolatable, then we might be able to experimentally remove *some* of the "conditions of sensible appearances" without affecting others. This might allow us to get an idea of what happens to our inputs on their way to becoming outputs. Neuroscience is full of such examples already but it is still not clear exactly what overall patterns are emerging (see (Milner & Rugg 1992) for an overview and many examples). It may transpire that the operation of different embodied networks in the brain are not cleanly separable and thus the case is as Kant describes it;

the contribution of our categories of perception operates either all at once or not at all. If the former, we cannot assume anything from our outputs. If the latter, we cannot be really said to be "perceiving" anything. This is the essence of the position held by what Churchland calls the "boggled skeptics"[22] who believe that the brain does indeed contain all of the clues necessary for a correct view of ontology and semantics but who think that the brain is too complex for brains to be able to fathom. Thus, there would be a definite answer to the question of whether the brain modifies our sensory input in a particular way but the complexity of this is so great that we could never determine it. This is however a matter for the history of neuroscience to answer at a later date and not something Cognitive Science can attack in an *a priori* manner in its work in natural language semantics or linguistics.

8.4 The POSI and its Implications

So, my views regarding the status of the different supposed levels of explanation in Cognitive Science can be summarised in the Principle of Semantic Indistinguishability:

> The contribution of the environment to our cognition is epistemologically opaque. This is empirically supported. Given this, the output of our cognitive apparatus can be as much a part of the environment as that which we are more used to calling "real". Therefore, "content" is a concept only applicable to patterns of our cognitive output and is therefore useless in an investigation of the causal factors contributing to cognition. *a fortiori*, it is useless to posit a level of explanation characterised by the notion of "semantic content". This, however, is common to all of the formal, propositional based accounts commonly proposed in Cognitive Science.

[22](Churchland 1986*b*) p. 315

The question is: how does this principle manifest itself in the brain? The onus is on me to present some more concrete picture of how this works in terms of the neuroscience discussed. One thing must be clear firstly. I am not obliged to give semantic accounts of the usual models, fragments or examples since these are, I have argued, misplaced not in their execution but rather their whole conception. They are attempts to learn about causes from undecomposable effects and this is therefore an impossible task. My task is rather to describe the way in which we contribute to the environment: the way in which the output of our cognition - the categorical judgements - is constantly absorbed into the fabric of "reality" such that it becomes an indistinguishable part of our inputs.

One hears about certain experiences that have changed people's lives forever. A certain experience changes, as it were, the parameters by which the world is measured – it changed the criteria for future cognition and perception. I think that these experiences occur daily and are the reason why out cognitive life is so inscrutable. There are, I think, two forms of this experience. The first is of a major experience that changes one's life. This is like a large change of view, occurring over a short space of time and is a relatively rare event. The other type of this event is due to ubiquity. The constant occurrence of a perception will, over time, tend to alter one's view, particularly if this perception is so ubiquitous as to appear not worthy of conscious attention. [23] This sort of constant, repetitive and small-scale environmental invariant has the same effect, at a much slower speed, as the what we might term "Eureka!" experiences mentioned earlier. However, the effect is the same: the criteria for our cognition and perception are altered. In a sense, this effect is simply an aspect of our ability to learn from the environment. Let us digress a moment into this area in order to make

[23]This is, I think, exactly how advertising works. Billboards, television advertisements, radio advertising etc. are so much part of the furniture of life that they melt into the background *because* of their ubiquity and are no less effective for that.

some observations that bear directly on the POSI.

Humans are able to learn indirectly. That is, they can learn from books, from videos, from tapes or from being told what to do. They can even learn by thinking about how to do something. This is, I think , related to the POSI. If POSI were not true, this adaptability in learning would be impossible since the source of information would matter far more. The reason that we can learn from sources so different from direct experience is because, I argue, the POSI ensures that, as far as much of cognition goes, books, videos etc. *are* direct experience. Epistemologically, we fool ourselves into an extended notion of reality. I say "as far as *much* cognition goes" as we famously cannot learn physical skills very well from books etc. Well, like most evolutionary effects, POSI is not designed to be all-encompassing and perfect. Indeed, it is not designed at all and simply serves as the most useful cognitive architecture so far. Indeed, this sort of integrated learning ability means we often try to learn things in only one way e.g. from books, which, as every carpenter or computer consultant or will tell you is nonsense.

So, in terms of the abstract geometrical picture I have urged, what is responsible for the POSI? That is, what elements of this theory are responsible for the constant embedding of our continuing cognition in the environment in such in inscrutable way? The issue is one of malleability: the axes of any abstract space that activates upon stimulation from the environment must not be rigid. They must be able to change in response to experience and thus to change the significance of similar experience in the future. Thus, we must not see the axes of the spaces as being so different from the tensors that they define. If we allow such malleability in the very structure of our cognition then there is the following consequence: the reason our cognition is so stable under normal conditions is *because the conditions are normal.* We are capable of quite radical conceptual changes and experiences but we seldom have such not because the brain is so sta-

ble but because our experiences are usually so narrow and repetitive. Thus, our conception of reality is such due to the intimate relationship between our cognitive structure and "the world". This is the bottom line: our experience determines how we take our future experience by altering the structure which makes sense of our experience. In terms of a tensorial theory, the metric tensor for the particular space would need to be modified by every impression coming from what we take epistemologically as "empirical reality"'. This could be accomplished by a constant tensor product of impression "inputs" with the tensor representing the metric for the space. (By convention, matrices are double overlined, vectors single overlined).

$$\overline{\overline{g}}^{ij} \cdot \overline{v}_j$$

Metrics are established by relations of covariant and contravariant expressions of the same physical vector (see, for example, (Wrede 1972) p. 82). That is, by a relation of $\overline{v}^j$ to $\overline{v}_j$. This is very important as it corroborates the idea that on this model, it is natural to expect that the metric is dynamic since part of my whole premise is that cognitive inputs are continuously effected by previous cognitive outputs and by the state of the metric. It is the former effect that accounts for notions of "content" and the latter that accounts for the perceptually and cognitively prior categories that shape our thought. Normally, the metric is seen as a stable part of a system, relating different vectorial expressions of invariants. However, I am urging that the metric is constantly changing in response to the activation that passes through it. That is, given that we determine the metric tensor by relating different expressions of the same invariant, we must accept that what is taken as an "invariant" is determined by the metric itself. This is the most important part of the geometrical picture I advocate and thus I shall state it explicitly.

We should be careful not to force prosaic interpretations onto the notion of a metric. A metric is often said to define the notion of "distance" in a geometrical space. Now, we are used to this corresponding to our usual notion of "distance" for Euclidean spaces but this is simply a special case where, for one, the axes are orthogonal. In more complex spaces, "distance" can be a highly abstract property indicating some relationship between elements in the space. Also, different metrics give different ways of measuring it. Our usual notion is something like "the shortest route between two things". Well, it depends what is meant by "shortest". There are, for example, "city-block metrics" that define distance in terms of routes taken only on combinations of paths parallel to the axes of the space. So, the notion of metric is a very general notion and this should not be forgotten. The usual mathematical treatment of tensor relationships supposes that we have an invariant, physical state of affairs that can be represented by different vectors. That is, *the same* event can be multiply represented in different spaces of different dimensions etc. This much is obvious as far as it goes since, for example, a ball being caught is represented in motor space and visual space on quite different criteria. The whole of the tensorial picture of cerebellar activity presented so far starts from this. However, in the case of human cognition, we must accept that the notion of "the same" is unavoidably one that is determined by our current conceptual apparatus. On this picture, crucial to an understanding of this apparatus is the notion of the metrics which govern their activity. Therefore the metrics are responsible for the generation of the metrics. The dynamics of the system becomes apparent when we consider the embodiment of a tensorial picture in human cognition where our outputs are opaquely taken as inputs too. This is exactly the POSI.

A metric tensor is required in order for the notion of "distance" to have any meaning. Einstein famously made very cryptic remarks about the metric underlying everything in his model because "distance" between things,

however abstract, is such an essential part of our way of understanding things, any picture of a geometrical nature therefore must note the importance of the metric. The geometrical interpretation of tensor analysis has it that a tensor with contravariant indices is represented in the usual way by the parallelogram rule we are familiar with from vectors. It is easy to associate a notion of "distance" or "magnitude" with vectors when they are represented this way. However, a tensor with covariant indices in an n dimensional space is usually represented by an $n-1$ dimensional plane and thus there is no way to define "magnitude" or "distance" for them. It simply has no meaning and thus the metric, which allows the lowering and raising of indices as noted earlier, essentially allows the reinterpretation of covariants as contravariants. In fact, this was the essence of the Pellionisz and Llinas model discussed in Chapter 5. However, as we note below, since most tensors are "mixed", that is to say they have both covariant and contravariant indices, a metric is essential if "magnitude" is to have a meaning. It is clear that indeed it must if we are to be able to, at least, coordinate our motor responses and so it seems to be required as an integral part of our cognitive makeup.

8.5 Why tensors are the right way to think about cognition

Consider the problems discussed previously with the usual geometrical approaches to cognition. It was argued that the criticisms levelled by opponents such as Fodor and Lepore were justified because the advocates of a geometrical picture had conceded too much to the methodological standards of their opponents. As a result, their models were too simplistic and designed with adherence to established criteria of understandability (which underpin the very theories they seek to displace) in mind. The no-

tion of concepts as points in conceptual spaces is not a happy one as it begs an *a priori* specification of the components of such a point which is exactly the same as a decomposition or meaning postulate. Thus such geometrical theories are no different from traditional approaches. All that has happened is a transliteration of terms, not the genesis of a new theory. Now, if we are to take say, concepts as tensors, we see immediately that this approach is quite different[24]. The components of a tensor are not values along axes but rather are *functions* whose values are the components of vectors. So, a tensor in this respect is a highly abstract entity composed of all of the sets of functions defined on vector components *for all reference frames.* A tensor is a set of cardinality c of sets of functions that describe transformations into c reference frames. Now this brings to mind the notion of metaphor. Metaphors are enabled by concepts playing similar roles in different situations. What we have lacked in investigating metaphors is a suitable notion of "similar". Concepts as tensors gives an obvious benefit in that their nature is of some abstract object that manifests as ways of operating on objects in any given space *in the same way.* Here, we have a notion of "same" we can really flesh out since it is mathematically defined by the invariants of a system. Of course, that is not to say that we can state this in ordinary language for a given concept: to do this would be to go against everything I have argued so far. The complexity of the spaces involved may mean we can never so this for any given example; this is an empirical matter and not for philosophers or Cognitive Scientists to make *a priori* models of. Indeed, this model makes our attempts to "define" concepts much harder since they are much more abstract things than we would, according to traditional methodology, like. What we might hope for is a classification of

[24]We now discuss the tensor concept in full generality whereas before, we have considered it in the same sense as Einstein i.e. tensors as being analogous to vectors that transform according to certain restrictions. This is really a convenience of discussion and I revert to the more modern treatment largely due to Weyl in order to bring out the abstract nature of the concept. The simplification is due to the fact that vectors can be conveniently considered as tensors of rank one. See (Sokolnikoff 1951)

concepts according to their tensorial rank and covariant/contraviant index. That is, their level of abstractness and how they treat their arguments. I venture to speculate on this in the spirit of constructive exposition and not in any manner that should be taken as suggesting an *a priori* model.

Consider the covariant law (this is only given for tensors of rank one e.g. vectors, for clarity):

$$B_i = \frac{\delta x^\alpha}{\delta y^i} A_\alpha$$

This means that the tensor A – which is a set of functions that transform coordinates in one reference frame – is *the same* object represented by the set of functions B in another reference frame and the relation is one governed by the partial differential where the new coordinates y^i are the divisor (hence "covariance"). This is a certain type of transformation which we may well fruitfully correlate with a certain type of conceptual shift: that of a metaphor whose concept relates its arguments in the same way in its metaphorical sense (i.e. the transformation *covaries* the argument roles). "Big hearted" may be said to be of this type since the literal use of "big" covaries intuitively with its sense in the metaphor. "Small minded", "big headed", "hugely interesting", "tiny mind" etc. are all of this type too, from which we might conclude that metaphors and analogies to do with physical size are all governed by concepts that are covariant tensors since they map into their metaphorical uses in an obvious way that covaries with their "literal" usage. Consider the contravariant law:

$$B^i = \frac{\delta y^i}{\delta x^\alpha} A^\alpha$$

This says the same as the covariant law but here, the old coordinates are the divisor in the partial differential that governs the transformation and thus "contravariant". We would expect metaphors and conceptual

shifts based on sarcasm to follow this pattern. For example "He was as brave as a mouse" has an obvious opposite shift of meaning. In our terms, the transformation of "brave" into this usage is contravariant. Now, it is obvious that this is simplistic in the sense that we cannot hope to include all conceptual nuances into these two polar opposites. Tensor theory has the notion of a "mixed" tensor however, where some components transform covariantly and some contravariantly. In fact, the covariant and contravariant laws are just special cases where all components transform the same way. Mixed tensors are usually employed to treat real problems since they allow an arbitrarily complex relation. The transformation law in its full generality for mixed tensors is:

$$B^{j_1\cdots j_s}_{i_1\cdots i_r} = \frac{\delta x^{\alpha_1}}{\delta y^{i_1}}\cdots\frac{\delta x^{\alpha_r}}{\delta y^{i_r}}\cdot\frac{\delta y^{j_1}}{\delta x^{\beta_1}}\cdots\frac{\delta y^{j_s}}{\delta x^{\beta_s}}A^{\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_r}$$

This shows that the transformation law is for tensors of covariant rank r and contravariant rank s. The contravariant and covariant laws expressed earlier fall out as special cases (when expressed in full generality for tensors of any rank) when the contravariant or covariant rank is 0 respectively. Thus, this can be seen as an extremely powerful way of expressing transformations of concepts of any mixture of covariance/contravariance. That is, certain things about a conceptual change might result in some aspects of the concept being covariant and some contravariant. Phrases like "as safe as can be expected" seem to fall under this umbrella as they covary with the literal positive connotations of "safe" but contravary with them due to the qualifier. The notion of rank allows us to have concepts of arbitrary degrees of abstractness. That is, we can have concepts that deal with transformations of concepts etc. This is necessary to be able to account for meta-conceptual schemes. For example, "conceptually shaky" seems to be a covariant concept about concepts. i.e we might say it is a rank two covariant tensor.

I do not want to advocate a detailed catalogue of conceptual types at all. That would according to my criteria set out previously, be little more than an intellectual game. What I have wished to do is to suggest how the notion of a tensor does justice to some of our intuitions. The serious obstacles to normal theorising that I have insisted upon throughout are of paramount importance still.

Another thing of importance that this geometrical picture shows is that the metric - that which our cognition relies upon for its speed and relations of ideas according to this model - is a dynamic and evolving determinant of our minds. Since, on the picture we outlined in Chapter 5, the metric is responsible for creating contravariant execution vectors (behaviour) from covariant input vectors and thus this malleability in the metrics for our cognitive spaces will result in changing behavioural patterns. Naturally, how we see things influences how we act. The strength of the experience would correspond to the quantitative strength of the input tensors and this serves to model the notion of importance of impression. Many small but repeated impressions would alter the metric by successive product. Large impressions would alter it swiftly because their strength - their vectorial magnitude - would cause a larger alteration in the metric. Of course, it is implied by my view that what constituted a "small'" or "large'" impression or input would depend on the state of the cognitive agent. Now, it is natural to question this malleability of our cognitive apparatus when put in such an extreme form. There are surely, it might be said, certain limits to malleability. This is certainly the case and is what, I think, defines important cognitive difference, for example, between species. Kant's program was to identify the fundamental categories that cannot be bypassed by anything we would want to call "cognition". This is indeed a real limit and we must therefore accept a *scale* of categories ranging from the very malleable to the almost mandatory. For example, for the human species;

Hard categories time, space, causality

Firm categories religion , moral belief

Soft categories tastes, manners

As examples, these suggest the types of things that might appear at different places on a continuum of malleable categories. We even might not want to take the notion of "hard-coded" categories such as space and time too literally in the face of, say, drug-related experiences in which time and space are sometimes reported as having changed qualities. What is clear however, is that this sort of malleability is much less common and much more serious in its implications for the whole of cognition. That there is a scale of malleability seems to me to be beyond doubt; that it can be accounted for in terms of an abstract geometrical theory seems to be very important.

8.6 Connections With Meta-Systems Transition Theory

It is necessary here to consider a theory developed by Valentin Turchin called "Meta-Systems Transition Theory" (MST)[25]. This notion is similar in spirit to the POSI and it would be well to compare and contrast the two ideas in order to contribute to an understanding of my position. I think MST is an ideal discussion point since it encompasses many modern views which, whilst not really falling into the traditional symbolic formalist camp, still holds to some notion of nomologically imperative "levels of explanation". Thus, this discussion is crucial in setting my ideas off against the central problems that I have mentioned in Cognitive Science. MST is the idea that, as a naturally occurring phenomenon, there is a process whereby an organisation of multiple subsystems occurs, leading to a con-

[25](Turchin 1977)

trolling subsystem taking over and creating more order. The controlling system is then a meta-system and the move from the initial state to the next, more organised state is the transition. Turchin[26] uses this notion to account for the whole of evolution in that simple systems are integrated by the emergence of controlling systems which further enable the multiplication of the subsystems since their very organisation makes their multiplication now more useful to the organism. An example is the appearance of movement:

> "The first metasystem transition we discern in the history of animals is the appearance of movement. The integrated subsystems are the parts of the cell that ensure metabolism and reproduction. The position of these parts in space is random and uncontrolled until, at a certain time, there appear organs that connect separate parts of the cell and put them into motion: cell membranes, cilia, flagella. A metasystem transition occurs which may be defined by the formula: control of position = movement".[27]

Turchin builds up the history of evolution as a history of such transitions, even going so far as to include language:

> "Language emerges when the phenomena of reality are encoded in linguistic objects. But after its origin language itself becomes a phenomenon of reality. Linguistic objects become very important elements of social activity and are included in human life like tools and household accessories. And just as the human being creates new tools for the manufacture and refinement of other tools so he creates new linguistic objects to describe the reality which already contains linguistic objects. A metasystem transition within the system of language occurs. Because the new linguistic objects are in their turn elements of reality and may become objects of encoding, the metasytem transition may be repeated an unlimited number of times. Like other cybernetic systems we have considered in this book, language, is a part of the developing universe and is developing itself. And like other systems, language–and together with it thinking–is undergoing

[26]Ibid Chap. 3
[27]Ibid, Chap 3

> qualitative changes through metasystem transitions of varying scale, that is to say, transitions which encompass more or less important subsystems of the language system."[28]

It will be apparent that this bears similarities to the POSI in that it allows the outputs of our perceptions of the world to be, literally, a part of the world. However, there is an assumption built into this theory that does not carry over well into the cognitive domain. This is the idea that such a transition is a unique event that monotonically establishes a system as *the* controlling system for some set of subsystems. If this were the case, then the structure of cognition would be a static tree of systems, related in established ways. It is clear from what has come before that this is not an adequate picture and that we must take seriously the dynamic aspect of cognition. Thus, in terms of the notion of a metric defined on the abstract spaces of our cognition, we said that this must change constantly in order to account for the way in which our cognition changes and alters the world. In terms of MST then, we must allow that the transitions are constantly happening and that they may reverse themselves and instantiate completely different control structures for the same subsystems. Meta-System Transitions may also occur where elements of a controlled subsystem might become the controlling system. For example, one day we might believe that marketing is simply pandering to people's already existing tastes, the next day we might realise that the tastes are often caused by the marketing. The reversal of causation in beliefs is a good example in general of a transition where an organising principle is actually underdetermined and so can sway with different evidence. In fact, it is the crux of the matter that MST theory underestimates the amount of underdetermination for any particular belief. Given this, there will be many transitions, often cancelling each other out. In plain terms, this is simply the fact that our cognition is controlled by different principles at different times. Some-

[28]Ibid, Chap. 7

times one idea governs our thought, sometimes another and this tends to change rapidly as is obvious to all in cases where we are preoccupied with a thought that colours the rest of our thoughts. This is very volatile and the governing thought passes away as the meta-systemic governor, changing places often to be governed itself. It may well be useful to describe evolution as a certain static path of Meta-Systems Transitions and in one sense, since it actually took only one path, it is. This is not adequate as a model for ongoing cognition though as the very point with this is that, every moment, it takes different paths, backtracks, changes direction and often performs radical u-turns.

More recent proponents of MST are coming to this sort of view I feel. For example, C. Henry's paper at a recent MST conference[29] says:

> "Biological structures determine cognitive strategies. Meaning, truth, and observational accuracy are embodied in a complex branching system that performs, in part, through deconstructing objects and phenomena into particular characteristics and then reassembles these parts into an often different whole."

Strikingly, this echoes the tensorial model I have explored in its insistence of deconstruction and reconstruction of the same data. This is clearly a more dynamic approach to MST than Turchin originally expressed.

So, to summarise. Formal systems based on symbolic, logical notions are like sailor's classification of patterns of waves. This is a classification that helps them do things. It does not explain where the patterns come from. Sailors may sensitively correlate wave patterns with wind and temperature. This still does not explain how the patterns arise – it is pragmatic knowledge. Pragmatic knowledge may become filigreed with spare time left over from surviving and may appear to be too complex and clever to be mere pragmatic knowledge. But filigreed pragmatic knowledge is still, at heart, pragmatic and is for doing things. In the case of filigreed pragmatic

[29](Heylighen & Aerts 1996)

knowledge, it is simply no longer clear anymore what it is that it is for doing. The real explanations of things seldom enter into our ways of doing things. The real explanations are too detailed and complex. Hence we have what Churchland calls "folk Psychology" which cannot, contrary to what Churchland thinks[30], be replaced by a more rigorous physicalist vocabulary since folk language is essentially a pragmatic language that allows us to do things. There is simply no reason why the language of explaining things is of any use in doing things. In fact, an implication of my position is that the language of explanation cannot be a language of action, of doing, of pragmatics. This is because it cannot even be a language of static models and of formalisms whose central thrust is the desire to render an explanation as canonically "understandable" by some, unfortunately explanatorily irrelevant, historical criteria. We must simply take this strong medicine and accept that knowing how something works does not help you do it better and conversely, knowing how to do something does not mean you know how you do it. The mistake that has been made is that this state of affairs has led to theorists thinking that there are different "levels" of explanation: those that correspond more closely to what we do and those that are closer to how we do it. There is only one "explanation" and that is the explanation of how we do something. How we talk, think and theorise about what we do may bear no relation to this and need not effect us once we realise the incredible complexity of the transitions of information that take place in perception and cognition. Patterns in behaviour may come from two sources: the world and the machinery that deals with it. If we cannot tell these two apart in experience, as the POSI states, then the patterns cannot be used as evidence for theories that take patterns in a realist manner. The upshot of the POSI is that there is an epistemological barrier to model-building in Cognitive Science. This barrier, as Kant saw, is uncrossable. I have argued that this has important implications for methodology

[30]See (Churchland 1984)

in Cognitive Science.

Bibliography

Barwise, J. (1989), *The Situation in Logic I*, CSLI.

Barwise, J. & Perry, J. (1983), *Situations and Attitudes*, Bradford Books Series, MIT Press.

Barwise, J. & Seligman, J. (1992), 'The rights and wrongs of natural regularity', *Philosophical Perspectives*.

Barwise, J. & Seligman, J. (1993), 'Channel theory: Towards a mathematics of imperfect information flow'. Manuscript.

Churchland, P. M. (1984), *Matter and Consciousness*, MIT Press.

Churchland, P. M. (1986*a*), 'Some reductive strategies in cognitive neurobiology', *Mind* **XCV**(379), 279–309.

Churchland, P. S. (1986*b*), *Neurophilosophy*, MIT Press.

Cooper, R. (1988), Facts in situation theory: Representation, psychology, or reality?, *in* R. M. Kempson, ed., 'Mental Representations', Cambridge University Press, chapter 2, pp. 49–61.

Cooper, R. (1991), 'A working person's guide to situation theory', HCRC Publications: Research Paper HCRC/RP-24.

Crangle, C. & Suppes, P. (1989), Geometrical semantics for spatial presuppositions, *in* 'Midwest Studies in Philosophy', Vol. XIV, University of Notre Dame Press, pp. 399–422.

Davidson, D. (1967), 'Truth and meaning', *Synthese*.

Davidson, D. (1985), On the very idea of a conceptual scheme, *in* 'Inquiries Into Truth and Interpretation', Blackwell.

Davidson, D. (1990), Meaning, truth and evidence, *in* R. Barrett & R. Gibson, eds, 'Perspectives on Quine', Blackwell, pp. 68–79.

Dretske, F. (1981), *Knowledge and the Flow of Information*, MIT Press.

Dreyfus, H. L. (1992), *What Computers Still Can't Do*, MIT Press. Revised edition of "What Comupters Can't Do" (1972).

Dreyfus, H. L. & Dreyfus, S. E. (1988), 'Making a mind versus modelling the brain: Artificial intelligence back at a branch-point', *Artificial Intelligence*.

Fodor, J. & Lepore, E. (1992), *Holism: A Shopper's Guide*, Blackwell.

Fodor, J. A. (1987), Why there still has to be a language of thought, *in* 'Psychosemantics', MIT Press, pp. 135–167.

Fodor, J., Fodor, J. & Garrett, M. (1975), 'The psychological unreality of semantic representations', *Linguistic Enquiry* **VI**(4), 515–531.

Fodor, J., Garrett, M., Walker, E. & Parkes, C. (1980), 'Against definitions', *Cognition* **8**, 263–367.

Frege, G. (1882), 'Über den wissenschaftliche berechtigung einer begriffsschrift (on the scientific justification of a conceptual notation)', *Zeitschrift für Philosophie und philosophische Kritik*. English translation by Bartlett in *Mind* 73, 1964.

Frege, G. (1892), 'Über sinn und bedeutung (on sense and reference)', *Zeitschrift für Philosophie und philosophische Kritik*.

Frege, G. (1918), 'The thought: A logical enquiry'. Translation: Quinton and Quinton in *Mind* 65, 1956.

Gamut, L. (1991), *Intensional Logic and Logical Grammar*, Logic, Language and Meaning, University of Chicago Press.

Gärdenfors, P. (1990), 'Induction, conceptual spaces and AI', *Philosophy of Science* **57**, 78–95.

Gärdenfors, P. (1993*a*), Conceptual spaces as a basis for cognitive semantics, Lund internal paper.

Gärdenfors, P. (1993*b*), 'The emergence of meaning', *Linguistics and Philosophy* **16**, 285–309.

Gregory, R. L. (1966), *Eye and Brain: The Psychology of Seeing*, 4th (1990) edn, Oxford University Press.

Harnad, S. (1992), Connecting object to symbol in modeling cognition, *in* A. Clarke & R. Lutz, eds, 'Connectionism in Context', Springer Verlag.

Harnad, S. (1993), 'Grounding symbols in the analog world with neural nets', *Think*. Special Issue on Machine Learning (in press).

Hesse, M. (1970), Is there an independant observation language?, *in* R. Colodny, ed., 'The Nature and Function of Scientific Theories', University of Pittsburgh Press, pp. 36–77.

Heylighen, F. & Aerts, D., eds (1996), *The Evolution of Complexity*, Vol. 8, Principia Cybernetica Project. Symposium in "Einstein Meets Magritte" conference.

Hospers, J. (1956), *An Introduction to Philosophical Analysis*, Routledge.

Jackendoff, R. (1983), *Semantics and Cognition*, Current Studies in Linguistics Series, MIT Press.

Jackendoff, R. (1985), 'Information is in the mind of the beholder', *Linguistics and Philosophy* **8**(1), 23–34.

Jackendoff, R. (1988), Conceptual semantics, *in* U. Eco, M. Santambrogio & P. Violi, eds, 'Meaning and Mental Representation', Indiana University Press. Advances in Semiotics Series.

Janssen, T. (1983), *Foundations and Applications of Montague Grammar*, Mathematisch Centrum, Amsterdam.

Kamp, H. & Reyle, U. (1993), *From Discourse to Logic*, Vol. 42 of *Studies in Linguistics and Philosophy*, Kluwer Academic Press.

Kant, I. (1787), *Kritik der reinen Vernunft*, Macmillan.

Katz, J. J. (1972), *Semantic Theory*, Harper.

Katz, J. J. (1990), The refutation of indeterminacy, *in* R. Barrett & R. Gibson, eds, 'Perspectives on Quine', Blackwell, pp. 177–197.

Keenan, E. L. (1976), Toward a universal definition of "subject", *in* 'Subject and Topic', Academic Press.

Keenan, E. L. (1987), Semantic case theory, *in* J. Groenendijk, M. Stokhof & F. Veltman, eds, 'Proceedings of the 6th Amsterdam colloquium'.

Keenan, E. L. (1996), Logical objects, UCLA working paper.

Keenan, E. L. & Faltz, L. M. (1978), Logical types for natural languages, Technical report, University of California at Los Angeles.

Kime, P. L. (1996), Reinventing the square wheel: the nature of the crisis in cognitive science, *in* S. O'Nuallain, P. McKevitt & E. M. Aogain, eds, 'Two Sciences of Mind: Readings in Cognitive Science and Consciousness', John Benjamins, pp. 5–17.

Kron, G. (1939), *Tensor Analysis of Networks*, Wiley.

Kuhn, T. S. (1963), The function of dogma in scientific research, *in* A. Crombie, ed., 'Scientific Change', Basic Books, chapter 11.

Lakoff, G. (1972), Linguistics and natural logic, *in* Harman & Davidson, eds, 'Semantics of Natural Language', Dordrecht-Holland, pp. 545–665.

Lakoff, G. (1987), *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*, University of Chicago Press.

Lakoff, G. (1988), A suggestion for a linguistics with connectionist foundations, *in* H. Touretzy & Sejnowski, eds, 'Proceedings of the 1988 Connectionist Models Summer School, CMU', Morgan Kaufmann, pp. 301–314.

Lappin, S. (1997), 'Semantic types for natural languages'. Inaugural Address.

Lewis, D. (1984), 'Putnam's paradox', *Australasian Journal of Philosophy*.

McDowell, J. (1994), *Mind and World*, Harvard University Press.

Milner, A. D. & Rugg, M. D., eds (1992), *The Neuropsychology of Consciousness*, Foundations of Neuropsychology Series, Academic Press.

Nuallain, S. O. (1995), *The Search for Mind*, Ablex Publishing Corporation.

Pellionisz, A. (1983), 'Brain theory: Connecting neurobiology to robotics. Tensor analysis: Utilizing intrinsic coordinates to describe, understand and engineer functional geometries of intelligent organisms', *Journal of Theoretical Neurobiology* **2**, 185–211.

Pellionisz, A. (1984), 'Coordinates: a vector-matrix description of transformations of overcomplete CNS coordinates and a tensorial solution using the moore-penrose generalised inverse', *Journal of Theoretical Neurobiology* **110**, 353–375.

Pellionisz, A. & Llinas, R. (1979), 'Brain modeling by tensor network theory and computer simulation. The cerebellum: Distributed processor for predictive coordination', *Neuroscience* **4**, 323–348.

Pellionisz, A. & Llinas, R. (1980), 'Tensorial approach to the geometry of brain function: Cerebellar coordination *via* a metric tensor', *Neuroscience* **5**, 1125–1136.

Pellionisz, A. & Llinas, R. (1982), 'Space-time representation in the brain. The cerebellum as a predictive space-time metric tensor', *Neuroscience* **7**, 2949–2970.

Pellionisz, A. & Llinas, R. (1985), 'Tensor network theory of the metaorganization of functional geometries in the central nervous system', *Neuroscience* **16**(2), 245–273.

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

Pylyshyn, Z. W. (1984), *Computation and Cognition*, MIT Press.

Quine, W. (1953), On mental entities, *in* 'Contributions to the Analysis and Synthesis of Knowledge', MIT Press.

Quine, W. (1960), *Word and Object*, MIT Press.

Quine, W. (1969), Epistemology naturalised, *in* 'Ontological Relativity and Other Essays', Columbia University Press.

Quine, W. (1972), Methodological reflections on current linguistic theory, *in* Harman & Davidson, eds, 'Semantics of Natural Language', D. Reidel, pp. 442–454.

Quine, W. (1990), *Pusuit of Truth*, Harvard University Press.

Shepard, R. N. (1980), 'Multidimensional scaling, tree-fitting and clustering', *Science*.

Smolensky, P. (1988), 'On the proper treatment of connectionism', *Behavioural and Brain Science* **11**, 1–74.

Sokolnikoff, I. S. (1951), *Tensor Analysis: Theory and Applications*, Applied Mathematics Series, Wiley.

Tarski, A. (1931), 'The concept of truth in formalised languages'. In (Tarski, 1956).

Tarski, A. (1944), 'The semantic conception of truth', *Philosophy and Phenomenological Research*.

Tarski, A. (1956), *Logic, Semantics and Metamathematics*, Oxford University Press. Translation by Woodger.

Turchin, V. F. (1977), *The Phenomenon of Science*, Columbia University Press.

van Fraassen, B. C. (1966), PhD Thesis, PhD thesis, University of Pittsburgh.

van Fraassen, B. C. (1967), 'Meaning relations among predicates', *Noûs*.

von Eckardt, B. (1993), *What is Cognitive Science?*, MIT Press.

Winograd, T. (1985), 'Moving the semantic fulcrum', *Linguistics and Philosophy* **8**(1), 91–104.

Wrede, R. C. (1972), *Introduction to Vector and Tensor Analysis*, Dover.