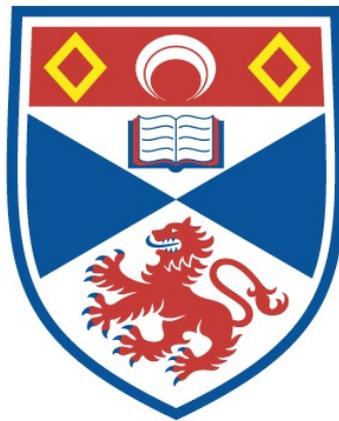


# ANALOGY, RULE-FOLLOWING AND MEANING

Andrew Simeon Williams

A Thesis Submitted for the Degree of MPhil  
at the  
University of St Andrews



2018

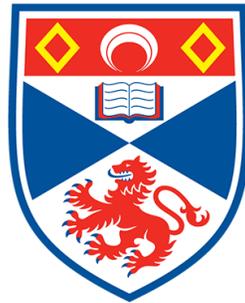
Full metadata for this item is available in  
St Andrews Research Repository  
at:  
<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:  
<http://hdl.handle.net/10023/18627>

This item is protected by original copyright

# Analogy, Rule-Following and Meaning

Andrew Simeon Williams



University of  
St Andrews

This thesis is submitted in partial fulfilment for the degree of  
Master of Philosophy (MPhil)  
at the University of St Andrews

September 2017

## **Abstract**

*In this thesis, I argue that meaning something by a linguistic expression should be thought to consist, not in following a rule, but in drawing an analogy. I argue that using a linguistic expression meaningfully involves analogically extrapolating from our past experience of that expression, by observing a similarity between the present instance (that which the expression is being applied to) and previous instances (our past experiences involving the use of that expression). This is in opposition to the classical account of meaning, according to which meaning something by a linguistic expression involves following a rule – a rule stating necessary and jointly sufficient conditions for the expression to apply. I argue that there is no one thing that all meaningful instances of many linguistic expressions have in common, such that its meaning could be captured in the form of a rule. I claim that the categories corresponding to our linguistic expressions (containing the objects that any given expression is true of), are defined not in terms of necessary and jointly sufficient conditions, but in terms of family resemblance relationships between particular instances, such that analogy plays an essential role in every meaningful application of a linguistic expression. I argue that metaphorical uses of linguistic expressions are clearly dependent on analogy, and that as we are unable to uphold a principled distinction between the literal and the metaphorical, that all meaningful uses of language should be thought to depend on analogy. I argue that conceiving of meaning as governed by analogy, rather than rules, helps to diffuse the rule-following paradox, laid out by Saul Kripke. I claim that the meaning of every linguistic expression is governed by analogy, including the word ‘rule’, such that there may be scope to speak of ‘rules’ of meaning, after all.*

### **Candidate's declaration**

I, Andrew Simeon Williams, do hereby certify that this thesis, submitted for the degree of MPhil, which is approximately 39,728 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2015.

I confirm that no funding was received for this work.

Date

Signature of candidate

### **Supervisor's declaration**

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of MPhil in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date

Signature of supervisor

### **Permission for publication**

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Andrew Simeon Williams, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

**Printed copy**

No embargo on print copy.

**Electronic copy**

No embargo on electronic copy.

Date

Signature of candidate

Date

Signature of supervisor

**Underpinning Research Data or Digital Outputs****Candidate's declaration**

I, Andrew Simeon Williams, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date

Signature of candidate

## **Contents**

Introduction – page 1

Chapter 1 – Motivating Examples – page 5

Chapter 2 – Prototype Theory – page 26

Chapter 3 – Metaphor – page 54

Chapter 4 – The Rule-Following Paradox – page 77

Conclusion – page 102

Bibliography – page 105

## Introduction

In their book ‘Surfaces and Essences’, Douglas Hofstadter and Emmanuel Sander claim that ‘analogy-making’ and ‘categorisation’ denote the very same cognitive process.<sup>1</sup> Specifically, the idea is that we categorise an object –  $x$  – as belonging to a category –  $A$  – because we observe an analogical resemblance between  $x$  and other objects we have previously assigned to the category  $A$ . This contrasts with the classical view of category-membership, according to which we categorise an object  $x$  as a member of  $A$ , because it meets the necessary and jointly sufficient conditions for being a member of  $A$ . I want to explore and develop this claim, specifically with respect to those categories that are denoted by linguistic labels. Construed as a claim about meaning, this idea can be understood as the claim that when we apply a word meaningfully to some newly-encountered object, it is necessary that we draw an analogy between objects we have previously applied that word to (or heard or seen it applied to) and the newly-encountered object. I’ll call this the ‘analogical account’ of meaning. This contrasts with the classical view of meaning as a rule-governed practice (where words apply to objects on the basis of those objects meeting the necessary and jointly sufficient conditions for the word to apply).

In chapter 1, I will look at some motivating examples for the analogical view. I will consider a wide variety of word-types, and will observe that it is very hard to pin down precisely what it is that every object, action, or situation, to which any given word can be meaningfully applied, has in common. However, I will point out that it is clear that the instances of any given word do resemble each other in various ways, such that the perception of analogical resemblances between particular instances may be what governs meaningful language use, rather than rules. I will indicate that the analogical account of meaning is supposed to be universal in its scope – thus every word that has a use in a language, has a meaning in that language, and such meaning is governed by analogy, rather than rules. As I will indicate –

---

<sup>1</sup> Hofstadter&Sander(2013)p.19

this even applies to the key terms in the discussion – ‘analogy’ and ‘rule’.

In chapter 2, in light of the examples considered in chapter 1, I will claim that insofar as each meaningful word can be thought to stand for a category, containing the kind of things to which the word can be meaningfully applied, these categories must be rather different than the classical account of categories supposes (on which category-membership is defined by necessary and jointly sufficient conditions). I will explore Eleanor Rosch’s proposal for an alternative account of categories – known as prototype theory. According to this account, while there are family resemblances between the instances of any given category, there is no one thing that all category-members have in common. On this account, category-membership is not an all-or-nothing matter, with some objects being regarded as belonging more fully to a category than others, by being more prototypical members of the category. I will explore a series of experiments that were conducted in order to demonstrate that categories are structured in this way, and specifically, that the degree to which is an item is judged to be prototypical, correlates with the degree to which that item has properties in common with other members of the category. I will argue that while it is possible to interpret the experimental evidence differently, there are good reasons to adopt Rosch’s account, and that it works well alongside an analogical account of meaning.

In chapter 3, I will look at the question of how a classical, rule-following account of meaning may conceive of metaphorical uses of linguistic expressions, in contrast to literal uses. I will argue that classical accounts will equate literal uses of expressions with uses that are governed by the rule, but that they must conceive of metaphorical uses in some other way. One option is to claim that there is no such thing as metaphorical meaning, in contrast to literal meaning, with the difference between literal and metaphorical uses explained pragmatically, rather than by a difference in meaning. I will argue that the only alternative for a classical account, is to admit that there is such a thing as metaphorical meaning, but that such meaning could not plausibly be governed by rules, so must be thought of as governed by analogy. I will then consider the phenomenon of dead metaphors – expressions that were once used

metaphorically, but through repeated use, have gone on to develop a new literal meaning. I will argue that dead metaphors present a problem both for the classical account, which denies that there is such a thing as metaphorical meaning, and for the other account on which metaphorical meaning, in contrast to literal meaning, is governed by analogy. I will argue that the commonsense view that metaphors die slowly, should give us reason to believe that there is no sharp distinction in kind between the literal and the metaphorical use of an expression. I will then claim that even when it comes to using language in the present, it is hard to uphold the idea that there is a sharp distinction between the literal and the metaphorical, seeing as many of the things we say on a daily basis are neither clearly literal, nor clearly metaphorical. I will argue that literal and metaphorical uses of linguistic expressions alike should thus be thought to be governed by analogy.

In chapter 4, I will consider the rule-following paradox – a paradox about meaning, interpreted by Saul Kripke, from Ludwig Wittgenstein, in his book ‘Wittgenstein: On Rules and Private Language’. This paradox sets out on the idea that meaning something by an expression must consist in following the rule for the meaning of that expression, and leads to the absurd conclusion that language is meaningless. This is because, for any given usage of an expression, there is not some fact that we can cite to demonstrate that it was governed by one rule, rather than any of indefinitely many other rules. I will argue that there is not some fault in the paradoxical reasoning, but that the paradox may be solved, or at least, diffused, if we conceive of meaningful language use, not as governed by rules, but as governed by analogy. The analogical solution that I will propose has significant points of contact with the skeptical solution, which Kripke himself lays out. However, according to the skeptical solution, while there is not some fact that we can cite to demonstrate that a particular rule is being followed, it is judged to be appropriate, in some circumstances, to speak of what someone means by a word as if it were governed by a rule. I will then return to the point raised at the end of chapter 1, that given that the analogical account is universal in its scope, it also applies to the key terms in the discussion – ‘analogy’ and ‘rule’. Given that even the word ‘rule’ is governed by analogy, I will argue that, there may, after all, be good reason for us to speak of the

‘rules’ that govern meaning in some circumstances, insofar as that which governs meaning is seen to have properties in common with some of those things we typically call ‘rules’. However, I will argue that there are also important dissimilarities, between that which governs meaning, and various kinds of rule, such that meaning something by a linguistic expression is best characterised analogically, rather than in terms of rule-following.

I will then conclude, that meaning something by a linguistic expression should not be thought to consist in following a rule, but should be thought of as governed by analogy. If meaning was governed by rules, then we should expect to be able to identify just what it is that every member of a category corresponding to any given word has in common. Chapter 1 demonstrated that this is, at best, extremely difficult. Chapter 2 indicated an alternative way in which we may conceive of categories, in light of the examples considered in chapter 1, which lends strong support to the analogical account of meaning. Chapter 3 looked at how classical accounts may conceive of metaphorical uses of expressions, in contrast to literal ones, and argued that the distinction was hard to uphold, but that both literal and metaphorical uses alike could be said to be governed by analogy, rather than rules. Chapter 4 looked at the rule-following paradox, which demonstrates that insofar as meaning something by an expression is thought to consist in following rules, then we can never know which rule, if any, we are following. I propose that an analogical account of meaning can diffuse the paradox, before returning to the idea that the analogical account is universal in its scope, and even applies to the words ‘analogy’ and ‘rule’. I argue that it may thus be legitimate to talk of the ‘rules’ of meaning after all, but that the best way to characterise meaning something by an expression, is as governed by analogy, rather than rules.

## Chapter 1 – Motivating Examples

### Introduction

In this section, I will consider a large variety of word-types, and argue that there is no one thing that all instances of any given word of each type has in common, such that all instances may be captured in terms of a rule, stating necessary and jointly sufficient conditions for the meaningful application of the word. However, I will point out that there are undeniable similarities between the instances of any given word, such that the meaning of any given expression might be thought to be governed by analogy, rather than by a rule. Specifically, I will consider nouns, verbs, connectives, articles, expressions of greeting and gratitude, and expressives. Each meaningful expression may be thought to stand for a category, containing the kind of objects, actions and situations that the expression can be meaningfully applied to. In the next section, I will argue, that if we are adopt an analogical account of meaning, as the word-types considered in this section may lead us to – then a different account of categories is needed, rather than the classical account, on which category-membership is governed by formal membership criteria.

### Nouns

When we recognise something as a table, as something to which the linguistic label ‘table’ may be appropriately applied, it is not the case that we unconsciously, or otherwise, go through some checklist to ensure that the object before us meets the necessary and jointly sufficient conditions for being a table, for having the word ‘table’ be appropriately applied to it. When we make such a recognition, causing us to think or utter “that is a table”, to suggest that our meaningful use of the word ‘table’ is the result of our having followed the rule for the meaningful use of that word, would imply that what leads us from our perception of the object before us to producing that sentence (in thought or speech) is that we infer that that object meets the conditions for tablehood, or ‘fits’ the definition of ‘table’. However, it seems to me that rather than inferring that the object before us meets the definition of ‘table’, what causes our

recognition and our meaningful application of the word ‘table’, is simply that we see a resemblance – an analogy, between that object and previous objects we have designated ‘tables’ (the history of which goes right back to the first time we were told of some object: “that is a table”.) Some might claim that, of course, when we recognise something as a table, and apply the word ‘table’ to it, we are observing or inferring, probably unconsciously, that it does meet certain conditions that must be met for something to be a table. One dictionary definition of ‘table’ is the following: ‘a piece of furniture with a flat top and legs, for eating, writing, or working at.’<sup>2</sup> Perhaps, the thought might go, when we recognise something as a table, we are observing, or inferring, that it fits some such definition. There are problems with any such putative definition, however.

Putting aside the issue that the surface of no piece of furniture is strictly speaking flat (at the molecular level), it seems that many things we would be disinclined to call ‘tables’ fit the above definition. Take, for example, ironing boards. They are pieces of furniture, presumably, they tend to have flat tops and legs, and are designed for working at (assuming ironing can be construed as work). Another example may be desks - are some, or all, desks tables, or are they mutually exclusive? Equally, many things we would unhesitatingly call ‘tables’ do not fit this definition. An antique table in a museum may explicitly not be for eating, writing and working at. A table may have a bumpy surface (by design or decay). Something could presumably still be a table even if it didn’t have legs – say if it was a solid block, or was held up by criss-crossing boards. Even if someone insisted that following such a definition, or an improved definition, to the letter would successfully sort all the tables from the non-tables, the problem arises again for the words that constitute this definition. It is not universally agreed what counts as furniture, what counts as working, what counts as flat, and so on. So even if we recognise something as a table because we perceive that it has certain properties, we also recognise these properties (say the property of having a flat surface, legs, etc.) by analogy with our past experience of perceiving objects with such properties. Furthermore, if we asked a number of competent speakers proficient with the word ‘table’ to offer some

---

<sup>2</sup> Little Oxford English Dictionary, Eighth Edition, Oxford University Press, 2002, p.717

definition of 'table' (without consulting a dictionary), these definitions would be idiosyncratic and stipulative, and would disagree on what exactly does count as a table, but this would not seem to undermine the linguistic competence of those who gave these definitions.

One option, for those who insist that something qualifies as a table by virtue of meeting the necessary and jointly sufficient conditions for tablehood, would be to give a purely functional definition of 'table'. So a table is some raised platform that is used for eating, writing, working, perhaps placing drinks, or flowers, or books on. This would allow us to say that even a cardboard box could be a table, if a group of people sat around it and ate a meal on its surface. This would seem to be an improvement, but there is still the question of what constitutes a legitimate function for something to be used for if it is to count as a table. And some people may be uncomfortable with the consequences of such a definition – could a bookcase really count as a table, if I ate breakfast off its top? Would it still be a bookcase while it was being used as a table, or only if it was being used to store books? Could a bed be a table? Could an object in a region that was spatio-temporally isolated from our universe (and from any life), and hence could never function as a table, be a table? In any case, no matter how sophisticated a functional definition we may offer for something to count as a table, there are still going to be things that not everyone will agree does count as a table included in the extension of 'table' by such definitions.

It seems to be no easy question just what it is that every table (objects which may appropriately be called 'tables') has in common. Yet, in spite of this, we all seem pretty confident when we go around calling things 'tables'. It seems naïve to think that there is in fact some determinable number of tables in the universe at any one moment. But most of the time, speakers of a common language are unlikely to have long and protracted disputes about whether some object counts as a table, or not. A related question does not concern whether or not some given object counts as a table, but what the meaning of the abstract concept *table* actually is. There are still problem cases that arise when we discuss tables in the abstract. Is a table that has lost one of its legs still a table? Is a table that still has its legs, but that would collapse if any object were to be placed on its surface still a table? The problem of what constitutes a table arises just

as much when we discuss the concept in the abstract, for we can use other equally abstract notions to create problem cases. It seems that the word 'table' is both hard to define, and that whatever definition we attempt to provide, be it descriptive or functional in character, is prone to have borderline cases. Thus any putative definition will include things that many or most competent speakers would not feel comfortable calling a 'table', and perhaps to exclude others which most speakers would insist were tables. The same is true of many other common nouns, like 'chair' and 'bookcase'.

Every object, action and situation is unique, yet we call many collections of such objects, actions or situations, by a common linguistic label. Suppose someone were to press the point for common nouns such as 'table', 'chair', and 'bookcase', and insist that we can express precisely what it is that every instance of such a concept has in common (I do not think this can be done – much less in terms of the expression of a rule that is shared, implicitly or otherwise, by every competent speaker of a language). Other nouns such as 'mess', 'heap' or 'game', would seem to leave the advocate of necessary and jointly sufficient conditions for the meaning of each noun in a real mess, their aspirations in a heap, realising they're playing a futile game. A bedroom, a political party, a romantic relationship, or a philosophical project can be a mess. The ancient version of the Sorites paradox demonstrates that it is arbitrary and stipulative to insist that only collections of objects (e.g. grains of sand) over a certain number can constitute a heap (leaving aside the issues of how these objects must be organised). Games can be played with cards, balls, or linguistic expressions, among countless other things, and it's far from clear what all these games have in common. It seems utterly impossible to capture all such instances or potential instances of games in terms of necessary and jointly sufficient conditions - in such a way that all non-instances or potential non-instances are excluded. So it seems that there are, at the very least, many nouns, the meaning of which cannot be understood in terms of each instance of that noun following the rule for the meaningful application of that word. I will turn now to verbs (and just what is it that every instance of *turning* has in common?).

## Verbs

As indicated at the end of the last paragraph, verbs such as ‘turn’ also apply to many things, such that it is impossible to state some general principle that every instance of turning has in common. One can turn by changing the physical direction in which one is moving, or else by changing the focus of one’s attention. One can turn from one topic to another, in the course of a conversation or text. It is not only agents that can turn - a road can turn (even though the road itself remains essentially stationary), wheels can turn, the weather can turn, or even the tables can turn (though situations described by such an expression will be likely to involve agents, not tables). It seems absurd to propose that, at this moment, if we looked at the entire past history of the universe, there would be some determinate number of instances of turning, in any or all of the above senses (and many others unmentioned) that we could identify (supposing we were omniscient). It does not seem that every instance of turning obeys some principle that can be isolated, such that we can determine precisely how many instances there have been. Some might claim that only some of the above instances of turning are literally, or strictly-speaking, instances of turning. Perhaps they would claim that genuine turning must involve some kind of change in physical direction, while the idea of turning from one subject to another, or of the tables turning, must be construed as loose-talk, as a figurative way of speaking that we should not expect to draw a sharp boundary between instances and non-instances, unlike those of literal instances of turning. However, it seems to me that there is no sharp boundary between literal and figurative applications of the verb ‘turn’ – it seems unclear, for example, whether a road can really turn, in the supposedly literal sense of changing physical direction, seeing as it is surely only those objects or agents that move along the road that turn, while the road itself remains static.

Is it loose talk to say that I will now turn from one subject to another, or is this a perfectly literal use of the word ‘turn’, despite it not involving a change in physical direction (but rather a change in subject-matter)? It would seem arbitrary to make this insistence. Another issue might be that even if we could

isolate some kind of literal meaning of ‘turning’, thereby pushing away the supposedly figurative, loose meanings, what constitutes one turn even in this sense is totally stipulative and arbitrary. If I turn forty-five degrees when I am walking along, do I just turn the one time, or do I turn one degree, and then another one degree, and yet another... etc, until I reach the forty-five degree pivot? Even if we were to constrain the literal meaning of ‘turn’ to changes in physical direction, it seems a matter of choice, rather than a matter of fact, how many turns we should construe each relevant course of action as consisting in. So it seems that verbs, no less than nouns, cannot be thought to apply to actions and situations by virtue of those actions and situations meeting the necessary and jointly sufficient conditions for the verb to apply. It seems that what all the instances of turning discussed above have in common, is, to quote Wittgenstein (in his ‘Philosophical Investigations’): “a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail.”<sup>3</sup> We call something an instance of ‘turning’ because it has something in common with other instances of turning we have previously experienced. This cannot be captured in terms of a rule that all instances and potential instances of turning instantiate, and that all non-instances or potential non-instances do not.

I have discussed nouns and verbs, and how I think the objects, actions and situations such words can be applied to are best accounted for analogically, rather than in terms of precise membership criteria. However, the thesis under consideration – that each and every meaningful use of a word requires the user of that word to draw an analogy – is more far-reaching than that. While it is nouns, and to a lesser extent verbs, that philosophers have often attempted to spell out in terms of there being necessary and jointly sufficient conditions for the word to apply, I think that any word that has a use in a language has a meaning in that language, and thus that the meaning of all words that have a use in a language can be accounted for analogically. This includes words that seem to play a purely logical role, such as connectives (the likes of ‘and’ and ‘but’), and articles (‘a’ and ‘the’). This also includes very common words that serve a social function, which do not seem to be representational in character, such as ‘hello’,

---

<sup>3</sup> Wittgenstein(1996)p.32

‘goodbye’, ‘thanks’ and ‘sorry’. This also, significantly for the thesis under consideration, includes expressions such as ‘analogy’ and ‘rule’, which I will indicate at the end of this section, before returning to this point in the final section.

### Connectives

As Saul Kripke highlights in his paper “‘And’ and ‘But’”, most philosophers hold that the words ‘and’ and ‘but’ are “truth-functional equivalents” – and thus that they mean the same - with any difference affected by using one word rather than the other residing at the level of pragmatics, rather than semantics.<sup>4</sup> Specifically, such accounts claim that while sentences like ‘A and B’ and ‘A but B’ mean the same, the latter sentence, featuring the word ‘but’, involves a conventional implicature that B is unlikely or unexpected given A, while the former sentence, featuring the word ‘and’, involves no such conventional implicature.<sup>5</sup> The idea is that in using the word ‘but’ to connect the two clauses, we take the speaker to be intending to convey something beyond the meaning of the sentence they uttered – namely that the latter clause, B, somehow conflicts with our expectations, in light of the former clause, A. Kripke claims that while such accounts; of the kind of contrast between ‘A’ and ‘B’ that the use of the connective ‘but’ implies, differ in emphasis, they tend to agree with the spirit of Frege’s account of the matter – “a speaker uses ‘but’ when he wants to hint that what follows is different from what might at first be supposed.”<sup>6</sup> The use of word ‘and’, by contrast, is thought to involve no such hint of a contrast between the clauses. I will consider Kripke’s arguments against this widespread, Fregean-style view, and find them to be conclusive, before suggesting, in line with Hofstadter and Sander, that the difference between ‘and’ and ‘but’ is a difference in meaning, and that this difference in meaning can be accounted for analogically.

Kripke considers a number of examples to show how the widespread, Fregean-style view on the

---

<sup>4</sup> Kripke(2017)p.102

<sup>5</sup> Kripke(2017)p.103

<sup>6</sup> Kripke(2017)p.102

difference between ‘and’ and ‘but’ is wrong. He cites Michael Dummett’s example, in which the possibility of inviting a speaker to talk at a conference is being considered:

(1) He is always a very popular and successful speaker, but he is in America for the year.<sup>7</sup>

We’ll call the first part of this sentence (‘He is always a very popular and successful speaker’) ‘A’, and the latter part of this sentence (‘he is in America for the year’) ‘B’. Kripke points out that, contrary to the widespread view of the difference between ‘and’ and ‘but’, the use of the word ‘but’ in this sentence does not imply (pragmatically or otherwise) that B is unlikely, or unexpected, given A, nor is there any “intrinsic” conflict between the two clauses.<sup>8</sup> Instead, clause A gives a reason in favour of inviting the speaker, while clause B gives a reason against the invitation. This suggests that the link between the clauses that the use of the word ‘but’ introduces, can be rather more subtle than the widespread, Fregean-style view claims. Kripke takes this example to demonstrate that the Fregean account of the word ‘but’ does not provide a necessary condition for its use.<sup>9</sup> He then considers another sentence to demonstrate that the Fregean account also fails to provide a sufficient condition for the use of the word ‘but’:

(2) She entered the lottery, but she won.<sup>10</sup>

Kripke points out that while B (‘she won’) in this case, is unlikely to be true given A (‘She entered the lottery’), insofar as the overwhelming majority of people who enter the lottery do not win – there is something wrong with this sentence – it involves an inappropriate use of the word ‘but’ (where ‘and’ could appropriately have been used instead, contrary to the Fregean account).<sup>11</sup> Thus even if one clause, B, is unlikely or unexpected given another – A, this is no guarantee that the word ‘but’ can be appropriately used to connect the two clauses in the sentence ‘A but B’.

Kripke’s arguments against the Fregean-style view of the difference between ‘and’ and ‘but’ reveal that there is not some straightforward conventional implicature that the word ‘but’ introduces, which can

---

<sup>7</sup> Kripke(2017)p.102

<sup>8</sup> Kripke(2017)p.102

<sup>9</sup> Kripke(2017)p.102

<sup>10</sup> Kripke(2017)p.102

<sup>11</sup> Kripke(2017)p.103

successfully distinguish between sentences where the word ‘but’ can be appropriately used, and sentences where its use would be inappropriate (and perhaps ‘and’ would be appropriate instead). Furthermore, Kripke goes on to look at examples which demonstrate that in at least some sentences using the word ‘but’ rather than the word ‘and’ would make a difference to the meaning of the sentence – and by using one word rather than the other, a speaker is indicating that only their particular word choice, be it ‘and’ or ‘but’ respectively, appropriately conveys their attitude. Kripke gives the example of an anti-Semitic classics department, considering the appointment of someone to an academic position. An anti-Semite in such a department might utter:

(3) He is an excellent classicist, but he is a Jew.<sup>12</sup>

While this sentence need not imply that there is any contradiction between the predicates expressed by the two clauses (A – ‘being an excellent classicist’, B – ‘being Jewish’), as a Fregean style account would predict, the first clause gives a reason for the appointment, and the second clause gives a reason against it (much like in sentence (1)). The use of the word ‘but’ in this sentence expresses the anti-Semitic attitude of the speaker. Conversely, Kripke considers what the sentence would express if the word ‘and’ had been used instead:

(4) He is an excellent classicist, and he is a Jew.<sup>13</sup>

Kripke claims that such a sentence would imply that both clauses gave reasons for the appointment (and thus conveyed that the speaker had a positive attitude towards appointing Jews, as well as a positive attitude towards appointing excellent classicists to the department).<sup>14</sup> Thus if the use of the word ‘but’ introduces a conventional implicature of some kind, then the word ‘and’ can be said to also – since it can imply that both clauses (A and B) provide reason to do something (and not merely that both are true).

While it may be possible, therefore, to explain the significance of using ‘and’ rather than ‘but’ (or vice versa) by claiming that each introduces a conflicting conventional implicature (albeit one rather more

---

<sup>12</sup> Kripke(2017)p.103

<sup>13</sup> Kripke(2017)p.104

<sup>14</sup> Kripke(2017)p.104

complicated than that suggested by Frege for the word ‘but’), as Kripke suggests<sup>15</sup> – it seems to me that a better case can be made for the claim that there is a genuine, semantic difference between the words ‘and’ and ‘but’, rather than a merely pragmatic one. The words ‘and’ and ‘but’ are simply not interchangeable – and changing one for the other can make a genuine difference to the meaning of the sentence. Hofstadter and Sander provide a positive account of the difference in meaning between ‘and’ and ‘but’ – and emphasise that the difference in the appropriateness of using one word rather than the other really should be regarded as a difference in meaning – for in using the word ‘and’ or ‘but’, one is in fact making categorisations, just as one is in using the word ‘dog’ or ‘cat’. In using the words ‘and’ and ‘but’, they claim, we are categorising our own speech, and the way that our sentences are connected together. We are, as they put it, categorising “motion in the space of discourse”.<sup>16</sup> We use the word ‘and’ when “motion in the space of discourse continues smoothly along a pathway that has already been established”, while we use the word ‘but’ when “motion in the space of discourse makes a sudden, unexpected swerve”.<sup>17</sup> More specifically, in sentences of the form ‘A but B’, they claim, “the first part of the sentence [A] suggests a pattern and the second part [B] states an exception to the pattern”.<sup>18</sup> This characterisation of the appropriateness of using ‘and’ or ‘but’, might seem to agree with the Fregean-style accounts in some sense (in which some kind of conflict between the two conjoined clauses is implied pragmatically). However, on Hofstadter and Sander’s account, the appropriateness of one word rather than the other depends on our linguistic categorisations, and thus on the appropriateness of the meanings of the words ‘and’ and ‘but’ to that which we are categorising. Furthermore, their account can handle all of Kripke’s data, which served as counterexamples to the Fregean account.

In Kripke’s sentence (1) – clause A was said to provide a reason for inviting a speaker, while clause B was said to provide a reason against the invitation. Hofstadter and Sander’s account can accommodate

---

<sup>15</sup> Kripke(2017)p.102

<sup>16</sup> Hofstadter&Sander(2013)p.72

<sup>17</sup> Hofstadter&Sander(2013)p.72

<sup>18</sup> Hofstadter&Sander(2013)p.73

such sentences – the word ‘but’ is appropriate here because one of the ways in which clauses or sentences can be categorised is as “providing reason for X”, and another is as “providing reasons against X”. Thus clause A was categorised as providing reason for the invitation, and clause B as providing reasons against the invitation, and therefore, the connection between the two clauses in the ‘space of discourse’ can be categorised as a kind of swerve, or the stating of a pattern, followed by the stating of an exception to that pattern. Kripke’s sentence (3) can be dealt with in the same way (as stating a reason for the appointment, followed by a reason against the appointment), and sentence (4), using ‘and’ analogously (with the word ‘and’ implying that both clauses provide reasons for the appointment). There is a subject matter we are categorising here – the way that our clauses and sentences are connected together, and we can categorise this subject matter appropriately or inappropriately, by using (and meaning) ‘and’ or ‘but’. Hofstadter and Sander give an example of how someone may do this inappropriately, pointing out how odd it would sound if someone, on returning from a visit to Paris, was to say ‘I like Paris but I like Parisians’ (and how natural, conversely, it would sound for them to say ‘I like Paris and I like Parisians’).<sup>19</sup> They point out that if the sentence ‘I like Paris but I like Parisians’ was to show up in an e-mail, we would perhaps expect the writer to have made a typo – surely they meant ‘I like Paris but I don’t like Parisians’ (or conversely ‘I don’t like Paris but I like Parisians’).<sup>20</sup> This difference in the appropriateness of using ‘and’ or ‘but’ can be understood as a difference resulting from the meaning of the words, on the basis of which we make our categorisations.

On the face of it, it might seem as though Hofstadter and Sander’s account will have as much trouble with sentence (2) as the Fregean account. It might seem as though clause A (‘she entered the lottery’) states a pattern with which clause B (‘she won’) conflicts, given the extreme unlikelihood of any given person who enters the lottery winning the lottery. However, perhaps given that entering the lottery is a necessary condition of winning the lottery (but a long way from being a sufficient condition!), people do

---

<sup>19</sup> Hofstadter&Sander(2013)p.72

<sup>20</sup> Hofstadter&Sander(2013)p.72

not tend to hear clause A as stating a pattern to which clause B states an exception. By contrast, “She entered the lottery, but she didn’t win” sounds all right, in spite of the overwhelming likelihood of the two clauses both being true, when stated of any given person (though I think that “and she didn’t win” would sound all right too). As Kripke suggests<sup>21</sup>, explicitly mentioning the improbability of winning in clause A makes sentences like (2) sound appropriate, as in:

(5) She entered a competition she had very little chance of winning, but she won.

As Hofstadter and Sander point out, the ‘swerve’ in discourse space required by the use of the word ‘but’ can be rather subtle (they give the example of “He has big ears, but he’s a really nice guy” sounding all right, despite there being no correlation between ear size and niceness).<sup>22</sup> There is certainly real subtlety to the manner in which we choose between ‘and’ and ‘but’, and whether either word sounds right in any given sentence, but despite this subtlety, it seems as though we are categorising a real subject matter – how our discourse flows - and we choose to use one word rather than the other, because each word has a different meaning, which guides us in our categorisation of discourse. Indeed, this subtlety demonstrates that there is not one rule for the meaning of ‘and’ and another rule for the meaning of ‘but’ – rather, while the two words differ in meaning, each word is applied in any particular instance (in categorising the flow of discourse) based on analogical extrapolation from previous examples encountered.

This characterisation allows us to explain the apparent difference between ‘and’ and ‘but’ as a genuine difference in meaning – and, of course, we identify whether we are in an ‘and’-situation or a ‘but’-situation by analogy. Children are never taught that the word ‘and’ means one thing, while the word ‘but’ means another (as they are taught by example, say, that the word ‘cat’ means one thing, and ‘dog’ another) – rather, they encounter a large number of examples of ‘and’ and ‘but’-situations, and gradually they get it, and are able to use the words ‘and’ and ‘but’ appropriately, by extrapolating from their growing body of experience. They learn to categorise the logical flow of their own discourse by observing

---

<sup>21</sup> Kripke(2017)p.103

<sup>22</sup> Hofstadter&Sander(2013)p.73

others doing the same, just as they learn to categorise cats and dogs by observing other people's categorisations using the words 'cat' and 'dog'. Hofstadter and Sander point out that in the process of acquiring the word 'and', children will be likely to initially encounter the word being used to link people or objects (such as 'Mommy and Daddy') before they encounter the word linking people and objects ('Sally and her toy') and progressively onto even more abstract combinations, linking times ('I went out and looked') and even causal links ('It fell and broke').<sup>23</sup> Thus, just like the other words we have considered, the word 'and' is acquired by encountering examples of it, and extended by analogy with the examples encountered. I will now discuss the articles (the words 'a' and 'the') words whose contribution to the meaning of a sentence has generally been thought to be precise and unambiguous, which I will question, in favour of explaining the meaning of these expressions, too, analogically.

### Articles

While the articles 'a' and 'the' are often not regarded as meaning anything in themselves, but merely as grammatical devices, it is clear that the use of one word rather than the other can make a difference to the meaning of a sentence, or phrase. As Hofstadter and Sander point out: 'a survivor died' means something different to 'the survivor died', and 'I married the man in the photo' means something different to 'I married a man in a photo'.<sup>24</sup> Specifically, the word 'the' in 'the survivor died' tells us that there was only one survivor (in whatever accident or incident is at issue), and that the only survivor has died, while the word 'a' in 'a survivor died' implies that there was more than one survivor, but that someone who was a survivor has died. Bertrand Russell's theory of descriptions sought to provide an analysis of the truth conditions of sentences involving the word 'the' (the definite article). Russell's theory tells us that when a sentence contains a definite description ('the F') – a description preceded by the definite article, such as 'The F is G', then this sentence is true if and only there is exactly one thing that is F, and that thing is also

---

<sup>23</sup> Hofstadter&Sander(2013)p.71

<sup>24</sup> Hofstadter&Sander(2013)p.76

G.<sup>25</sup> This contrasts with sentences containing the indefinite article ('an F') of the form 'An F is G', which do not imply that there is exactly one thing that is F, but just that something that is F is G. Russell's theory identifies the difference in meaning indicated by Hofstadter and Sander, while providing, it would seem, necessary and jointly sufficient conditions to use the words 'the' and 'a' appropriately (insofar as one wishes to assert what is true).

However, it is possible to provide counterexamples to the theory of descriptions – uses of the words 'the' and 'a' which sound fine to competent speakers, even though they do not carry the truth conditional implications suggested by Russell's theory. For example, we can say 'I've walked the dog' without implying that there is only one dog in the world (and it would sound very odd to say 'I've walked a dog', instead). Some attempts have been made to save the theory of descriptions from such counterexamples by claiming that in such sentences, the definite description ('the F') refers not to the only F in the world, but to the only contextually-relevant F. Thus such an account could claim that there are necessary and jointly sufficient conditions for the use of the articles, it's just that these conditions – and specifically the uniqueness condition (that there is exactly one F) may be relative to a context. However, if many uses of the word 'the' do require contextual supplementation, in order for Russell's account to work, we might ask what it is that determines the relevant context in any given usage of the word 'the' – could it be that there's some rule that governs the relevant context for any given use of the word 'the'? It doesn't seem as though there could be some rule, stating, that when reference is made to 'the dog', this refers to the only dog in the house, while when reference is made to 'the milk', this refers to the only milk in the fridge, and so on. It seems as though the contextual restriction from which Russell's theory requires supplementation, cannot itself be captured in terms of rules stating necessary and sufficient conditions which determine the relevant context for any given usage. Rather, it seems as though the context that a definite description quantifies over is governed by analogy. We learn, by encountering examples of it, just what kind of objects are being quantified over by any given use of a definite description.

---

<sup>25</sup> Russell (1910), p.127

On top of the need for contextual supplementation for many expressions of the form “The F”, there are other problematic cases for Russell’s theory too – we may say that someone was ‘punched in the ear’ (even though they have two ears), or that ‘the son of the President is obnoxious’ (even if the President has more than one son). Assuming we take such uses to be legitimate uses of the word ‘the’, then it becomes clear that just like the nouns and verbs we have considered, there are not clear membership criteria, or rules, regarding the use of the articles ‘the’ and ‘a’ respectively. The theory of descriptions cannot successfully sort all uses of the word ‘the’ into appropriate and inappropriate uses – at least, if what sounds right to competent speakers is a good test of whether a word is being used appropriately, or not. Further, as I argued, in the many cases where contextual supplementation is required, it seems that the context itself cannot be supplied by rules, but must be supplied by analogy. We learn to use the words ‘the’ and ‘a’, just like any other words, by hearing them used, and then, as we acquire the term, using them in other situations, and sentence-combinations that strike us as similar to those we have experienced in some way. We extend the words ‘the’ and ‘a’ into new contexts by analogy, rather than by following a rule. Having explained that even connectives and articles, words which seem to serve a purely logical, grammatical function, can be explained analogically, I will now consider what the analogical account can say about words that serve a social function, but that do not seem to be representative in nature, such as ‘hello’ and ‘sorry’.

#### Expressions of greeting and gratitude

Nobody (to my knowledge) has ever attempted to claim that the word ‘hello’ can only be used correctly if certain necessary and jointly sufficient conditions are met. Nonetheless, there are some situations in which the word ‘hello’ seems like an appropriate thing to say, and others in which it seems strange. It would seem an appropriate thing to say first of all upon meeting someone, for example, but a very odd thing to say to someone half-way through a conversation, or at the end of a conversation, when both conversational partners are about to go their separate ways (in which case, ‘goodbye’ would seem more appropriate). Many theories of meaning will have little to say about such words, and the

significance and appropriateness of their use in different contexts will likely be said to be accounted for pragmatically, rather than semantically. However, I will argue that they too should be treated as meaning something, and that an analogy underlies each and every meaningful use of these words, just like the word-types I have already discussed.

Hofstadter and Sander point out that analogy plays a vital role when we greet someone, or thank someone, just as it does when we identify something as a ‘dog’ or a ‘table’. They discuss how Hofstadter, having travelled to an Italian university to work as a visiting professor, came to learn which Italian greeting was appropriate for each of the people he met in the corridors.<sup>26</sup> In Italian, there are (at least) three different common greetings, each of which has a different degree of formality, and so assumes a different degree of familiarity with the person to whom it is addressed. These three greetings are ‘Ciao’; which is the most informal, and is used to address people one knows well, ‘Salve’; which is slightly less informal, and is used to address people one knows less well, and ‘Buongiorno’; which is most formal, and is used to address people one either doesn’t know, or knows hardly at all. Hofstadter learnt to use the appropriate greetings in the appropriate situations by treating certain people as prototypical members of each category – that is of the categories – people it is appropriate to greet with ‘Ciao’, ‘Salve’, and ‘Buongiorno’, respectively. Hofstadter explains that when he greeted someone in the corridor, his reasoning would go something like this: “Hmm ... This fellow who’s approaching me, I know him roughly as well as I know that tall curly-haired administrator”<sup>27</sup> – leading him to produce the greeting ‘Salve’. Hofstadter observed how other people greeted him, and once he had at least one good example of a person belonging to each category (category of people it is appropriate to greet with such-and-such), he would greet other people he met by analogy – by comparing how well he knew these people with how well he knew the people who had greeted him with ‘Ciao’, ‘Salve’ and ‘Buongiorno’, respectively. Of course, over time, each of the categories *people it is appropriate to greet with*

---

<sup>26</sup> Hofstadter&Sander(2013)p.46

<sup>27</sup> Hofstadter&Sander(2013)p.46

*'Ciao'/'Salve'/'Buongiorno'* would grow in size by accepting more members as Hofstadter's grasp of Italian greetings grew, and he would no longer have to think explicitly of the first person who had greeted him 'Salve', for example, when uttering that greeting to someone he met. Nonetheless, each utterance of one of these three greetings would depend on Hofstadter making an analogy to other people he has had greeting-in-Italian experiences with. Hofstadter would choose the appropriate greeting based upon his growing body of experience.

There is a clear analogy here, between learning to use the appropriate Italian greeting for each person one meets, and someone's learning to use any other word, such as the word 'table', appropriately. Originally, they will call objects 'tables' because they resemble some of the early, prototypical objects they have been told are called 'tables'. Over time, as more and more objects are accepted into the category 'table', by this person's interaction with other members of the community (and various objects the community calls 'tables'), they will not have to even implicitly compare the object before them to that first object they were told is called a 'table', but their word-choice will nonetheless depend on their body-of-experience of people calling objects 'tables'. Just as we call an object a 'table' because it resembles other objects we have heard called by that word, we greet someone in a particular way because the greeting-situation in which we find ourselves, resembles other greeting-situations we have experienced. Even if we didn't have three different greetings to choose between, as Hofstadter did when learning to use Italian greetings appropriately, simply identifying a situation as one in which a greeting is appropriate ('Hello', 'Hi', 'Good morning', etc.) is also dependent on analogy. We learn that a greeting is appropriate upon meeting someone, rather than say, halfway through our conversation with them, or upon parting, by experiencing greetings and growing familiar with the kind of situations they are used in, such that we eventually become experts at identifying greeting situations, and seemingly automatically produce the word 'hello', or whatever, upon meeting someone. Thus, even the word 'hello' can be used meaningfully, and its use can be accounted for analogically – we use the word 'hello' because the situation we find ourselves in resembles other 'hello'-situations we have experienced, and it is only by extrapolating from

this body of past ‘hello’-experiences that we manage to use the word ‘hello’ appropriately. Clearly, much the same goes for the words ‘goodbye’, or ‘bye’ or ‘see you later’ – we originally learn to use these expressions based on particular experiences, overtime we develop a vast body of experiences involving these words, and it is because a situation resembles other situations in which we have experienced these words that we go about using these expressions appropriately – specifically, when we are parting from a conversational interlocutor.

Hofstadter and Sander point out that when someone learns English, either as their native language, or as a second (or third, etc.) language, they too will have to learn to use various social expressions appropriately based on analogy, much as Hofstadter did with the Italian greetings ‘Ciao’, ‘Salve’ and ‘Buongiorno’. They give the example of the different expressions we have in English in order to express gratitude – ‘thanks’, ‘thanks a lot’, ‘thank you’, ‘thank you very much’, ‘many thanks’ etc.<sup>28</sup> While all of these expressions are used to express gratitude (unless, say, they are used sarcastically), there are subtle differences between situations in which we would tend to produce one of these expressions rather than another, and rather than learn formal membership criteria for ‘thanks’-situations, ‘thank you very much’ situations, ‘many thanks’-situations, and so on, respectively – we use these different expressions of gratitude by analogically extrapolating from our vast bodies of experience concerning their use. While it is not plausible that the appropriate use of the word ‘thanks’ (as opposed to, say, the expression ‘thank you very much’) can be spelled out in terms of necessary and jointly sufficient conditions, it is clear that the word has a slightly different use (to ‘thank you very much’) – and thus we can say that it means something different, and the appropriateness of its use, rather than the use of a rival expression-of-gratitude phrase, is determined by analogically extrapolating from our past experiences. Much the same goes for expressions of apology – ‘sorry’ might be appropriate if you have just accidentally walked into someone, but something stronger might be called for if you have just carelessly ran over and killed their pet dog (‘I’m very sorry’, perhaps). Different expressions of apology are suited to different levels of

---

<sup>28</sup> Hofstadter&Sander(2013)p.46

formality ('sincerest apologies' vs. 'sorry about that'), and to different levels of the importance of that which is being apologised for (killing a pet, vs. spilling a drink, and so on). We learn to use these words appropriately (just as we learn to use all words of our language appropriately) by extrapolating from our past experiences of hearing these words, and we extrapolate in this way by perceiving resemblances between the present situation and past situations we have experienced. We tend to use a particular expression of gratitude, or apology, in a given situation, rather than any of a whole host of others ones that would seem, in some way, to be less appropriate, and the analogical account of meaning tells us that this is because they mean something different, even though all that this difference in meaning consists in is our tendency to use each expression in certain contexts rather than others. I will now discuss the words 'oops' and 'ouch', which have generally not been regarded as meaning anything, but which David Kaplan has argued should be regarded as having a meaning.

### Expressives

In his lecture 'On the Meaning of 'Ouch' and 'Oops'', David Kaplan claimed that many word-types that have often been overlooked by traditional theories of meaning, should be thought of as having a meaning, insofar as there are appropriate and inappropriate ways in which each such expression can be used.<sup>29</sup> Two of the words that Kaplan focuses on, are the expressives 'Oops' and 'Ouch'. He claims that the conditions in which the word 'Oops' can be appropriately used can be vaguely characterised as situations in which someone has "just observed as minor mishap".<sup>30</sup> Thus, he points out, it would be appropriate to use the word 'Oops' having just observed someone knock over a display of glasses in a shop. However, if, somehow, their knocking over of the display had caused the entire shop to collapse, killing numerous people, the word 'Oops' would be entirely inappropriate, given the severity of the consequences of the mistake observed.<sup>31</sup> If the word 'Oops' is regarded as having a meaning in this way, then it doesn't seem that such a meaning could be captured in terms of necessary and jointly sufficient

<sup>29</sup> Kaplan(2004), <http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf> ,p.3

<sup>30</sup> Kaplan(2004), <http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf> ,p.12

<sup>31</sup> Kaplan(2004), <http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf> ,p.12

conditions. Rather, it seems that all cases in which the word ‘Oops’ can be appropriately used, do not have one sharply defined property in common, but a network of similarities between instances. Thus, if ‘Oops’ has a meaning, it seems this meaning may be governed by analogy, rather than rules. Kaplan claims that situations in which the word ‘Ouch’ can be appropriately used can be vaguely characterised as situations in which the speaker is in pain.<sup>32</sup> However, he points out that the word can also be appropriately used empathetically – not because the speaker is in pain, but because they observe somebody else in pain.<sup>33</sup> If the word ‘Ouch’ is regarded as meaning something, it once again seems difficult to explain how this meaning could be captured by a rule. It seems that there is no sharp distinction between cases in which the word ‘Ouch’ can be appropriately used, and cases in which its use would be inappropriate, but nonetheless there do seem to be analogical resemblances between particular uses of the expression, such that its meaning might be thought to be governed by analogy, rather than by a rule.

### Conclusion

I have looked at a wide variety of word-types – nouns, verbs, connectives, articles, expressions of greeting and gratitude, and expressives – and argued that we cannot sharply distinguish between the appropriate and inappropriate use of any such word, in terms of a rule. However, I have pointed out that there are undeniable similarities between the instances of any such word, such that the meaning of such terms should be thought to be governed, not by rules, but by analogy. This reflects the way in which any such expression is acquired – we encounter examples of a word, and eventually we begin to apply the word ourselves, extending it into new contexts, based on the perception of an analogical resemblance between the new instance, and previous instances of the expression we have experienced. As I indicated earlier – the analogical account of meaning is universal in its scope – thus any word that has a use in a natural language, is regarded as having a meaning in that language, and such meaning is said to be governed by analogy. However – significantly for the account that I am proposing, even the meaning of

---

<sup>32</sup> Kaplan(2004), <http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf> ,p.14

<sup>33</sup> Kaplan(2004), <http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf> ,p.11

the key terms in the discussion – ‘analogy’ and ‘rule’, are said to be governed by analogy, rather than rules. Thus, there is not one thing that all ‘analogies’ or all ‘rules’ have in common, but just a series of resemblances between particular instances. I will return to this issue later – for if the words ‘analogy’ and ‘rule’ are governed by analogy, rather than by rules, then this makes a significant difference to what exactly is being proposed. Every word that has a meaning can be understood as standing for a category, containing the kinds of objects, actions, and situations, that the word can be appropriately applied to. However, given that I have argued that those things that a word can be appropriately applied to cannot be captured in terms of a rule, this means that the categories corresponding to our words cannot be defined in terms of precise membership criteria, contrary to the classical account of categories. Thus, in the next section, I will propose an alternative account of categories, that can both work alongside, and provide support for, the analogical account of meaning.

## Chapter 2 – Prototype Theory

### Introduction

In the previous section, I argued that for many of our word-types, it is, at least, extremely difficult to state some rule specifying precisely what objects, actions or situations words of each type can be meaningfully applied to. For any candidate rule, there are likely to be counterexamples (objects that the word can be meaningfully applied to, but that don't conform with the rule, or objects that the word can't be meaningfully applied to, that do conform with the rule) and borderline cases (objects for which it is unclear whether or not they conform with the rule). Thus, if each meaningful expression is understood to stand for a category, containing the objects, actions and situations it can be meaningfully applied to, then it seems that the categories corresponding to many of our words do not have sharp boundaries (or if they do have sharp boundaries – we have an extremely hard time finding them). On the classical account of categories, the intension of a category (the membership criteria) is supposed to precisely and unambiguously determine the extension of the category (the objects that belong to the category) and so determine a sharp boundary between the category's extension and anti-extension. However, while I noted it was hard to specify one thing shared by every member of a category (i.e. one set of necessary and jointly sufficient conditions for category membership), I pointed out that there are undeniable similarities between category-members, such that we might assign something to a category due to its perceived analogical resemblance to other category-members, rather than by applying a rule. If we assign things to categories in this way, we should not expect there to be a sharp boundary between things that do and do not belong to the category. If categories are thought of like this, then category-membership is not defined by each member of the category having one thing in common (i.e. by formal membership criteria), but by each member of the category having something in common with (some) other members of the category. Thus, if meaning something by a linguistic expression consists in making an analogy, rather than in following a rule, then a different account of categories is needed. Eleanor Rosch has proposed one such

account, known as prototype theory.

In this section, I will first outline the Wittgensteinian roots of Rosch's theory, focussing on Wittgenstein's notion of family resemblance, and idea that our categories lack sharp boundaries. I will then explore Rosch's claim that we should, thus, conceive of categories, not in terms of their boundaries, but in terms of their clear cases – their most prototypical instances. I will outline Rosch and Mervis's hypothesis that there will be a correlation between the degree to which an object is judged to be a prototypical member of a category, and that object's family resemblance to other objects, in both the same and other categories. I'll then detail the experiments they carried out in order to test their hypothesis, for a variety of categories of concrete object, and for artificial categories. I will find the experimental results to offer compelling evidence for their hypotheses, and will agree with Rosch and Mervis, that the results also help to explain why it is that we tend to believe that categories must be sharply defined in terms of criterial attributes, in spite of the contrary evidence. I will then outline Rosch's explanation of why our categories would have a prototype structure, focussing on her claim that there are two fundamental psychological principles that govern the nature of human category-systems. I will take this to offer a good explanation of why our categories would be this way. I will then claim that, while the experiments focussed on categories corresponding to concrete nouns, there may be good reason to believe that the categories corresponding to many of our other word-types could have a prototype structure as well. I will focus on adjectives, and in particular, colour categories.

Having explained how Rosch's prototype theory might be extended to other categories, I will then look at whether there's any way of interpreting the experimental data that is consistent with a classical account of categories. I will claim that it is possible to claim that in spite of the evidence, our categories do have sharp boundaries, and categories could be defined by membership criteria, but that the only viable way of doing this would be to adopt an epistemicist account of meaning. Thus, I will argue that there are two ways of interpreting the evidence – either as providing support for the analogical account of

meaning, by demonstrating that categories are defined by family resemblance relationships, rather than membership criteria, or as providing support for an epistemicist account, on which our categories are sharply defined by membership criteria, albeit in an extremely complicated way, which captures the family resemblance relationships noted by the experiments. I will claim that we should favour an analogical account of meaning, over an epistemicist account, and that a picture on which categories have a prototype structure works well alongside the analogical account of meaning that I am advocating.

### Family Resemblances and Prototypicality

Wittgenstein points out that when we consider all of the different things which we call ‘games’, we will see that there is not something common to all of them, but just “a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail”<sup>34</sup>. He characterises these similarities as “family resemblances”, indicating that just as the resemblance between the members of family cannot be captured in terms of necessary and jointly sufficient conditions, but only in a number of particular overlapping features, such as “build, features, colour of eyes, gait, temperament, etc.”, so too this may be the case for the resemblance between particular instances of the category ‘game’.<sup>35</sup> He points out that some but not all games involve winning and losing, some involve luck, others involve skill, some involve a combination of the two, some involve cards, others involve balls – and so on, but that there is not one thing that all games have in common.<sup>36</sup> Thus, on Wittgenstein’s account, the instances of an expression like ‘game’ need not all share one thing in common, which can be captured in terms of criterial attributes, in order for the word to be used meaningfully. Nor should we expect a word, like ‘game’, by virtue of having a meaning in our language, to draw a sharp boundary between those things to which it does and does not apply.

---

<sup>34</sup> Wittgenstein(1996)p.32

<sup>35</sup> Wittgenstein(1996)p.32

<sup>36</sup> Wittgenstein(1996)p.31-32

Wittgenstein claims that the concept of game is not bounded.<sup>37</sup> He points out that we teach someone the meaning of the word ‘game’, roughly, by showing them particular examples of games, and saying “This *and similar things* are called games”<sup>38</sup>, rather than by teaching them anything of the boundary between games and non-games. He claims that our knowledge of what a ‘game’ is, is “completely expressed” in the explanations we can provide – namely by “describing examples of various kinds of game; shewing how all sorts of other games can be constructed on the analogy of these; saying that I should scarcely include this or this among games; and so on.”<sup>39</sup> While Frege had claimed that concepts must have sharp boundaries, Wittgenstein claims that the idea of a concept, like game, having blurred edges, should not trouble us, because when describing our use of language an “indistinct picture[is] ...often exactly what we need”.<sup>40</sup> Wittgenstein indicates that we can use a word meaningfully, without being able to draw a sharp boundary between instances and non-instances. We exhibit our understanding of a concept, like ‘game’, not by being able to stipulate a boundary between games and non-games, but by being able to provide and identify examples of games (and of non-games). Thus our understanding of the concept ‘game’ is shown by our ability to identify clear-cases of the concept, rather than its boundaries.

Eleanor Rosch has built upon these ideas of Wittgenstein’s, to propose an alternative to defining categories in terms of criterial attributes. The key idea she (along with co-author, Carolyn B. Mervis) gets from Wittgenstein is the idea that the referents of a word (such as the word ‘game’) need not all share one thing in common in order for that word to be used meaningfully.<sup>41</sup> Rather, categories (such as the category corresponding to the word ‘game’) could be structured based on the family resemblance relationships obtaining between particular instances. Rosch and Mervis write that, in a category structured by such family resemblance relationships, “each item has at least one, and probably several, elements in

---

<sup>37</sup> Wittgenstein(1996)p.33

<sup>38</sup> Wittgenstein(1996)p.33

<sup>39</sup> Wittgenstein(1996)p.35

<sup>40</sup> Wittgenstein(1996)p.34

<sup>41</sup> Rosch&Mervis(1975)p.574

common with one or more other items, but no, or few, elements are common to all items".<sup>42</sup> When categories are defined by such family resemblance relationships, they do not have sharp boundaries, since membership in the category is not defined by the possession of a simple set of criterial attributes. Rosch cites experimental evidence in support of the idea that our categories lack sharp boundaries.<sup>43</sup>

Experimental subjects tend to disagree, or even change their minds, when asked certain questions of the form 'Is an X a Y?'. For example, as Hofstadter and Sander report, subjects tend to disagree over whether a *wig* is an *item of clothing*, or whether *Big Ben* is a *clock*, and an individual subject may even change their mind when asked the same question a few days apart, as, reportedly, in the case of the question 'Is a pillow an article of furniture?'<sup>44</sup> It thus seems that many of our categories have borderline cases – objects for which it's unclear whether or not they belong to that category, and such categories therefore appear to lack sharp boundaries.

Rosch takes Wittgenstein's central insight to be that we should conceive of categories, not in terms of their boundaries, but in terms of their clear cases.<sup>45</sup> She claims that, contrary to the classical account, category-membership could be a matter of degree. She proposes that categories can have stronger and weaker members, ranging from the most prototypical members of the category – the best examples of the category, to the least prototypical members - the poorest examples. Rosch cites psychological evidence that subjects tend to overwhelmingly agree when it comes to questions about how good an example of a category a given object is.<sup>46</sup> For example, people tend to agree that a table is a better example of an item belonging to the category *furniture*, than a radio is, and that a sparrow is a better example of an item belonging to the category *bird*, than a penguin is. Rosch thinks that these judgments of prototypicality (of how good an example of a category a given item is), and their uniformity across experimental subjects, reveal something important about the structure of these categories. Specifically, she (along with co-author

---

<sup>42</sup> Rosch&Mervis(1975)p.575

<sup>43</sup> Rosch(1999),p.197

<sup>44</sup> Hofstadter&Sander(2013)p.57

<sup>45</sup> Rosch(1999),p.196

<sup>46</sup> Rosch(1999),p.197

Carolyn B. Mervis) thinks that those items judged to be the most prototypical belong most fully to the category, while those items judged to be less prototypical belong to the category to a lesser degree.<sup>47</sup> She thinks that this prototype structure will not arise by coincidence, and hypothesises that it may be importantly connected to Wittgenstein's notion of family resemblance. I will now outline Rosch and Mervis's specific hypotheses, and the experiments they carried out in order to test them.

Rosch and Mervis propose that there may be a correlation between the prototypicality rating of a given item within a category, and that item's degree of family resemblance to both members of the same category, and to members of other, contrasting, categories.<sup>48</sup> The degree of family resemblance between two items is understood to be the degree to which those items share attributes in common. They propose that the most prototypical items within a category will be those that possess the highest degree of family resemblance to (or share the most attributes in common with) other members of the category. Secondly, she proposes that the most prototypical items within a category will be those which possess the lowest degree of family resemblance to members of other, contrasting, categories.<sup>49</sup> Thus – the most prototypical members of a category will be those items that are most representative of the category, and least representative of other categories, to which the items do not belong. Eleanor Rosch and Carolyn B. Mervis conducted a series of experiments to test these hypotheses. They conducted these experiments using superordinate categories (very general categories of object – such as *fruit*, *animal*, and *furniture*, more abstract than the 'basic level' that experimental subjects tend to categorise objects at), basic-level categories (categories like *banana*, and *dog*, and *table*, the level of abstraction at which most of our day-to-day categorisations tend to be made), and for artificial categories (categories invented by the experimenters, composed of strings of letters, e.g. *JXPHM*). I will now outline these experiments in turn.

---

<sup>47</sup> Rosch&Mervis(1975)p.574

<sup>48</sup> Rosch&Mervis(1975)p.575

<sup>49</sup> Rosch&Mervis(1975)p.575

### Experimental evidence

The first experiment concerned superordinate categories. Rosch and Mervis note that these are of particular interest, because there seem to be few, if any attributes common to the members of many superordinate categories, meaning they may be structured entirely based on the principle of family resemblance.<sup>50</sup> The superordinate categories used in this experiment were the following: *furniture*, *vehicle*, *fruit*, *weapon*, *vegetable*, and *clothing*.<sup>51</sup> In this experiment, subjects were presented with members of these categories (for example, for the category *furniture*, they may be presented with category members like table, chair and radio) – and they were asked to list the attributes that these category-members possessed. This was done for 20 objects belonging to each superordinate category, and each of the attributes listed for some of the objects was weighted according to how many of the 20 objects it was true of.<sup>52</sup> Thus, the most highly weighted attribute was that which was shared by the most objects, and the least highly weighted, was that which was shared by the fewest. Those objects which had the most highly weighted collection of common attributes were considered to be those which possessed the highest degree of family resemblance to other category members. It turned out those objects that possessed the highest degree of family resemblance, were just those objects that are regarded as the most prototypical members of the category by experimental subjects, thereby establishing a correlation between judgments of prototypicality and degree of family resemblance.<sup>53</sup>

Rosch and Mervis conducted a second experiment concerning superordinate categories, in order to test the second part of the hypothesis – that the most prototypical items within a category will be those which share the least family resemblance to members of other, contrasting, categories. They note that it was only possible to test this hypothesis indirectly, for superordinate categories, by measuring the extent to which items judged to be the most prototypical members of one category, were judged to be members of other

---

<sup>50</sup> Rosch&Mervis(1975)p.576

<sup>51</sup> Rosch&Mervis(1975)p.582

<sup>52</sup> Rosch&Mervis(1975)p.578-580

<sup>53</sup> Rosch&Mervis(1975)p.582

superordinate categories, as well.<sup>54</sup> Subjects were presented with nouns, and made to list three categories to which each noun belonged. The categories for each noun were weighted, based on where they appeared in the list, and the category-dominance of each category over each noun was calculated, by subtracting the weightings of a noun's dominance in other categories, from the weighting of its membership in any given category.<sup>55</sup> The results confirmed the hypothesis – the most prototypical items were those which were the most dominant members of those categories, those items least likely to be regarded as belonging to another category, instead.

Rosch and Mervis conducted a similar pair of experiments concerning basic-level objects. They note that basic-level objects are of particular interest, because they are the level of abstraction at which most of our day-to-day categorisations tend to be made.<sup>56</sup> The categories studied were: *car, truck, aeroplane, chair, table, lamp*.<sup>57</sup> Subjects were presented with pictures representing members of basic-level categories (so, for example, for the basic-level category 'chair', they would be presented with a series of pictures of chairs), and they were asked to make a list of the attributes possessed by each item with which they were presented. Another group was asked to rank the objects with which the subjects were being presented in terms of their degree of prototypicality – that is, of how good an example of the category at issue each item was.<sup>58</sup> While the results of this experiment contrasted somewhat with the experiment concerning superordinate categories, in that there were more properties that were shared in common by all members of any given category, the results once again confirmed Rosch and Mervis's hypothesis, in that a strong correlation was established between the subject's judgments of degree of prototypicality, and the number of attributes each item shared with other items in the same category.<sup>59</sup> Just as predicted, the most prototypical members of each category were those items which shared the most properties in common

---

<sup>54</sup> Rosch&Mervis(1975)p.585

<sup>55</sup> Rosch&Mervis(1975)p.585

<sup>56</sup> Rosch&Mervis(1975)p.586

<sup>57</sup> Rosch&Mervis(1975)p.588

<sup>58</sup> Rosch&Mervis(1975)p.588

<sup>59</sup> Rosch&Mervis(1975)p.589

with other members of that category.

Another experiment was conducted to test the second part of their hypothesis, which states that the most prototypical members of a category will be those items which share the fewest properties in common with members of other categories. Subjects were once again made to list attributes possessed by members of various contrasting categories, and the number of attributes shared in common between the members of these different categories was calculated.<sup>60</sup> Once again, a strong correlation was established – only this time, a negative one, between the degree of prototypicality of an item belonging to a given category, and the number of attributes it shared in common with members of other categories.<sup>61</sup> This experiment confirmed, that just as in the case of superordinate categories, so too is it the case for basic-level categories, that the most prototypical members of each category are those that possess the highest degree of family resemblance to other members of that category, and the lowest degree of family resemblance to members of other categories.

Having conducted the experiments concerning the correlation between family resemblance and prototypicality ratings for concrete nouns at both the superordinate and the basic-level of abstraction, Rosch and Mervis conducted a similar pair of experiments using artificial categories – strings of letters, invented by the experimenters, such as *JXPHM*.<sup>62</sup> The purpose of these experiments was to see if the same principles governed the formation of categories in novel domains of categorisation, as well as in the categories already established in natural language. For the first experiment, subjects were presented with a number of categories, containing letter strings, which possessed a varying degree of family resemblance to each other, in terms of the number of overlapping attributes they possessed (e.g. ‘contains the letter ‘J’’).<sup>63</sup> They hypothesised that those items possessing the highest degree of family resemblance within a

---

<sup>60</sup> Rosch&Mervis(1975)p.591

<sup>61</sup> Rosch&Mervis(1975)p.591

<sup>62</sup> Rosch&Mervis(1975)p.593

<sup>63</sup> Rosch&Mervis(1975)p.592

category, would be judged to be the most prototypical members of that category, would be the first category-members that were learned, and would take the smallest amount of time for people to judge to belong to the category.<sup>64</sup>

Subjects were presented with two sets of letter categories at a time (with no family resemblance between the two categories), and made to sort them into the two categories. They were corrected if they made an error, until they had correctly sorted the letter strings into two categories.<sup>65</sup> The rate at which they learnt to categorise each item was measured by the number of errors they made. The results of the experiment confirmed that those items which possess the highest degree of family resemblance, are judged to be the most prototypical, are the category members that are learnt most quickly, and are the items that are categorised most rapidly.<sup>66</sup> A second experiment confirmed that, when contrasting categories are involved, the higher the degree to which an item possesses family resemblance to members of other categories, the less prototypical it is judged to be, the longer it takes subjects to learn such category-members, and the less rapidly such items are categorised.<sup>67</sup> These experiments thus suggest that the most prototypical members of a category play an important role in how we learn to categorise things – they are the first category members we tend to learn, and are identified as category-members more rapidly. Rosch also cites evidence that when it comes to learning the names of objects, children tend to learn to categorise the most prototypical members of a category first.<sup>68</sup> Given the correlation between prototypicality and degree of family resemblance, this makes sense – children first learn the most representative members of a category, before learning to categorise other objects that do not share so many of the attributes that are distinctive of the category as a whole. These experiments suggests that prototypes may play an important role, both in the formation of categories and the learning of categories, as well as in the structure of many of the natural language categories existing in our culture today.

---

<sup>64</sup> Rosch&Mervis(1975)p.591

<sup>65</sup> Rosch&Mervis(1975)p.594

<sup>66</sup> Rosch&Mervis(1975)p.595-596

<sup>67</sup> Rosch&Mervis(1975)p.598

<sup>68</sup> Rosch(1999),p.198

As well as providing evidence for the prototype structure of categories, Rosch and Mervis's experiments also offer us some explanation of why it is that people tend to assume that there is some sharply defined, criterial feature, shared by each and every member of a category, rather than simply a criss-crossing network of similarities. Rosch cites evidence that when we are presented with the name of a category, we tend to think of the most prototypical members of that category, rather than the least prototypical ones.<sup>69</sup> The most prototypical members of a category are those that have the most attributes in common with other members of that category – and thus the most prototypical members of a category do tend to have a lot of attributes in common with each other. Rosch compared the number of attributes common to the five most, and the five least, prototypical items in each category, and found that the five most prototypical items in a category did tend to share many attributes in common, unlike the five least prototypical items. For example, Rosch and Mervis notes that the 5 most typical members of the category *furniture* shared 13 attributes in common, while the 5 least typical members shared only 2 properties in common, and the 5 most typical members of the category *clothing* shared 21 properties in common, while the 5 least typical members shared no properties in common at all.<sup>70</sup> Thus when we think about what it is that the members of a category have in common, our tendency to focus our attention on the most prototypical members of a category misleads us into thinking that there must be some set of necessary and jointly sufficient conditions that defines membership in that category. Were we to think also of the much less prototypical members of the category, the idea that there is some formal principle that we can isolate that defines category membership in a binary, all-or-nothing way, would be much less likely to take hold. Rosch and Mervis note that even after receiving feedback from the experiment, subjects tended to insist that there must be some properties common to all instances of a category, and take this to be down to their focus on the most prototypical instances.<sup>71</sup>

These experiments show, for a variety of categories of concrete object, that people judge an object to

---

<sup>69</sup> Rosch(1999),p.199

<sup>70</sup> Rosch&Mervis(1975)p.582

<sup>71</sup> Rosch&Mervis(1975)p.582

be a more prototypical member of a category, when that object has more properties in common with other members of the category, and fewer properties in common with members of other categories. Rosch claims that these experiments show that “it is a fact that both representativeness within a category and distinctiveness from contrast categories are correlated with prototypicality in real categories.”<sup>72</sup> She proposes that this reveals an important fact about the structure of our categories, and suggests that the reason why our categories are structured this way can be explained by considering the two fundamental principles that drive our categorisations.

### Two Fundamental Psychological Principles

Rosch claims that the study of categorisation deals with “the fact that unique particular objects or events can be treated equivalently”.<sup>73</sup> She claims that the nature of the categories existing within a culture at any particular point in time, should not be considered an “accident of history”, but, rather, can be explained by considering the two fundamental psychological principles that govern human categorisation.<sup>74</sup> The first of these principles is the principle of cognitive economy. The principle of cognitive economy states that we want our categories to provide us with as much information as possible, while conserving finite resources as much as possible.<sup>75</sup> Thus, we want to be able to understand and predict as much as we can about something, by virtue of placing it into one of our categories. However, as Rosch indicates, while a greater number of finely-distinguished categories may provide us with more information about that which we categorise, we also don’t want to have too many, too finely distinguished categories, since this would make categorisation too difficult. She indicates that there are infinite differences amongst stimuli, and that in order to for our categories to be suitably user-friendly, we must “reduce the infinite differences amongst stimuli to behaviourally and cognitively usable

---

<sup>72</sup> Rosch(1999),p.198

<sup>73</sup> Rosch(1987)p.151

<sup>74</sup> Rosch(1999),p.189

<sup>75</sup> Rosch(1999),p.190

proportions”.<sup>76</sup> For example, it is useful to us, for various purposes, to categorise people based on their age in years – while we could categorise people by their age in seconds, instead, such categorisations would not be behaviourally or cognitive useful. Thus, the principle of cognitive economy drives us to form categories that provide us with as much information as possible, while ignoring differences between stimuli that are irrelevant to our purposes. As Rosch indicates, by placing something into a category, we are simultaneously excluding it from various other categories, and the principle of cognitive economy drives us to make our categories as easily discriminable as possible, to aid in the task of categorisation.<sup>77</sup>

The second fundamental psychological principle that Rosch takes to govern the nature of human category-systems, is the principle of perceived world structure. According to this principle, the perceived world “comes as structured information rather than as arbitrary or unpredictable attributes”.<sup>78</sup> Thus, that which we categorise – the perceived world – possesses a highly correlational structure, and we want our categorisations to reflect this correlational structure as much as possible. By virtue of having a highly correlational structure, this means that the attributes that we perceive in the world do not occur uniformly, but rather some attributes are more likely to occur with certain attributes than they are with others. For example, as Rosch writes, something that possesses the attribute of “having wings” is more likely to possess the attribute of “having feathers” than it is to possess the attribute of “having fur”.<sup>79</sup> Therefore, there really are combinations of attributes in the perceived world, and we want our category-systems to be able to track these clusters. However, Rosch notes that the attributes that we do perceive, may not in fact be “inherent in the real world”, but may be “constructs of the perceiver”<sup>80</sup>, insofar as they depend, in order to be perceived, on the existence of the category system in place. For example, Rosch notes that one of the attributes we perceive chairs to possess, is the attribute of ‘being a seat’, yet if we did not have a cultural and linguistic category called ‘chairs’, the attribute of ‘being a seat’ would not have meaning for

---

<sup>76</sup> Rosch(1999),p.190

<sup>77</sup> Rosch(1999),p.190

<sup>78</sup> Rosch(1999),p.190

<sup>79</sup> Rosch(1999),p.191

<sup>80</sup> Rosch(1999),p.191

us.<sup>81</sup> Nonetheless, even if at least some of the attributes that we perceive are constructs of the perceiver, Rosch notes that it is still a fact that these attributes do not occur uniformly, and thus our categorisations are constrained by the principle of perceived world structure.<sup>82</sup>

Rosch proposes that these two psychological principles lead us to form categories with a prototype structure, based on the family resemblance relationships we perceive between stimuli. Rosch proposes that in order to meet the principle of cognitive economy, and in particular, to increase the “distinctiveness and flexibility” of our categories, our categories tend to be defined and structured around their most prototypical instances.<sup>83</sup> As Rosch and Mervis’s experiments showed, the most prototypical instances in a category (or at least in the categories of concrete object studied) are those members of the category that share the most properties in common with other members of the category, and the least properties in common with members of other categories. Therefore, the most prototypical items of a category are those which are most representative of the category, and least representative of other categories. By structuring our categories around these most prototypical instances, therefore, we increase the distinctiveness of our categories, since the most prototypical members of a category are those objects that are the most distinctive members of that category. We also increase the flexibility of our categories, since objects are categorised because of the perception of family resemblance to other category-members (and especially the most prototypical ones, since these share the most properties in common with other category members), rather than by the rigid application of some rule. Categories structured around a prototype do not have sharp boundaries, since there is no explicit restriction on the manner in which, and the degree to which something must share family resemblances with members of a category in order to be regarded as a member of that category.

By having categories with a prototype structure, we are also able to meet the principle of perceived

---

<sup>81</sup> Rosch(1999),p.201

<sup>82</sup> Rosch(1999),p.191

<sup>83</sup> Rosch(1999),p.191

world structure. This is because the most prototypical members of a category, are, by definition, those category members that possess the highest degree of correlational structure, by sharing the most attributes in common with other members of the category, and least attributes in common with members of other categories. Thus, our categories as a whole, reflect clusters of perceived attributes, and the most prototypical category members are those which best reflect these clusters, by virtue of possessing the highest degree of properties that are also common to other category members. Rosch notes that the clusters of perceived attributes, which we want our categories to be able to track, are “not ... necessarily discontinuous”.<sup>84</sup> Thus these clusters of attributes, and the categories we form to represent them, may not have sharp boundaries. Rosch claims that by representing categories in terms of their most structured portions – their most prototypical instances, we are able to form categories that are as separate and clear-cut as possible from each other, even in the absence of a sharp boundary between instances and non-instances.<sup>85</sup> Rosch takes these two fundamental psychological principles to explain how it is that we came to have the categories that we do have. The practical constraint of the principle of cognitive economy leads us to form categories that are easily distinguishable from each other, while providing us with as much information as possible, and the real-world constraint of the principle of perceived world structure leads us to form categories that track clusters of perceived attributes, even if these clusters may not be discontinuous. Rosch claims that forming categories with a prototype structure, based on the perception of family resemblance, is one viable way in which we could meet these constraints.<sup>86</sup>

#### From nouns to other word types

Rosch’s experiments looked at categories of concrete objects, both at the superordinate level, and at the basic-level of abstraction. Thus the results suggest that the categories denoted by (concrete) nouns have a prototype structure. This poses a challenge to the traditional account of meaning, which also tends to focus on nouns, according to which, associated with each noun is a set of necessary and jointly

---

<sup>84</sup> Rosch(1999),p.196

<sup>85</sup> Rosch(1999),p.196

<sup>86</sup> Rosch(1999),p.196

sufficient conditions which determines precisely which objects fall under the extension of the noun, and which ones don't. However, as I indicated above, I think that a common explanation can be given, in terms of categorisation being driven by analogy, for the meanings of many word types, including, though not limited to, nouns. While Rosch's experiments are limited to nouns (which is also the favoured example of the traditional account of meaning, which she is challenging), I think that there are good reasons to believe that many of our word-types have a prototype structure. The case can be most clearly made for adjectives, which I will focus on. However, in the previous section I noted that the categories corresponding to many of our word-types seem to lack sharp boundaries – even the categories corresponding to conjunctions and articles, greetings, and expressions of gratitude, and this could be because all that instances of each category have in common is family resemblances between particular instances. It is possible that even the categories corresponding to such words have more and less prototypical instances, given that in each case there are clear cases in which we should use one expression rather than another, but no sharp boundary between categories. In general, if a category lacks boundaries, but contains clear cases, this may be reason to believe that such a category has a prototype structure, based on the principle of family resemblance.

As Rosch indicates, the domain in which the prototype structure of categories has been most clearly demonstrated, is in fact that of colour.<sup>87</sup> She cites evidence that people's judgements about the boundaries of colour categories tend to disagree, or to even be inconsistent.<sup>88</sup> Thus someone may categorise a particular shade of colour as 'red' in one context, and 'orange' in another context. She notes that the members of a given colour category do not all share one attribute in common, seeing as "the physical properties of light, such as wavelength, vary continuously".<sup>89</sup> However, Rosch proposes that there are clearly more and less prototypical instances of a colour, such as blue – and that the most prototypical

---

<sup>87</sup> Rosch(1987)p.154

<sup>88</sup> Rosch(1987)p.154

<sup>89</sup> Rosch(1987)p.154

instances of a colour category occur at “physiologically salient points in the colour space”.<sup>90</sup> Structuring our colour categories around such prototypes allows us to place things in colour categories, based on their resemblance to these prototypes, in spite of the fact that the physical properties underlying our perception of colour are not discontinuous, such that our colour categories lack sharp boundaries. It seems that many other adjectives too may have such a prototype structure. Thus, while there is no sharp boundary between people who are ‘tall’ and people who are ‘not tall’ (stated, say, in centimetres), there are clearly prototypical instances of ‘tall’ people, and prototypical instances of ‘non-tall’ people, with a whole range of more and less prototypical instances lying inbetween. While people would tend to agree on clear cases of ‘tall’ and ‘non-tall’ people, we should not expect there to be agreement on a precise boundary, thus such categories too may have a prototype structure. Thus, it is not only concrete nouns, but some adjectives too, that have been shown to have a prototype structure. If both the names for concrete objects, and the names for the properties of concrete objects have a prototype structure, then this suggests that prototype structure is pervasive in our category system.

### Intepreting the Evidence

Rosch and Mervis’s experiments showed, that for several categories of concrete object, both at the superordinate and the basic-level of categorisation, there is a positive correlation between the degree to which the members of that category share properties in common with each other, and the degree to which items are regarded as prototypical members of that category. The experiments also showed that there is a negative correlation between the prototypicality of an item in one category, and its family resemblance to members of other categories. Rosch takes this to indicate that membership in such categories is defined by the presence of family resemblance relationships between category-members, and that membership in such categories is a matter of degree, with the most prototypical items belonging most fully to the category, and the least prototypical items belonging less fully to the category. Another consequence of

---

<sup>90</sup> Rosch(1987)p.154

this account is that categories do not have sharp boundaries - which Rosch took to be demonstrated by the experimental evidence that people's judgements about boundaries tend to disagree. However, some may argue that the experimental evidence produced by Rosch and Mervis, do not prove that the categories at issue have a prototype structure. It is possible to claim that the categories at issue could still be defined by membership criteria, and have sharp boundaries, despite the experimental evidence. I will now examine this possibility.

It's possible for someone advocating a classical account of categories; in which category-membership is an all-or-nothing matter, governed by the application of formal membership criteria, which draws a sharp boundary between instances and non-instances, to argue that the experimental evidence does not contradict such an account. Rosch took the fact that experimental subjects tend to disagree about the boundaries of categories to indicate that such categories do not have sharp boundaries. However, a proponent of a classical account could claim that all that this evidence shows is that we are ignorant of the boundaries of our categories, not that our categories do not have sharp boundaries. Rosch took the fact that experimental subjects tend to agree in their prototypicality judgements, to indicate that category-membership is a matter of degree. However, someone defending a classical account of categories could simply claim that people's prototypicality judgements have nothing to do with the extent to which objects belong to a category, but simply reflect the strength of psychological association between a category and its particular members. They could claim that just because chairs are regarded as more prototypical members of the category *furniture* than radios are, both chairs and radios belong fully and equally to the category *furniture*. However, while its possible for a proponent of a classical account of categories to interpret the evidence concerning people's boundary-judgements and prototypicality judgements in this way, it seems that Rosch and Mervis's experiments demonstrate something that cannot be dismissed so easily. The experiments demonstrate that there is no one thing that all members of the categories studied share in common, but that there are family resemblances between particular instances. If we accept that such categories really are defined by such family resemblance relationships, then it's hard to see how

such categories could have formal membership criteria, as the classical account insists they must. I will now look at how a proponent of a classical account may interpret the experimental results.

As I indicated, it's possible for a proponent of the classical account of categories to claim that the fact that people's boundary-judgements disagree, reflects our ignorance about the boundaries of our categories, not that our categories lack boundaries. Someone could claim that our categories are, in fact, defined by formal membership criteria which determine a sharp boundary between instances and non-instances, but that we are ignorant of the boundary, because we either do not know these formal membership criteria, or we apply them imperfectly in our categorisations. It seems that if our categories do have formal membership criteria, but that we don't know them, or are at least not able to apply them straightforwardly – then such membership criteria must be of considerable complexity. Presumably, if the categories corresponding to many of our words do have sharp boundaries, but we are ignorant of these boundaries, this must be because the boundaries of our categories, and the formal membership criteria that govern these boundaries, were determined by the manner in which that word was used over a long period of time, amongst the members of a given linguistic community. It's hard to see how else our categories could have established sharp boundaries, in spite of our ignorance of their whereabouts.

Timothy Williamson has proposed one such account, known as epistemicism, according to which the meaning of a linguistic expressions supervenes on (is determined by) the way that expression is used, but in an “unsurveyably chaotic way”<sup>91</sup>, such that we should not expect to know exactly what some of our words mean. Williamson gives the example of the adjective ‘thin’ (in application to a person) – claiming that the way that that expression has been used in our linguistic community determines a meaning for that expression, which is precise, and draws a sharp boundary between instances and non-instances. Thus, he suggests, the sentence ‘Timothy Williamson is thin’ must have a determinate truth value, even if most of

---

<sup>91</sup> Williamson(2001)p.209

us would regard Williamson as a borderline case of thinness.<sup>92</sup> Analogously, it could be claimed that the sentence “a pillow is an article of furniture” has a determinate truth value, in spite of people’s judgements disagreeing, because the way that the word ‘furniture’ has been used has determined a sharp boundary for the category. It seems that if we want to insist that our categories do have sharp boundaries, in spite of our ignorance of them, then something like Williamson’s epistemicist account may be a viable option. Thus a classical account of categories can be upheld in spite of the evidence about people judgments about the boundaries of categories disagreeing. However, the experimental evidence concerning prototypicality and family resemblances, is harder to dismiss.

I noted that a classical account of meaning can simply claim that people’s prototypicality judgements have nothing to do with the meaning of a word, or the degree to which items belong to the category corresponding to the word. It could be claimed that the most prototypical items in a category are simply those we happen to psychologically associate most strongly with the category-name, but that in reality all category-members, be they more or less prototypical, belong fully and equally to the category. In some cases, such a reaction might seem appropriate. For example – Rosch states that a robin is judged to be a more prototypical bird than a penguin<sup>93</sup>, but we may wish to insist that both penguins and robins are wholly birds, even if robins happen be more typical examples than penguins. Though Rosch notes that while we can say that a penguin is “technically a bird”, we can’t say the same thing about a robin, since robins are “more than just technically a bird ... it is a bird par excellence”.<sup>94</sup> However, while it’s possible to dispute the claim that the prototypicality ratings determine an item’s degree of membership in a category, Rosch and Mervis were able to provide an analysis of prototypicality in terms of degree of family resemblance, which is rather harder to dismiss. Rosch and Mervis demonstrated, that for the categories studied, there was no one thing that all category members had in common, but that there were family resemblances between particular instances, and the degree to which an item possessed family

---

<sup>92</sup> Williamson(2001)p.196

<sup>93</sup> Rosch(1999)p.199

<sup>94</sup> Rosch(1999)p.199

resemblance to other items in the category correlated with its prototypicality rating. Thus, it seems that whether or not an item's prototypicality rating determines its degree of membership in the category, it does correlate with that item's representativeness within that category, spelled out in terms of degree of family resemblance. However, a proponent of a classical account may still insist that such categories are defined by criterial attributes.

For the categories studied, it seems there really is no one thing that all category-members have in common. It also seems clear that there are family resemblances between particular instances. As long as we accept that the kind of attributes listed by subjects in the experiments, are the kind of attributes that do define membership in each given category, then it seems that the only way to uphold a classical account of meaning, in light of the evidence, is to claim that the membership criteria for our categories are extremely complicated. It seems as though if all that the members of the categories studied have in common is this network of family resemblance relationships, then the membership criteria for the category, if there are any, must state that items must possess a certain critical sum of a cluster of attributes, shared by some but not all category members, in order to belong to a category. If we insist, that, in spite of the evidence, there must be something common to the members of each category, all we are seemingly left with is a series of family resemblances between particular instances. As Wittgenstein suggested, having listed some attributes common to some but not all *games*, such an insistence can seem artificial: "But if someone wished to say: "There is something common to all these constructions – namely the disjunction of all their common properties" – I should reply: Now you are only playing with words. One might as well say: "Something runs through the whole thread – namely the continuous overlapping of those fibres"."<sup>95</sup>

I have suggested that while a proponent of a classical account of meaning can dismiss the fact that people disagree about boundaries, by adopting an epistemicist account, it is much harder for such an

---

<sup>95</sup> Wittgenstein(1996)p.32

account to dismiss the evidence that many of our categories seem to be defined by family resemblance relationships. While it is possible for such an account to claim that categories are sharply defined, in terms of these family resemblance relationships, with each item belonging to the category having to possess a certain combination of these attributes, common to some but not all category members, such a claim seems extremely artificial. We're left with a picture on which the meaning of at least many of our expressions consists in some precise matrix of properties, some of which must be present in some particular combination, in order for an expression to be meaningfully applied to some object. Classical accounts of meaning have tended to insist that the formal criteria that define category-membership are simple, and thus easily applicable for language users, whereas on this account, the formal criteria for category-membership ends up being extremely complicated. Such accounts must claim that our categories do have sharp boundaries, and that these are determined by membership criteria stating some precise matrix of properties, some of which must be possessed in a particular combination in order for an object to belong to the category, with the nature of this matrix of properties, and the ways in which it can be instantiated, determined by the chaotic patterns in which the word denoting the category has been used over time.

In light of the experimental evidence, and in particular the demonstration that many of our categories are defined by a series of family resemblance relationships, I think that we are left with two viable options for a theory of meaning. We can adopt an epistemicist account of meaning of the kind just outlined, in which our categories do have sharp boundaries, and the family-resemblance relationships obtaining between particular category-members is captured in terms of a rule. Alternatively, as I advocate, we can adopt an analogical account of meaning – on which meaning something by an expression requires the user of that expression to draw an analogy between the present instance, and previous objects that have been designated by the expression. On this account, we assign something to a category, because we perceive an analogy between it and other objects we have assigned to that category – and the nature of the analogies that underlie our categorisations can be understood in terms of Rosch's notion of family resemblance. I

will now outline how the analogical account of meaning can work alongside a prototype theory of categories, such as Rosch's. I will then compare the two theories I have identified as viable responses to the experimental data – the epistemicist account of meaning, and the analogical account, arguing that we should favour the latter.

Rosch's findings about the prototype structure of categories provide strong support for the analogical theory of meaning. The fact that we tend to think of the most prototypical members of a category when we think of a word denoting that category suggests that prototypes play a key role, not only in how we think of categories, but in how we categorise other stimuli. It seems that we will very often assign something to a category, because we perceive a family resemblance between that object and the most prototypical members of the category. The most prototypical members of a category are just those members that have the most attributes in common with other members of the category – and thus are most representative of that category, and so the perception of analogical resemblances between a freshly perceived stimulus, and category-members we regard as prototypical, will often play a vital role in categorisation. That is not to say that the less prototypical members of a category will not have any role at all to play in categorisations driven by analogical comparison – for we may well reason, for example, on our first encounter with an emu (supposing we have never seen or heard of one before), that it should be categorised as a member of the category 'bird' because it resembles other things we have already assigned to that category – say ostriches, for example, even though ostriches are not very prototypical birds at all. However, the more prototypical members of a category do have a more central role to play, generally speaking, in our assignments of other members to that category, than the less prototypical members do, given that they are more representative of the category as a whole. The most important point for the analogical account, however, is that categories are structured based on family resemblance relationships, rather than membership criteria (however intricate).

On Rosch's account, categories simply are collections of stimuli, or objects, that we consider to be

equivalent for the purposes of categorisation<sup>96</sup>, and the reason why we group them together in this way is because of the family resemblance-relationships obtaining between the members of a category. When we assign something to a category, we do so because we perceive an analogical resemblance between this freshly perceived stimulus, and previous things we have assigned to that category. Our analogies that underlie the meaningful application of a word to a newly encountered object, are based on the perception of there being overlapping properties, at some level, between this newly encountered object, and objects we have previously experienced that word being applied to. On Rosch's view, our categories themselves are structured based on the analogies we perceive between objects (which amount to the family-resemblance relationships that we observe), which leads us to consider them as 'the same kind of thing', and thus members of a common category. Thus, if categories have a prototype structure which is correlated with degree of family resemblance between category members, then this kind of account of categories explains how analogy could play such a central role in linguistic meaning. We've seen that there are problems with the classical account of the categories corresponding to many of our words, according to which such categories are defined by necessary and jointly sufficient conditions, but we are able to overcome these problems by adopting a different account of categories, which is also backed by considerable psychological evidence. If categories are defined in the way that Rosch suggests, then it seems that meaning something by a word corresponding to one of these categories, should not be thought to consist in following a rule, but in drawing an analogy – or perceiving a family-resemblance between stimuli.

I have suggested that meaning something by a linguistic expression requires us to draw an analogy between the present instance, and previous things that the expression has been applied to – and thus something is meaningfully referred to by a linguistic expression, by virtue of the perception of a family resemblance between that object, and previous objects that have been referred to by that expression. One doubt that may be raised against such an account, is that if meaning something by a word simply

---

<sup>96</sup> Rosch(1999)p.190

requires someone to make an analogy, rather than to follow a rule, then such meanings must be private, rather than public, and cannot serve as a normative constraint on the use of each individual speaker. While on such an account, every speaker is extrapolating from their own experience – that is, drawing analogies between their own past experience of an expression, and the present instance – as we have seen, the members of a linguistic community tend to agree in their judgments of prototypicality, and thus on which items are the most representative of any given category. Therefore, while my meaningful use of an expression is driven by my analogies, and your meaningful use of an expression is driven by your analogies, the psychological evidence concerning prototypicality judgments helps explain how people are able to communicate successfully, despite each person extrapolating from distinct (though analogous) bodies of experience. Such an account is also supported by the evidence that people do not tend to agree on the boundaries of categories – the fact that categories are defined by family resemblance relationships, rather than criterial features, means that the phenomenon of borderline cases does not present a problem.

While I acknowledged that one doubt that may be raised against the analogical account is the idea that when someone means something by an expression, in so doing, they are extrapolating from their own body of experience – and thus some doubt may be cast on the claim that meaning is public, and a normative constraint on the use of each individual – it seems that the alternative epistemicist account of meaning, which has it that meaning is external (and therefore, in theory – public and shared) runs into its own set of problems here. If category-membership is defined by the presence of some sufficient quantity of a matrix of disjunctive properties, then, as observed, it seems unlikely that we will be able to work out the exact nature of this matrix, and furthermore, it seems extremely hard to see how meaning, so conceived, could have any kind of normative force, when it comes to the usage of individual speakers. On this account, given the chaotic nature in which meaning is determined by use, we could end up with individual speakers being totally insulated from the meanings of their expressions, as they are unable to cognitively access these meanings – which makes it hard to see how meaning could have the kind of normative force that we want it to. If there are rules for the meanings of our expressions, but they are

determined in the way that Williamson states, then it's hard to see how we could follow these rules, or how they could act as a constraint on our appropriate usage of linguistic expressions.

Conversely, the issues highlighted concerning how meaning could be public, and therefore, act as some kind of normative constraint on usage, on the analogical account of meaning, seem rather less insurmountable. The fact that our prototypicality judgments tend to converge, and that our prototypicality judgments are robustly related to degree of family resemblance, helps point the way towards how meaning could be public, and communication could be successful, in spite of the fact that each one of us is simply extrapolating from our own body of experience. We can say that each of us extrapolates from our own experience, but that our individual bodies of linguistic experience are strongly analogous to each other, which explains how communication can succeed, and why our prototypicality judgments tend to converge.

While it is possible to argue that this evidence is irrelevant when it comes to what our words actually mean, as someone like Williamson may suggest, in favour of an epistemicist account of meaning – it seems that if meaning is determined by use (as Williamson will agree), then the psychological mechanisms which underlie our categorisations, and give rise to our categories, suggest that the perception of family-resemblance relationships (or – analogies between stimuli) is what drives our categorisations, and our meaningful use of language. Williamson could accept this much, but still insist that the meanings of our expressions are a product of the linguistic behaviour of many members of a linguistic community over time, such that it is the overall pattern of use that gives to a particular meaning, which, given the complexity of this pattern, may be impossible for us to epistemically access. Williamson could claim that all the evidence reveals is something of the psychological mechanisms underlying any given individual's use of language, with meaning being a different kind of thing entirely. However, as I have indicated, if our categories do have sharp boundaries, of which we are ignorant, as the words corresponding to these categories are governed by rules of great complexity that are determined by

“unsurveyably chaotic”<sup>97</sup> patterns of use, then it’s hard to see how these kind of rules could have any kind of normative force. It’s hard to see how such rules could be public in any meaningful sense, seeing as they’d be cognitively inaccessible to individual speakers. However, the analogical account of meaning is able to reflect the fact that the meanings of our expressions are public in a meaningful sense, by indicating that our prototypicality judgements tend to agree, and that these judgements track the degree of family resemblance we perceive between objects.

### Conclusion

To conclude this section – I have explored an alternative way of thinking of categories – rather than being defined in terms of membership criteria, categories could be defined by family resemblance relationships, and could have a prototype structure. I have taken Rosch and Mervis’s experiments to offer compelling support for their hypothesis, that the most prototypical members of a category are those items that have the highest degree of family resemblance to other category-members, and the lowest degree of family resemblance to members of other categories. I have argued that while the experiments focussed on concrete nouns, there may be good reason to believe that the categories corresponding to many of our word-types have a prototype structure, as well. I claimed that given the psychological evidence, there are two ways of interpreting the evidence. Either it demonstrates that our categories do not have formal membership criteria, but are defined by family resemblance relationships, which fits along well with an analogical account of meaning, in which the analogies that underlie our meaningful use of an expression focus on such family resemblances. Alternatively, we could uphold a classical account of categories, in spite of the evidence, by adopting an epistemicist account of meaning. However, I argued that an epistemicist account of meaning seems to be a very ad hoc way of interpreting the family resemblance data, and that furthermore, such an account makes it hard to see how meaning could be public, and a normative constraint on the usage of each individual, in a meaningful way. On the epistemicist account, we end up being insulated from the meanings of many of our expressions, which makes it hard to see how

---

<sup>97</sup> Williamson(2001)p.209

such meanings could be public, and play this normative role. I argued that while one doubt that may be raised against the analogical account is just this – that it fails to uphold the idea that meaning is public and a normative constraint on our usage – the analogical account fares better than the epistemicist account in responding to this criticism. The analogical account of meaning, working alongside a prototype theory of category-structure, can cite the fact that our prototypicality judgements tend to agree, as reason to believe that meaning could still be public in some meaningful way. The analogical account can claim that while each of us, in using language meaningfully, is extrapolating by analogy from our own body of linguistic experience, our bodies of linguistic experience are strongly analogous to each other, which explains how communication can succeed, and our prototypicality judgements tend to agree.

## Chapter 3 – Metaphor

### Introduction

In this section, I will argue that there are only two plausible ways for how rule-following accounts of meaning can conceive of metaphorical uses of linguistic expressions in contrast to literal uses. They can either deny that there is such a thing as metaphorical meaning, in contrast to literal meaning, and explain the difference between metaphorical and literal uses pragmatically. Alternatively, they can claim that there is such a thing as metaphorical meaning, in contrast to literal meaning, but that such meaning must be governed by analogy, rather than rules. However, I will argue that both accounts run into trouble when made to account for dead metaphors – expressions that have gone from being used metaphorically to being used literally. Both accounts rely on there being a sharp distinction between the literal and the metaphorical, which, I will argue, is hard to uphold. I will propose that these difficulties can be overcome if we reject the idea that there must be a sharp distinction between the literal and the metaphorical, and propose that both literal and metaphorical meaning alike are governed by analogy, rather than rules.

### A rule-following account of Metaphor

Linguistic expressions can be used either in a way that is regarded as literal, or in a way that is regarded as metaphorical. For example, the word ‘prison’ used literally presumably refers to (as the dictionary states) “a building where criminals are kept as a punishment”<sup>98</sup>, while the word ‘prison’ used metaphorically can refer to all sorts of things that are not buildings in which criminals are legally held for reasons of punishment. Thus someone can say that their school, or their job, is a ‘prison’, without implying that it is literally a prison. It seems that in this latter case, when the word ‘prison’ is being used metaphorically, then what is being implied is not that the school or job in question is literally a prison, but that it has certain properties in common with some of those things that we apply the word ‘prison’

---

<sup>98</sup> Little Oxford English Dictionary, Eighth Edition, Oxford University Press, 2002, p.551

literally too. Specifically, such a use of the word 'prison', would seem to imply that the school or job in question was an unpleasant place to be, in which the 'prisoners' (or employees, or students) feel they are being held against their will. Evidently, rule-following accounts of meaning will equate literal uses of linguistic expressions with uses that are governed by the rule for the meaning of that expression. Thus the word 'prison' applies to a building in which prisoners are kept, because that building accords with the rule for the meaning of the word 'prison'. However, it is rather less obvious how rule-following accounts should treat metaphorical uses of linguistic expressions, in contrast to literal uses.

We're supposing that literal uses of expressions are to be identified with uses that are governed by the rule for the meaning of that expression. It seems obvious that metaphorical uses cannot be governed by the very same rule as literal uses. If there is some set of necessary and jointly sufficient conditions that determines whether or not something is literally, say, a 'prison', then it doesn't seem that these conditions can also determine whether or not the word 'prison' can be applied to something metaphorically. It seems as though there's a potentially limitless variety of ways in which an expression could be used metaphorically, and these are clearly not straightforwardly determined by the rule that determines literal uses of the expression. Metaphorical uses are distinguished from literal uses, and these categories are mutually exclusive, and thus they cannot both be determined by the same rule. To give another example – the word 'pig' is applied literally to non-human animals of a certain sort, but it can also be applied metaphorically, and is often so applied to human beings. Usually, the implication here is that the human being in question is either greedy, or messy, or perhaps both. These are properties that, rightly or wrongly, are stereotypically associated with literal instances of the word 'pig'. However, when we use the word 'pig' literally, this presumably involves no direct implication that the animal in question is either greedy or messy, but simply that it belongs to a certain biological species (or to a family of related species).

It seems that when an expression is applied to some object metaphorically, it must be applied to this object for a different reason (or, at least – an additional reason), than the reason that that expression would

be applied to some object literally. One way to explain this is to say that when the expression is used metaphorically, it means something different than that expression does when used literally. However, while it is evident that we apply an expression to some object metaphorically, for a different reason than we would apply that expression to some object literally, some have claimed that this difference does not reside at the level of meaning. Donald Davidson has claimed that there is no such thing as metaphorical meaning. Thus, on Davidson's account, metaphorical uses of linguistic expressions do mean the same as literal uses of those expressions, and thus both metaphorical and literal uses alike are governed by the same rule. I will now discuss his account.

Donald Davidson has claimed that sentences involving metaphors “mean what the words, in their most literal interpretation, mean, and nothing more.”<sup>99</sup> Thus, if someone says that Bob (a human being) is a pig, their sentence means that Bob is a certain kind of farmyard animal (rather than merely that he is greedy or messy). For this reason, Davidson concludes that most sentences involving metaphors are false – as the sentence, understood literally, is false.<sup>100</sup> Davidson does allow that we can learn something from metaphors, even though sentences involving them are nearly always false – but insists that what we learn is to be cashed out pragmatically. We hear a sentence that is literally false, and this leads us to realise that that which the speaker meant to communicate does not reside in the meaning of the words they produced, but in some other factor. While the metaphorical use of an expression (say the metaphorical use of the expression ‘pig’) may cause us to notice, and even be intended to cause us to notice, certain features of the object being described metaphorically (for example – Bob's greed) – Davidson thinks it is a mistake to confuse what words are used to do (say – to draw our attention to certain features of objects) with what they mean.<sup>101</sup>

However, it seems very strange to treat what words are used to do, as an entirely different matter to

---

<sup>99</sup> Davidson(1978)p.32

<sup>100</sup> Davidson(1978)p.41

<sup>101</sup> Davidson(1978)p.33

what they mean. If any given linguistic expression had never been used, it wouldn't mean anything to anyone, and, while I do not have room to go into it here, I think that there are good reasons to think that all linguistic meaning is determined by use. Thus, if a word is regularly used to cause people to notice some particular feature of objects, and with the intention of causing people to notice this particular feature, then it seems strange to insist that this does not impact what is meant by the word. Whether or not we insist that meaning is, in fact, supervenient on use, I think that there are good reasons to reject Davidson's account of metaphors, which I will soon explore. However, first I will examine another way in which rule-following accounts of meaning may distinguish metaphorical uses of expressions from literal uses. Rule-following accounts may accept, contrary to Davidson, that there is such a thing as metaphorical meaning, in contrast to literal meaning. However, I will argue that it's not plausible to suppose that such meaning is governed by a rule. I will claim that if rule-following accounts want to allow that there is such a thing as metaphorical meaning, then they will be forced to propose that such meaning is governed by analogy, rather than rules. Before I layout this positive alternative, for the rule-following account, I will dismiss the idea that metaphorical uses of linguistic expressions could be governed by a special kind of rule, distinct from that which governs literal uses – a dismissal, in fact, which Davidson would endorse.

One possibility for how metaphorical uses could be understood in contrast to literal uses, is that while literal uses are governed by the rule for the meaning of the expression, metaphorical uses could be understood to be governed by a different rule instead. Thus there is one rule that determines, say, whether or not some object is literally a 'prison' (stating that it must be a building in which people are legally held for punishment, or something of this sort), and another, distinct, rule which determines whether or not some object can be metaphorically called a 'prison'. Presumably, given the reasons we examined that usually lie behind such metaphorical uses, the rule that determines metaphorical uses in this case would state something like: something is metaphorically a 'prison' if it is an unpleasant place to be, that those inside wish they could leave. However, it seems to me that a rule along these lines, no matter how

detailed it is, cannot possibly determine for every possible object, whether or not the word ‘prison’ could be applied to it metaphorically. We could refer to a particular theory or belief system, metaphorically, as a ‘prison’ (implying that it impaired the intellectual freedom of those who believed it, perhaps), which demonstrates that not every possible metaphorical use of the word ‘prison’ must refer to an ‘unpleasant place’ of some sort. Metaphorical uses of expressions can be novel, and creative, and it doesn’t seem, therefore, that each such usage could be captured by some pre-existing rule. Another possibility for accounting for novel metaphorical uses, however, would be to claim that each novel metaphorical use of an expression is governed by a rule – though not a pre-existing rule, but one that the novel usage itself brings into play. I will now discuss and dismiss this account too.

It might be suggested that when a novel metaphor is used, it is governed by a special kind of rule - not a pre-existing rule, but one that comes into being with the creation of the metaphor. Davidson rejects an account like this – suggesting that while it may be tempting to think that when an expression is applied to some object in a novel metaphorical way, this tells us that the set of all objects the expression is literally true of, and the object that is being described metaphorically have some “special, surprising property ...in common”<sup>102</sup>, which can be equated with the meaning of the newly minted metaphor – we should reject any such account. I think that Davidson is right here – it couldn’t be that when an expression is used in some novel, metaphorical way, that this very action brings a new rule into play, which determines the extension of the metaphor. This would in fact mean that metaphorical meaning works just like literal meaning – though not in the way Davidson himself wants to suggest. If metaphors worked like this, there would be no distinction between literal and metaphorical language, it’d just be that literal uses of expressions relied on pre-existing rules, while (novel) metaphorical uses of expressions brought new rules into play. As Davidson writes – if this account of metaphor were correct then “to make a metaphor is to murder it”.<sup>103</sup> What we had called ‘metaphor’ would turn out to amount simply to the introduction of a

---

<sup>102</sup> Davidson(1978)p.34

<sup>103</sup> Davidson(1978)p.34

new term in our language – an expression with a new literal meaning.

This does not seem to be how metaphors work, however - it doesn't seem as though each and every novel metaphor could somehow bring a rule into play, which would tell us, for any possible object, whether or not the expression could be applied to it in this new metaphorical way. It seems bizarre to think that simply by applying an expression to an object - O, that that expression is not literally true of, we could thereby, automatically, generate a new, literal rule, which applies to O and all those objects the expression had, until that time, been literally true of, and specifies exactly what properties are shared by the members of this class. If this were so, there would be no such thing as applying an expression to an object incorrectly, or mistakenly – no matter how one applied expressions to objects, one would be right to do so, automatically, by virtue of these spontaneously generated rules. Thus it seems that if literal uses are rule-governed, then metaphorical uses must be thought of in some other way, seeing as they cannot be governed either by the same rule that governs literal uses, or by some other, distinct rule governing metaphorical uses.

A better picture of what goes on when we use an expression metaphorically (assuming that literal uses are equated with rule-governed uses), is that we apply the expression to some object metaphorically, not in accordance with any rule, but due to a perceived resemblance between the object in question, and some literal instances of the expression. That is – a metaphorical use is made, not in accordance with the rule, but because the object being referred to is seen to have certain properties in common with some of those literal instances, that really do accord with the rule. Thus, the word 'prison' gets applied metaphorically, say – to a school, not because there's some rule that determines that (some or all) schools fall under the metaphorical extension of the word 'prison', but simply because a resemblance is being noted between the school, and some of those things that really are 'prisons' – namely that both are unpleasant places to be, and so on. Specifically, it seems that the kind of properties that a metaphorical use of an expression tends to conjure up, are those properties that are stereotypically associated with literal instances of the

expression. Thus, if some place is metaphorically referred to as a ‘prison’, this implies that it is an unpleasant place to be, because our stereotypical idea of a prison tends to be of such an unpleasant place. Metaphorical uses can, therefore, be understood as relying on analogical resemblance to literal uses of the expression, and specifically to there being shared properties between the metaphorical instance of the expression, and literal instances of the expression (especially those that are thought of as stereotypical). If we equate literal uses with uses that are governed by the rule for the meaning of the expression, then it seems that metaphorical uses will be best explained as derivative on these literal uses, though ultimately not governed by any rule, but by analogy to some of these literal instances, instead.

The proposed account of how metaphorical uses of linguistic expressions work, and how they relate to literal uses, allows us to account both for established metaphorical uses of expressions, and for novel metaphorical uses of expressions, as well. For established metaphorical uses of expressions, we are able to highlight the kind of properties that such uses usually imply about the object being referred to metaphorically – for example, metaphorical uses of the word ‘pig’ usually imply that the object (or, most likely, person) being referred to is greedy and/or messy. The proposed distinction between using an expression literally and using an expression metaphorically, does not, however, merely apply to established metaphors, on which the kind of properties the metaphorical use of an expression tends to imply are already familiar to most speakers.

It applies equally to novel metaphorical usages, in which an expression with an established literal meaning is being used in a new and unfamiliar way, by being applied to some object metaphorically. For example, in the opening line of his novel ‘The Go-Between’, L. P. Hartley introduces a novel and intriguing metaphor: “The past is a foreign country: they do things differently there.”<sup>104</sup> Despite there not having been an established convention of referring to things metaphorically as a ‘foreign country’, Hartley makes quite clear what kind of implication this metaphorical usage is supposed to have. Just as,

---

<sup>104</sup> Hartley(2004)p.5

on visiting a foreign country, one's surroundings can initially seem unfamiliar, and perhaps even mysterious, in comparison to one's native (or resident) country, so too, when we look back on the past, even our own past, it can seem unfamiliar and mysterious, in comparison to our 'native' time.

Thus, novel metaphorical usages, too, are readily understood, and it seems that they too depend on there being properties in common between that which is being referred to metaphorically, and some of those things (especially the more stereotypical ones) that are referred to when the expression is used literally. This is, of course, what we should expect, given that every established metaphor was once novel. Thus, the proposed account is seemingly able to account for every actual and possible metaphorical use of an expression – for it seems that any metaphorical use will rely on there being some kind of property or properties in common between that which is referred to metaphorically, and some of those things that are referred to when the term is used literally. Thus we have seen that there are two alternatives for how a rule-following account of meaning may account for metaphorical uses of linguistic expressions, in contrast to literal uses. They may either deny that there is such a thing as metaphorical meaning, as Davidson does, or they may accept that there is metaphorical meaning, but propose that it is governed by analogy to literal instances of the expression, rather than rules.

I have argued that a rule-following account of meaning, will equate literal uses of linguistic expressions with uses that are governed by the rule for the meaning of that expression. I have outlined two alternatives for how a rule-following account may conceive of metaphorical uses of linguistic expressions, in contrast to literal uses. They might either claim, as Davidson does, that there is no such thing as metaphorical meaning as opposed to literal meaning, and that that which metaphorical uses of linguistic expressions communicate above and beyond the literal meaning of the words is to be understood pragmatically. Or they might claim, that there is such a thing as metaphorical meaning in contrast to literal meaning – but that it works differently to literal meaning. I have argued that it is not plausible to suppose that metaphorical uses of linguistic expressions are governed by rules, so that if the

rule-following account of meaning wishes to posit a special kind of metaphorical meaning, in contrast to literal meaning, then such meaning must be governed by analogy, rather than rules. Specifically, it seems that a metaphorical use of an expression relies on there being properties in common between that which is referred to metaphorically, and literal instances of the expression. However, I will argue that both of these options, for how a rule-following account should conceive of metaphorical uses of linguistic expressions in contrast to literal uses, run into serious problems. Specifically, both accounts have a hard time accounting for the phenomenon of dead metaphors. Further, both accounts rely on there being a sharp and principled distinction between the literal and the metaphorical use of an expression, which, I will argue, is hard to uphold. I will argue that the only way to overcome these problems is to deny that there is a sharp distinction between the literal and the metaphorical, and to argue instead that literality and metaphoricality are a matter of degree, with literal and metaphorical uses of language alike governed by the same principle – analogy, rather than rules.

### Dead Metaphors

According to both of the accounts I have considered, for how rule-following accounts of meaning may distinguish between the literal and the metaphorical use of an expression, metaphorical uses of expressions, are, in some way, derivative on literal uses of those expressions. On Davidson's account, this is because there is no such thing as metaphorical meaning in contrast to literal meaning, such that whatever is communicated by a metaphor, is explained fully by the literal meaning of those expressions. On the alternative account I considered, on which metaphorical uses, in contrast to literal uses, are governed by analogy – metaphorical uses are still derivative on literal uses, insofar as they require the perception of an analogy between that which is being referred to metaphorically, and literal instances of the expression. However, I will argue that both accounts thus run into trouble when made to account for the etymological development of our language. Both accounts hold that the metaphorical is derivative on the literal – such that we should expect an expression to develop a literal meaning, before it can go on to be used metaphorically. In short, the etymological development of any particular expression in our

language can seemingly only flow in one direction – from an expression being used literally, to that expression being used metaphorically. However, many expressions of our language have, at some point in the history of their use, gone from being used in a metaphorical way, to being used in a way that is regarded as literal, and it is necessary for an account of the literal/metaphorical distinction to be able to explain these etymological transitions, as well. Such expressions are referred to as dead metaphors. I will first discuss dead metaphors, before going on to discuss the problem that such expressions pose to Davidson's account. I will then propose that dead metaphors pose a threat to the alternative account on which metaphorical uses (in contrast to literal uses) are governed by analogy, as well.

Etymologically, many expressions in our language started out as metaphors, but are now used in a way that is thought of as perfectly literal. These are the *dead metaphors*. For example, we speak of table 'legs', the 'branches' of a bank, and the 'hands' of a clock. Presumably, the word 'legs' was originally applied to the poles that hold tables (and chairs, etc.) up, because of a perceived analogy between human and animal legs, and these poles, leading to them, initially, being metaphorically described as table 'legs'. Similarly, the word 'branch' was presumably originally applied to individual institutions in a bank, because of a perceived analogy between the role of individual institutions in a bank, and the role of branches in a tree. Finally, the 'hands' of a clock were presumably originally so called, because of a perceived analogy between these 'hands', and the hands possessed by humans, and various other primates – specifically, both are (or can be) used for pointing. Early reference to the 'legs' of a chair, the 'branches' of a bank, and the 'hands' of a clock were metaphorical, while when we speak of such things today, we take ourselves to be speaking literally, making reference to the 'legs' of chairs, the 'branches' of a bank, and the 'hands' of a clock, dead metaphors. In fact, a good deal of our linguistic expressions can be regarded as dead metaphors – to give a few more examples – a clock has a 'face', as well as 'hands', and a chair can have 'arms', as well as 'legs', while an essay can have a 'body', and a bed can have a 'foot'. The phenomenon of dead metaphors demonstrate an important etymological fact – expressions can go from being used in a way that is regarded as metaphorical, to being used in a way that is regarded as literal.

Rule-following accounts of meaning claim that literal uses of linguistic expressions are governed by rules. Thus, in accounting for dead metaphors, a rule-following account of meaning must distinguish between the rule governing the original, literal meaning of the expression, and the rule governing the new, dead metaphorical use of the expression. So to focus on the example of table ‘legs’ – there was a rule which restricted the application of the word ‘legs’ to human and animal legs – call this R1. The word ‘legs’ was then used metaphorically in reference to table ‘legs’, and now, by virtue of being a dead metaphor, it is now used literally in reference to table ‘legs’. Thus there is now a rule which governs the application of the word ‘legs’ to table ‘legs’ – call this R2. However, while such accounts can distinguish between the rule governing the original literal meaning of the expression, and the rule governing the new literal meaning of the expression, they must also offer us some explanation of what was going on when the expression was used metaphorically. One option is to claim, as Davidson does, that the expression, when used metaphorically, simply had the original literal meaning (and so was governed by R1), rather than any kind of metaphorical meaning. While Davidson claims, that the expression, when used metaphorically, causes us to “notice a likeness” (say between human and animal legs and table ‘legs’), in contrast to its use as a dead metaphor, at which point “there is nothing left to notice”<sup>105</sup>, he claims that the expression never had a metaphorical meaning – when it was a live metaphor, it was governed by R1, when it was a dead metaphor, it was governed by R2. However, Marga Reimer has claimed that the phenomenon of dead metaphors demonstrates that Davidson is wrong to deny that there is such a thing as metaphorical meaning. I will examine these arguments, finding them to be conclusive, before looking at how the alternative rule-following account (on which there is such a thing as metaphorical meaning, and it is governed by analogy) may account for dead metaphors.

Marga Reimer has claimed that the Davidsonian thesis that metaphors do not mean anything, over and above the literal meaning of the words, runs into serious trouble when made to account for the

---

<sup>105</sup> Davidson(1978)p.37

phenomenon of dead metaphors. She argues that a Davidsonian account of the meaning (or lack thereof) of metaphors makes it seem impossible to explain how it could be, that the dead metaphor, upon dying, acquired “the particular literal meaning that it did”.<sup>106</sup> Reimer looks at an example that Davidson considers extensively – the sentence “He was burned up”, which now has a literal meaning that “He was very angry”.<sup>107</sup> Once upon a time, the (only) literal meaning of this expression was that “He was destroyed by heat” – and this expression was then used to communicate the idea that “He was very angry” in a metaphorical way, and now today, by virtue of this expression being a dead metaphor, we take it to also possess the literal meaning that “He was very angry”.<sup>108</sup> Reimer asks how it is that the expression could have acquired this literal meaning – and, in fact, how it is, more generally, that any expression could come to possess a literal meaning.

Reimer proposes that the most plausible way to explain how it is that an expression comes to acquire a literal meaning, is a Gricean one, on which an expression must be regularly used in a particular way within a community of speakers, if it is to acquire a literal meaning. Specifically, she suggests “an expression *e* acquires literal meaning *m* in linguistic community *c* just in case *e* is regularly used, over a period of time, in *c*, to mean (roughly, to communicate) *m*.”<sup>109</sup> While Davidson may not accept such an account, given his insistence on making the distinction between what an expression means, and what it is used to do, the account does seem plausible. Surely, it is only by being used in a particular way, that an expression can come to be seen to have a particular meaning. Linguistic meaning, just like any other convention, must be established by a particular pattern of behaviour – specifically, by the linguistic expression being applied to objects in a particular kind of way. If we accept this seemingly plausible account of how literal meanings come about, and apply it specifically to dead metaphors, and the literal meanings they acquire upon dying, then Davidson’s account faces real problems.

---

<sup>106</sup> Reimer(1996)p.14

<sup>107</sup> Reimer(1996)p.14

<sup>108</sup> Reimer(1996)p.14

<sup>109</sup> Reimer(1996)p.14

Davidson claims that when an expression that is, today, a dead metaphor, was used metaphorically – its meaning consisted simply in the literal meaning of the words, at that time. Later on, when the expression is used literally, as a dead metaphor, then the meaning consists in the newly acquired literal meanings of the words. If we accept a Gricean account of how expressions acquire literal meanings – then this account tells us that a dead metaphor (say the sentence “He was burned up”) acquires a literal meaning (say – “He was very angry”), by being used, amongst a community of speakers, over a period of time, to mean that (i.e. “He was very angry”). Yet on Davidson’s account, it is only once the metaphor is already dead that it means anything, above and beyond the original, literal meaning of the words. On Davidson’s account, therefore, the newly acquired literal meaning of a dead metaphor seems to be spontaneously generated out of nothing, with no explanation of how it is that the formerly metaphorical use of the expression relates to the new literal meaning, upon the death of the metaphor. Davidson’s account does not seem to allow that the meaning of a dead metaphor could have been established gradually, since he must insist that the transition between the expression being used with its original literal meaning (“He was destroyed by heat”), and the expression being used with its new, dead metaphorical meaning (“He was very angry”), must be an abrupt, and binary transition. Thus Davidson seems to be committed to the idea that metaphors must die abruptly. Davidson is committed to the idea that there is a sharp cut off between the expression being used with its original literal meaning (in accordance with one rule – R1), and it being used with its new dead, dead-metaphorical meaning (in accordance with another rule – R2), given that there is no such thing as ‘metaphorical meaning’ to bridge the gap.

If we accept the Gricean account, then a word must be regularly used to mean or communicate something, before it can come to have that literal meaning. Thus in order for a metaphor to die, and for it to acquire a literal meaning, that metaphor must have been used to mean that which goes on to become the dead metaphor’s literal meaning. Thus, even when the metaphor is still alive, it must mean something,

otherwise the literal meaning of the dead metaphor could never have been established. So if we accept this account of how it is that literal meanings of linguistic expressions are established, then, contrary to Davidson's insistence, we must allow that there is such a thing as metaphorical meaning. On Davidson's account, when people referred metaphorically to table 'legs', the rule that governed their use of the expression, was R1 – which restricts the application of the word 'legs' to human and animal legs, while when people use the word 'legs' as a dead metaphor, to refer literally to table 'legs', their use is governed by R2. If there is no such thing as metaphorical meaning, then it's hard to see how R2 could have become established, seeing as no one had ever used the expression 'legs' to mean anything other than that captured by the rule R1, up until that point. It seems hard to explain how the metaphorical use of the expression preceding the death of the metaphor, related to the literal meaning that became established for the dead metaphor, if the expression never had a metaphorical meaning.

I think that Reimer is right to claim that the phenomenon of dead metaphors critically undermines Davidson's claim that metaphors mean, only what the words, interpreted literally, mean. However, while I think that the phenomenon of dead metaphors demonstrates that there must be such a thing as metaphorical meaning, contrary to Davidson's account, it also poses a serious threat to the alternative account of the literal and the metaphorical introduced earlier, on which metaphorical uses of linguistic expressions are governed by analogy to literal uses of those expressions. This is because both accounts rely on there being a sharp and principled distinction between the literal and the metaphorical. Davidson had to claim that live metaphorical uses of an expression were governed by one rule (R1), while dead metaphorical uses were governed by another (R2), and there must thus be a sharp cut-off between uses governed by one rule, and uses governed by the other. However, the alternative account, on which metaphorical uses are governed by analogy, also has to posit a sharp cut off. It has to propose that live metaphorical uses are governed by analogy to instances of one rule, while dead metaphorical uses are governed by a new rule, with any particular usage governed either by analogy or by a rule. This account is committed to there being a difference in kind between metaphorical and literal uses, and thus to the idea

that metaphors must die abruptly. I will argue that this is problematic, and should lead us to doubt that there is a sharp and principled distinction between the literal and the metaphorical use of an expression.

I suggested that a rule-following account of meaning, that equates literal uses of expressions with uses that are governed by the rule for the meaning of that expression, can think of metaphorical uses not as rule-governed, but as determined by analogy to some of those objects the expression is literally applied to. I will now examine how such an account can deal with the phenomenon of dead metaphors. On this account, when we think about the early metaphorical uses in the history of the use of an expression that is, today, a dead metaphor, then it seems that these uses will not be governed by any rule, but by analogy with some of those things that that expression, used literally, applied to at that point in history. For example, when people referred metaphorically to table 'legs', they presumably did so because they saw an analogical resemblance between human and animal legs, and the poles that support tables – both kinds of *leg* are parts of some object (or organism) that serve the function of raising it vertically from the ground, and they look similar in certain respects; both in their shape, and in their spatial position in the whole of which they are a part. These people referred to table 'legs', not by following some rule which determines what it is for something to be a 'leg', in this sense, but because they saw an analogy between some of those objects that really did accord with the 'leg'-rule (human and animal legs) in their community at the time, and the table 'legs' they were referring to metaphorically.

Thus, on the account of dead metaphors under consideration – we can distinguish between the rule governing the literal use of an expression (R-1), that restricts its application to objects of a particular kind - objects of kind P (i.e. human and animal legs), and the analogy that governs the metaphorical use of that expression (A(R-1)), that is used to refer to objects of another kind - objects of kind K (i.e. table 'legs'), by virtue of their perceived resemblance to objects of kind P. Once the expression is used as a dead metaphor, then a new rule comes in force (R-2), such that the expression in question can now be used to refer literally to objects of kind K, without requiring any kind of analogical comparison to objects of kind

P. It is not easy to see how it could be that the new rule (R-2) comes into force, and how dead metaphorical uses, those made in accordance with R-2, relate to those metaphorical uses of the expression that preceded them (those made in accordance with A(R-1)). As previously noted, there seems to be little difficulty in accounting for the introduction of a new metaphor, on the proposed distinction between literal and metaphorical uses. If metaphorical uses are derivative on literal uses, then it isn't hard to see how it could be that an expression could go from being used literally to being used metaphorically (from R1 to A(R1)). However, it is far from clear what the proposed account of the literal/metaphorical distinction should say about the death of a metaphor (from A(R1) to R2).

I proposed that a rule-following account of meaning can distinguish literal uses from metaphorical uses, by claiming that the former are governed by rules, while the latter are governed by analogy. However, if we argue that metaphorical meaning and literal meaning are fundamentally different in kind – as such an account of the matter suggests – then dead metaphors still present a problem. We're still faced with having to argue that a metaphor must die abruptly, since for any given usage of the expression, it must be that this expression has either a metaphorical meaning (governed by analogy) or a literal meaning (governed by rules). Thus, if metaphorical meaning and literal meaning are fundamentally different in kind, then for any expression that is today a dead metaphor, there must have been a first literal, rule-governed use of the expression as a dead metaphor, set apart from the metaphorical, analogy-governed uses of the expression that preceded it. We'll call this hypothetical, first rule-governed usage  $DM_1$ .

If we want to insist that live metaphors are governed by analogy, but that dead metaphors are governed by rules, then we're forced to the conclusion that there must have been such a first rule-governed, literal usage, of each dead metaphor. However, it's very hard to explain how there could have been such a sharp transition in the use of the expression. If there really was a  $DM_1$ , it must be the case, that something, either in the mind, or in the external environment, of that first person to use the expression as a dead metaphor, made it the case that this usage of the expression was the first usage made in accordance with

the rule that governs our use of the expression today. We've supposed that for each dead metaphor, we can distinguish between the original literal meaning of the expression, which is governed by one rule (R1), the analogy governing metaphorical uses (A(R1)) and another rule governing the literal meaning of the expression as a dead metaphor (R2). If we go along with Reimer's Gricean-style account of how literal meaning is established, then it seems that it must be the case that  $DM_1$  was individuated as such, because only at that point in history, had the expression been used regularly enough to mean a certain thing, such that it thereby acquired this as a literal meaning. Thus, the uses that were governed by A(R1) must have somehow laid the foundation for what would later become the meaning of the expression (R2).

However, according to the Gricean style account at issue, an expression must be used to mean a certain thing –  $m$ , before it acquires that as a literal meaning. Thus, it must be the case that some of those uses that were governed by A(R1), meant the same, as some of those uses later governed by R2 – both must have been used to mean or communicate  $m$ . It's hard to see how an expression with a metaphorical meaning, governed by analogy to instances of one rule (A(R1)) could mean the same, as a literal meaning, governed by an entirely different rule (R2). If a series of metaphorical uses of an expression, can give rise to a new literal meaning for that expression, and we conceive of metaphorical and literal meaning as totally different things (one is governed by analogy, the other governed by rules), then it's hard to see how the metaphorical usage of an expression can generate a literal meaning for that expression. If metaphorical and literal meaning are sharply distinguished, then we are committed to the idea that a metaphor must die abruptly, since its new literal use, must be differentiated from the metaphorical use that had preceded it. However, we need not be committed to the idea that literal meaning is sharply distinguished from metaphorical meaning – such that the former is governed by rules, and the latter by analogy, and so we need not be committed to the idea that metaphors must die abruptly. Indeed, it seems to be the commonsense view that metaphors die slowly – that as a particular metaphor is used more and more, it can become gradually less and less metaphorical, until the point that it no longer seems metaphorical at all, and is regarded as possessing a literal meaning. So if we can propose that literal and

metaphorical meaning, alike, work on a common principle, then it will be much easier to explain how it is that a metaphor could die slowly, rather than abruptly. If we claim that both literal and metaphorical meaning alike are governed by analogy, then we need no longer propose that the literal and the metaphorical use of an expression must be sharply distinguished, or that there must be a first literal, dead-metaphorical use of an expression, cut off from those live metaphorical uses that preceded it.

An analogical account of an expression that is regarded as a dead metaphor, does not need to suppose that there really was a DM<sub>1</sub> – a first literal, dead metaphorical usage of the expression that is sharply distinguished from all those uses preceding it. According to an analogical account of meaning, every meaningful use of an expression involves an act of analogical comparison – not just those uses we had called ‘metaphorical’ in the discussion above. It seems that if both literal and metaphorical uses of linguistic expressions alike are governed by analogy, then we can provide an account on which metaphors die slowly, since there is not a fundamental difference in kind between the literal and the metaphorical use of an expression. We can say that when a word is used as a live metaphor, this usage relies on analogical resemblance between objects, and that when it is used as a dead metaphor, this usage also relies on analogical resemblance, but on resemblance to different objects, and in different respects. For example, when table ‘legs’ were first referred to as ‘legs’, this relied on the perception of a quite abstract analogical resemblance between human and animal legs, and some particular table ‘legs’. As more and more things were referred to as table ‘legs’, the analogical resemblance such uses relied on became clearer, and similarities began to be noticed not just between human and animals, and table ‘legs’, but also between distinct particular objects that had been referred to as table ‘legs’. As more and more objects were referred to as table ‘legs’, and more and more similarities were noticed between these objects, reference to table ‘legs’ gradually became less and less metaphorical, as the particular resemblances to human and animal legs, on which this usage had first relied, became less salient.

We can propose that rather than literality and metaphoricality being a fundamental difference in kind,

it is a matter of degree, and thus we should not expect to find a sharp boundary between the metaphorical and the literal use of an expression today regarded as a dead metaphor. In fact, it may be that the categories of the 'literal' and the 'metaphorical' should be thought to have a prototype structure – since there are clear cases of each, but no sharp boundary between the two. It may be that as a metaphor dies, it becomes a less and less prototypical example of a metaphorical use of language, and a more and more prototypical example of a literal use of language. Given that live metaphorical uses, and dead metaphorical uses, alike, are both said to involve acts of analogical comparison, this allows us to respect the idea that a metaphor could die slowly, rather than abruptly. It also allows the change in meaning that occurs, when a live metaphor becomes a dead metaphor, to be a gradual one, in which we need not highlight a particular use by a particular speaker, as being particularly causally and developmentally significant, in the history of the use of the expression.

In accounting for dead metaphor in this way, an analogical account of meaning is rejecting the idea that there is a sharp transition between the literal and the metaphorical use of an expression. It is not the case, for example, that on the death of a metaphor, a new meaning-rule (R2) comes into play, in contrast to the meaning rule that had governed previous uses of the expression. This might seem to deny the reality of a seemingly evident distinction, between using an expression, today regarded as a dead metaphor, to refer to a particular kind of object metaphorically, and using that expression to refer to that particular kind of object literally. It might appear that while it may be vague just where the metaphorical usage meets the literal usage in the history of the use of the expression, this is not to say that there is not an evident distinction, that can be appreciated by considering clear cases, of using an expression metaphorically and using that expression literally. However, in order to challenge the idea that there is such a sharp distinction between the metaphorical and the literal usage of an expression, we need not merely speculate about hypothetical etymological transitions – we can find examples of uses of expressions for which it is a vague matter whether that usage is metaphorical or literal, even in many of the sentences we utter every day. I will now consider Stephen Yablo's arguments against there being a

sharp literal/metaphorical distinction, for without such a distinction, it is hard to maintain the idea that some uses of expressions (the literal ones) are rule-governed, and others (the metaphorical ones) are governed by something else instead – and it becomes more tempting to look for a common principle governing those uses we had called ‘literal’ and those we had called ‘metaphorical’, alike – which, I will go on to claim, is that all such uses are governed by analogical extrapolation.

### The Literal/Metaphorical Distinction

In his paper, ‘Does Ontology Rest On A Mistake?’, Stephen Yablo has argued that the boundaries between the literal and the metaphorical use of an expression are extremely blurry<sup>110</sup>. For the purposes of my argument, it seems that if the literal/metaphorical distinction cannot be upheld, then this puts another nail in the coffin of the idea that the meanings of expressions are spelled out in terms of necessary and jointly sufficient conditions. If, for example, we cannot isolate literal uses of the verb ‘turn’ from non-literal ones, then we’ll have to admit that not all meaningful uses of the verb can be captured in terms of necessary and jointly sufficient conditions. This will make the idea that what all meaningful uses have in common is analogical resemblances between particular instances seem all the more appealing. If the distinction could be spelled out, then it would be open to the advocate of necessary and jointly sufficient conditions to say that all literal instances of an expression obey a rigidly defined rule, while metaphorical instances might be driven by analogies, rather than rules. If it cannot be spelled out, however, then this might give us reason to think that every meaningful use of an expression is driven by analogy, rather than necessary and jointly sufficient conditions, given that we cannot isolate a bunch of clear cases of a category, from the murkier, less-than- literal ones.

Yablo points out that metaphorical (or ‘non-literal’, or ‘make-believe’) uses of linguistic expressions can nonetheless provide us with information about the world. He cites Kendall Walton’s examples that we could learn a substantial geographical fact by being told that the Italian town of Crotona is “on the arch of

---

<sup>110</sup> Yablo(1998)p.233

the Italian boot”, that we could identify a thundercloud to somebody by describing it as “the big, angry face near the horizon”<sup>111</sup>. These descriptions would seem very informative, in spite of their apparently non-literal nature. Yablo points out that by playing ‘make-believe’ games such as these, by treating various objects as props in a game (treating part of Italy as a ‘boot’, and a cloud as a ‘face’), we can, through playing this game, attribute game-independent properties to the objects we use as props in the game<sup>112</sup>. As Yablo points out, by speaking in such a way, we require that the world is “in a condition to make a pretence like that appropriate”<sup>113</sup>. He points out that very often in speech, if we tried to interpret things in a very literal way, our communicative abilities would falter (e.g. “‘that wasn’t such a great idea’, taken literally, leaves open that it was a very good idea.”<sup>114</sup>) He points out that metaphorical expressions have in addition to their literal content, a metaphorical content, each of which could be thought to determine a set of possible worlds in which the expression would be true<sup>115</sup>. He gives the example of ‘Jimi’s on fire’, determining, in terms of its literal content, a set of worlds in which that person Jimi has somehow caught fire, and in terms of its metaphorical content, a different set of worlds (presumably without much overlap!), in which Jimi is perhaps telling a long-stream of successful jokes, or winning some political or philosophical debate, or doing some activity successfully such that the metaphorical expression ‘is on fire’ could be thought to apply to him. He points out the indispensability of talking about ‘the average X’ – for example, “the average star has 2.4 planets”, which cannot be taken literally, seeing as there is no such thing as the average star (or the average X more generally), but it still conveys invaluable information about the world, about “what the (ordinary) stars are like on average”<sup>116</sup>.

Yablo’s arguments cited so far demonstrate the pervasiveness and usefulness of metaphorical language, in particular for conveying factual information. However, he argues further that we cannot easily determine which statements are to be taken as metaphors, and which are to be taken literally. He points

---

<sup>111</sup> Walton(1993)p.40-41

<sup>112</sup> Yablo(1998)p.246

<sup>113</sup> Yablo(1998)p.248

<sup>114</sup> Yablo(1998)p.248

<sup>115</sup> Yablo(1998)p.249

<sup>116</sup> Yablo(1998)p.250

out that it is hard to determine whether the sentence “after an extended delay, the game resumed”<sup>117</sup> is to be understood as literal or metaphorical – specifically whether the adjective “extended” can literally refer to times, and specifically delays, rather than merely to the boundaries of physical objects. He questions whether the adjective “calm” is literal or metaphorical when applied to people, bodies of water, and historical eras respectively, and whether the ‘backs’ and ‘fronts’ of “animals, houses, pieces of paper, and parades” is literal or metaphorical talk.<sup>118</sup> He suggests that such questions are “unanswerable”<sup>119</sup>. This means that for such sentences – ‘S’ – it is unclear whether the set of worlds in which they are true is to be understood as a function of their literal meaning, or their metaphorical meaning. Specifically, it is unclear whether particular words in S are to be understood as literal or metaphorical. If there is no boundary to be found between the literal and metaphorical use of a word, then this suggests that we will not be able to isolate some principle that every literal instance of the category denoted by the word obeys, given that the boundaries of the literal are so indeterminate. If we cannot isolate, the strictly-speaking, literal use of an expression, from the loose, less-than literal use of that expression, then the idea that there is some set of necessary and jointly sufficient conditions that every meaningful use of that expression must meet, seems very difficult to uphold. It is clear that there are not rigid rules for how metaphorical language must proceed, and to assume that there are rigid rules for how literal talk must proceed, is to assume that we can quarantine off the literal from the metaphorical, which, as Yablo demonstrates, cannot be done. It seems evident that if there is such a thing as metaphorical meaning (as the phenomenon of dead metaphors gives us cause to believe), then such meaning must be governed by analogy, and given that we cannot quarantine off the literal from the metaphorical, the idea that all meaningful uses of language are governed by analogy begins to seem all the more appealing.

---

<sup>117</sup> Yablo(1998)p.256

<sup>118</sup> Yablo(1998)p.256

<sup>119</sup> Yablo(1998)p.256

## Conclusion

To conclude this section, I argued that there are only two plausible accounts of how rule-following accounts of meaning can distinguish between the literal and the metaphorical use of an expression. They can either deny that there is such a thing as metaphorical meaning, in contrast to literal meaning, or they can accept that there is metaphorical meaning, but propose that it is governed by analogy, rather than rules. The phenomenon of dead metaphors demonstrates that there must be such a thing as metaphorical meaning. However, it seems to be the commonsense view that metaphors die gradually, and if we propose that metaphorical meaning is governed by analogy, while literal meaning is governed by rules, and that the literal and the metaphorical are thus fundamentally different, then we have a hard time explaining how it could be that a metaphor could die gradually, rather than abruptly. If we want to be able to respect the idea that metaphors could die slowly, then we have to drop the idea that there is a sharp distinction between the literal and the metaphorical, and propose that literality and metaphoricality, rather than being a difference in kind, is a matter of degree. If the literal and the metaphorical are understood in this way, then it must be that both kinds of meaning are governed by the same principle, which I have proposed, is analogy, rather than rules. I pointed out that while some people may find it objectionable to deny that there is a sharp distinction between the literal and the metaphorical, on the basis of hypothetical etymological transitions in the history of the use of expressions today regarded as dead metaphors, this is not the only reason we should reject the distinction. I cited Stephen Yablo's argument that even in many of the sentences we utter on a day to day basis, it's far from clear whether particular expressions within such sentences should be understood to be literal or metaphorical. Thus – this provides us with a good reason to propose that there is not a fundamental difference in kind between the literal and the metaphorical. I have argued that literality and metaphoricality can be considered a matter of degree, with literal and metaphorical uses alike governed by analogy.

## Chapter 4 – The Rule-Following Paradox

### Introduction

In this section, I will layout the rule-following paradox, that Saul Kripke interprets from Wittgenstein, in his book 'Wittgenstein : On Rules and Private Language'. This paradox casts doubt on the idea, that in using a linguistic expression meaningfully, we are following the rule for the meaning of that expression. This is because, when we look back on our past usage of an expression, we cannot cite some fact that demonstrates that this usage was governed by some particular rule, rather than any of an indefinite number of other rules, with which this usage was also compatible. Having cast doubt on the idea that we were following some particular rule in the past when we used a given meaningful expression, Kripke points out that doubt must also be cast on the idea that we are following a particular rule when we mean something by a linguistic expression in the present. I will first layout the paradox, as Kripke presents it, focusing initially on the mathematical example of *addition* on which he focuses, before indicating the generality of the problem. I will then propose that the analogical account of meaning can offer a good solution to the paradox. I will indicate that this is a skeptical solution to the paradox, insofar as it does not try to salvage the idea that we really are following rules, after all, but attempts to diffuse the problem, by explaining the meaningful use of language in terms of analogy, rather than rules. I will indicate the points of contact between the skeptical solution that Kripke outlines, and the analogical account I propose. I will suggest that the two accounts are compatible and complimentary. Kripke's solution has it while there is not some fact that demonstrates that someone meant, say, *plus*, by the word 'plus' (and thus that a particular rule was being followed), it is still assertable to say that someone does mean *plus* by 'plus', and thus that they are following a rule, insofar as doing so has a useful role in our lives as a community of language users. I will then return to the idea raised at the end of section 1, that even the words 'analogy' and 'rule' are to be understood analogically, on the analogical account, and thus that it may be useful for us, in certain circumstances, to categorise meaningful language use as 'rule-governed', after all. However,

I will claim that if we are to speak of meaning rules, then these rules must be quite unlike various other rules and rule-governed practices that we regard as typical. I will claim that analogy is the fundamental principle that we should take to govern meaningful language use, whether or not there is also scope to speak of ‘rules’ of one kind or another, in certain circumstances.

### The Rule-Following Paradox

In S201 of his ‘Philosophical Investigations’, Ludwig Wittgenstein raises a paradox about rule-following: “this was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule”.<sup>120</sup> In ‘Wittgenstein: On Rules and Private Language’, Saul Kripke develops this paradox so as to challenge the idea that we could ever mean something by a word or linguistic expression. An assumption that is built into the way that Kripke presents the paradox is that in order to mean something by a word (or atomically meaningful linguistic expression – hereafter, for brevity: ‘word’), it must be the case that we are following the rule for the meaningful use of that word. Furthermore, it is assumed that in order to use a word meaningfully in the present, we must accord with our past intentions regarding the meaningful use of that word<sup>121</sup> – we must follow the same rule in the present as we were following in the past when we used that word meaningfully. If we cannot determine what rule we were following in the past, then we cannot determine how we should use that word in the present if we wish to accord with our past intentions. It then becomes evident that the meaningfulness of our present usage must also be thrown into doubt, since later we can look back on what was now our present usage, and still we will not be able to determine what rule, if any, we were following. Seeing as our past usage of any given word forms only a finite sequence, however we go about using that word in the present (be it in an intuitively ‘meaningful’ way, or at random), we will not be able to find any evidence either that we are following the same rule or that we are not. To echo Wittgenstein: ‘every course of action can be made out to accord with the rule’ – and furthermore, every course of action can be

---

<sup>120</sup>Wittgenstein(1996),p.81

<sup>121</sup>Kripke(1993),p.9

made out *not* to accord with the rule. We are left with no criterion to determine whether any given instance of language use is meaningful. If meaning something by a word is a kind of rule-following, then our supposedly meaningful use of language in the present becomes “an unjustified stab in the dark”<sup>122</sup>, and the idea of meaning itself “vanishes into thin air”.<sup>123</sup> I will now outline Kripke’s presentation of the paradox in terms of a mathematical example – the addition function.

Kripke begins with the example of the word ‘plus’ (and the ‘+’ symbol), which we may suppose denotes the addition function. He points out that seeing as each one of us has performed only a finite number of addition problems, for each of us there is some number, such that we have never added any numbers larger than that number, stipulating, for ease of exposition, that we are to imagine that that number is 57.<sup>124</sup> Supposing that is so, he asks us how we know that we should answer ‘125’ rather than ‘5’, when we are presented with the novel addition problem ‘68+57’.<sup>125</sup> He is not intending to raise a skeptical doubt about mathematics, but about the meaning we attach to the word ‘plus’ and the ‘+’ symbol. Given we have only applied any word or symbol, including ‘plus’ and ‘+’, a finite number of times, he asks how it is that we know that when we apply this expression to some new instance (say ‘68+57’), that we are following the same rule in the present as we were in the past. The problem is that any finite sequence of instances could have been generated by an arbitrarily large number of distinct rules - rules which may diverge in their application once the sequence is extended, by being applied to some novel instance. Kripke asks how we know that ‘plus’ (‘+’) as we used it in the past denoted addition, rather than quaddition (a function denoted in our discussion by the ‘ $\oplus$ ’ symbol). Quaddition is defined like this:  $x \oplus y = x + y$ , iff  $x, y < 57$ , and always = 5, otherwise.<sup>126</sup> Seeing as we have each only performed finitely many addition problems, and never with numbers larger than 57 (by stipulation), for all we know when we used the word ‘plus’ in the past, it could have denoted the *quus* function rather than

---

<sup>122</sup>Kripke(1993),p.17

<sup>123</sup>Kripke(1993),p.22

<sup>124</sup>Kripke(1993),p.8

<sup>125</sup>Kripke(1993),p.8

<sup>126</sup>Kripke(1993),p.9

the *plus* function, since either function is entirely compatible with our past history of behaviour, mental states, and linguistic usage.

Kripke points out that we cannot evade the problem by appealing to more ‘basic’ rules. For example, someone may claim that using the word ‘plus’ meaningfully requires you to follow some kind of algorithm. Kripke suggests the following: “suppose you wish to add  $x$  and  $y$ . Take a huge bunch of marbles. First count out  $x$  marbles in one heap. Then count out  $y$  marbles in another. Put the two heaps together and count out the number of marbles in the union thus formed. The result is  $x + y$ .”<sup>127</sup> Someone might suggest that understanding a formula for addition along these lines (though perhaps more sophisticated) is what meaning addition by ‘plus’ consists in. However, Kripke points out that just like ‘plus’, we have applied the word ‘count’ only finitely many times, and our past usage of ‘count’ is equally compatible with the hypothesis that we meant *quount*, as it is with the hypothesis that we meant *count*. ‘Quount’ is defined as “to count it in the ordinary sense, unless the heap was formed as the union of two heaps, one of which has 57 or more items, in which case one must automatically give the answer ‘5’”.<sup>128</sup> This indicates the generality, and the inescapability of the problem – we cannot appeal to some other sequence of words, to some definition or formula, in order to demonstrate that we did mean some determinate function by ‘plus’ in the past – for we have only applied each and every word a finite number of times, and our past usage is compatible with these words; composing the definition, instantiating any of an indefinite number of distinct rules as well. As such, when such a definition is offered, we can simply interpret the words composing these definitions in a non-standard way – in particular so as to conform to the *quous* rule, and therefore they can be no help in demonstrating that we did mean *plus* by ‘plus’ rather than *quous*. Kripke points out that further attempts along these lines would be equally futile, for example – if ‘counting’ must be “independent of the composition of a heap in terms of sub-heaps”, we can interpret

---

<sup>127</sup>Kripke(1993),p.15

<sup>128</sup>Kripke(1993),p.16

‘independent’ as *quindendent* – which works out compatibly with the *quus* rule.<sup>129</sup> The upshot is that we cannot determine that we meant *plus* by ‘plus’, rather than *quus*, in the past, and appeal to more basic definitions offers no escape route.

If we can’t determine that we meant addition by ‘plus’ (rather than quaddition or any number of other functions that agrees with addition in terms of past instances, but that could diverge in the present), then the idea that we mean something by the word ‘plus’ in the present falls into question also, seeing as later we can ask of what was our present usage how we know which function we were following (perhaps with regard to various functions that diverge from addition after 58, for example). If we cannot determine whether or not we are following the same rule in the present as we were in the past, then our present usage is deprived of any foundation, not least because it will be (as Alexander Miller writes) “up for retrospective viewing in the future”<sup>130</sup>, leading us to conclude that we don’t know which rule we were following then, either. To solve this paradox, which threatens to undermine the very idea that language can be used meaningfully, Kripke asks us to produce some fact that “constitutes”<sup>131</sup> meaning addition, rather than some other function, by ‘plus’. He also insists that this fact must show that we are “justified”<sup>132</sup> in, for example, giving the answer ‘125’ to the addition problem ‘68 + 57’, for the meaning of ‘plus’ is supposed to dictate how we should respond, not necessarily how we will, in fact, respond.

Kripke points out that “the relevant skeptical problem applies to all meaningful uses of language”<sup>133</sup>, and spells this out by considering what our meaningful use of the word ‘table’ could be thought to consist in. Kripke asks us how we know, when we apply the word ‘table’ to some newly encountered object (say a piece of furniture with a flat top and legs, at the base of the Eiffel Tower, which we are visiting for the first time), that we are following the same ‘table’-rule now as we were in the past when we applied the

---

<sup>129</sup>Kripke(1993),p.17

<sup>130</sup>Miller (2002)p.2

<sup>131</sup>Kripke(1993),p.11

<sup>132</sup>Kripke(1993),p.11

<sup>133</sup>Kripke(1993),p.7

word ‘table’ to objects. Seeing as each of us has encountered only a finite number of objects that we have assigned to the category ‘table’ in the past, if meaning something by ‘table’ is to be understood as following the rule for the meaningful use of that word, then how do we know that in calling this newly encountered object a ‘table’, we are according with our past intentions that led to us designating things we encountered in the past ‘tables’? Kripke suggests that for all we know, we might have been following a rule for the concept *tabair*, where a *tabair* is a table that is “not found at the base of the Eiffel Tower, or a chair found there”<sup>134</sup>. Seeing as we have never visited the Eiffel Tower before, every past instance we have assigned to the category ‘table’ is equally compatible with the *tabair* rule as it is with any candidate *table* rule. The point, again, is that any finite sequence of instances is compatible with the hypothesis that it was generated by any of an arbitrarily large number of rules, and Kripke is asking us to produce some fact that establishes that we were following the rule for the concept *table* when we used the word ‘table’, rather than some other rule like the rule for *tabair*.

One way of attempting to escape this might be to offer some candidate definition for ‘table’, something that our past use of the word ‘table’ was derivative on, which would exclude the possibility that this object at the base of the Eiffel Tower cannot be designated a ‘table’, in accordance with our past intentions regarding the application of that word. However, suppose we use the definition suggested earlier (setting aside its inadequacies) “a piece of furniture with a flat top and legs, for eating, writing, or working at”.<sup>135</sup> Someone might claim – ‘well, this definition makes no explicit reference to the Eiffel Tower – so there! That’s how I know to call it a ‘table’’. But of course, the words composing this definition have only been applied a finite number of times – and perhaps by ‘furniture’ in the past we meant *furnair*, where a piece of *furnair* is a piece of furniture not found at the base of the Eiffel Tower, or an owl found there. If someone tries to evade this possibility that the rules we were following in the past excluded particular objects found at the base of the Eiffel Tower, by including a ‘whichever location they

---

<sup>134</sup>Kripke(1993),p.19

<sup>135</sup>Little Oxford English Dictionary, Eighth Edition, Oxford University Press, 2002, p.717

are found in' clause, we could of course insist that perhaps in the past by 'location' they meant *quocation*, where 'quocation' donates a place that is not at the base of the Eiffel Tower. This indicates the generality of the problem – just as we cannot determine what rule we were following in the past (and even whether we are following a particular rule in the present) in our use of the word 'plus', nor can we with the word 'table', or 'furniture', or 'location', or with any word whatsoever. If we cannot produce a fact that our meaning something by these words consists in – a fact which respects the idea that the meaning of words should justify our application of them – then we will have to give up the idea that language can be used meaningfully, which is a deeply troubling result.

### The Analogical Solution

'Straight' solutions to the rule-following paradox attempt to cite some fact that establishes that, when we used any given word meaningfully in the past, we were following a rule, and in using that same word meaningfully in the present, we are following the same rule. However, the account I wish to offer, in terms of analogy underlying the meaningful application of words, is not a 'straight' solution but a 'skeptical' one, insofar as it does not conceive of meaningful language use as a rule-following practice. So there is, on my account, no fault in the reasoning of the rule-following paradox – it's true that our attempts to find some fact that establishes that we're following the same rule every time we use some word meaningfully will be in vain. According to the analogical solution I propose, using words meaningfully in the present depends on extrapolating from our past experience by analogy. As such, there is not some fact that establishes that we mean the same thing now, as we did in the past, it is just that we apply words based on perceived resemblance to that we have encountered, and categorised using some linguistic label, in the past. I will now layout this account in more detail.

The fundamental point is that when we use words meaningfully by applying them in, or to, some newly encountered instance, this meaningfulness consists in our extrapolating from our past experience of using this word and experiencing its usage, and we extrapolate in this way by drawing an analogy.

According with our past intentions regarding the meaningful use of a word in the present, consists in seeing an analogy between a past instance, or past instances, and the present instance. So it's not that we were tacitly or otherwise following some precise rule in the past, which we must also follow in the present, if we wish to accord with our past intentions. To use a word meaningfully is not to do or think something (or even to be disposed to do or think something) that determines whether or not that word applies in, or to, each and every potential instance, in each and every possible world. It is bizarre to think that an individual token of a word, or even a sequence of such tokens, could have a potentially infinite (or merely arbitrarily enormous) number of consequences in this way. I will now spell this out, by looking at a number of examples.

If meaning something by a word is understood as following the rule for the meaningful use of that word, and the only evidence we have for what rule we have been following in the past is a finite sequence of instances, then, as Kripke highlights – when we encounter some item of furniture at the base of the Eiffel Tower for the first time, we cannot determine that in calling it a 'table' we are according with our past intentions regarding the use of that word (that we are following the same rule). So the question arises – what is it that makes our calling such an object a 'table', more than a mere leap in the dark – what makes our use of that word in this newly encountered instance meaningful? Well, it is simply that we see an analogy between this object and past objects that we have assigned to the category denoted by the label 'table'. We adopt a point of view on this object that is determined by its perceived similarity to objects we have encountered in the past, and, in this vein, assign it to the same category. Still, it might be thought – even if we unproblematically assign this object to the category *table* in this way, suppose the next day we visit the Arc de Triomphe and see an item of furniture at its base – how do we know than in calling this object a 'table', we are doing the same thing we did yesterday when we called the item of furniture at the base of the Eiffel Tower a 'table', by analogy? How do we know that we are making the same kind of analogy, continuing on in the same way, seeing things as similar to objects encountered in the past in the same way today as we did yesterday? I will now address this concern, for if it cannot be met, all the

analogical account of meaning seems to do is to reframe the paradox, without getting us any closer to a solution.

When I claim that we categorise objects by seeing an analogy between freshly encountered instances and objects we have assigned to that category in the past, I don't mean to say that we are following some kind of rule which determines in what sense two or more objects must resemble each other in order for them to count as similar. It is not just that we are not following the rule for the meaningful use of each word when we use that word meaningfully – nor are we following some general rule which determines what it is that makes distinct objects fall under the same linguistic label. If there was a rule for what was to count as similar when it comes to adding a newly encountered object to a category, containing various objects encountered in the past, then it would be a short step to actually defining what it is to belong to each and every given category by stating some rule – objects are 'tables' insofar as they resemble each other in respects A, B, and C, to at least degree D. This would create more problems than it would solve – bringing in skeptical doubts about what it is for two objects to resemble each other in respects A, B and C, respectively – and about how we know that we are following the same rule for determining the degree to which objects resemble each other, in each of these respects, as we were previously. I think that a good response to the rule-following paradox cannot be had simply by appealing to more and more rules, each one more fine-grained than the last – for as Wittgenstein writes “explanations come to an end somewhere”<sup>136</sup>, and as Kripke stresses, in appealing to more and more basic rules, we do nothing to address the question of what the meaningfulness of our use of language consists in. No matter how fine-grained a rule, it has only been applied a finite number of times, and this finite sequence of applications is compatible with having been generated by an arbitrarily large number of distinct rules, rules that would diverge in their respective applications to future instances. So how can we have meaningful use of language, driven by analogy-making, that does not appeal to the notion of rule-following? I will now turn to this.

---

<sup>136</sup>Wittgenstein(1996),p.3

On my account, it is best to understand the making of analogies not as any kind of rule-following, but, rather, as some kind of irrepressible instinct. We cannot help but, as Hofstadter and Sander put it: “to make sense of the new and unknown in terms of the old and known”<sup>137</sup>. When we encounter something unfamiliar, we try to make sense of it by drawing on our past experiences, namely those past experiences that strike us as similar to this newly encountered object in some way. As Willard V. Quine writes (in ‘Ontological Relativity’), we can understand our capacity for language (and our underlying capacity for categorisation) as relying on “an inborn propensity to find one stimulation qualitatively more akin to a second stimulation than to a third”<sup>138</sup>. This is simply how we learn from experience, we see things as similar in increasingly salient ways, which allows to interpret future experiences in a more useful and effective way. There is no rule for what it is for two objects to be similar such that they merit being called by a common linguistic label – it’s just that each and every one of us has an innate capacity to see objects, each of which is fundamentally unique, as resembling one another in various ways. Furthermore, each of us undergoes linguistic training – we are taught to make similar analogies to other members of our linguistic community – and perfectly legitimate analogies that each of us makes in the process of language acquisition (e.g. calling a dog a ‘cat’) are met with negative feedback. We are corrected when the analogies we make, in applying linguistic labels, are far off those that other members of our linguistic community are inclined to make, such that, once we are competent speakers, we tend to make pretty similar analogies to one another, at least when it comes to calling ordinary objects by a common linguistic label. So, calling a table a ‘table’ relies simply on our exercising our innate capacity to see analogies – a capacity which we have been trained to exercise, in the case of applying the word ‘table’, in a particular way, by being shown various objects and told they were ‘tables’, and thereafter, being corrected if we called something that others would not call a ‘table’ a ‘table’.

---

<sup>137</sup>Hofstadter&Sander(2013),p.3

<sup>138</sup>Quine(1968),p.188

There is no one thing that all tables have in common, which we can cash out by stating some general rule about what it is to use the word 'table' meaningfully. We need not even assume that the members of a given community of speakers would end up calling the very same objects 'tables', if they were shown some long list of candidate-tables, various members of which were far from prototypical. But in our day-to-day use of language, each of us draws on our own personal stock of experiences of calling objects 'tables', and we underwent some degree of training when we acquired this word – in that we were shown some objects, told they were 'tables', and corrected if we called some objects that those around us would not be inclined to call 'tables', 'tables'. So, we each have our own set of experiences that informs our use of any given word. But, there are commonalities, both between these sets of experiences, and between the manners in which we were corrected and trained to use each given word.

Whilst we do tend to resemble other members of our linguistic community in what it is that we see as similar, when it comes to applying ordinary words to objects by extrapolating from our past experience – this tendency to see resemblances should be seen as an innate capacity, an irrepressible instinct that has been fine-tuned by our interaction with other members of the community, rather than some general rule that informs what we should call newly encountered objects, based on some formula for what it is for distinct objects to count as relevantly similar. So when we encounter that table at the base of the Eiffel Tower – we call it a 'table' by exercising this innate capacity to see resemblances between the present and the past, and to apply linguistic labels on the basis of these resemblances. I think we needn't worry about questions of how we know that we are exercising our analogical capacity in a consistent way, since, I'm assuming that we are competent speakers of the language, who have been accepted into the linguistic community as such, given that our use of language tends to reflect the kind of resemblances seen, not only by ourselves, but by other members of our linguistic community.

As outlined earlier, Kripke asks us what fact it is about our past usage of the word 'plus' that establishes that, when we encounter a novel addition problem involving numbers larger than any of those

we have previously added together before – by stipulation a number larger than 57 – that we should answer in some fixed way (say ‘125’ to the problem ‘68 + 57’). Using words that refer to mathematical functions would seem surely, if any use of language is, to be an instance of rule-following. Yet, as also discussed previously, it seems very hard to cite some fact that proves that when we said ‘plus’ in the past we meant *plus* rather than, say, *quus*. Phillip Pettit says something, in the course of offering a very different kind of account of meaning, and response to the rule-following paradox, which may prove helpful in getting an analogical account, of what it is to mean *plus* rather than *quus* by ‘plus’, off the ground. Pettit suggests that while each of us has performed and seen only a finite number of addition problems in the past – while it is indeed the case that this finite sequence of examples instantiates many different rules – including the rule for *quus*, it just so happens that they only exemplify one rule – the rule for *plus*<sup>139</sup>. Pettit is in fact advancing an account of meaning on which meaning is not reducible to some other notion, or some other set of facts, and is offering a straight solution to the rule-following paradox. Pettit claims that while instantiation is a two-place relation that holds between a rule and a set of examples, exemplification is a three-place relation that holds between a rule, a set of examples, and a person<sup>140</sup>. This is certainly an interesting suggestion, and indicates the importance of the individual involved in acquiring some linguistic expression, when it comes to them extending their usage to new instances. However, while I am advancing a very different theory, I think his suggestion is useful at least in getting started – we’ve seen a bunch of examples of addition problems, and seen the word ‘plus’ used meaningfully in these instances – and when we apply the word ‘plus’ in some new instance, it just so happens that we see an analogy from these past instances that goes along with the addition function rather than the quaddition function. Perhaps in creatures and communities quite unlike ours, they would indeed see the quus-analogy, rather than the plus-analogy, that members of our community tend to see.

However, I think a little more can be said than simply that the examples of addition we have seen lead

---

<sup>139</sup>Pettit(2002),p.197

<sup>140</sup>Pettit(2002),p.197

us to see the plus-analogy rather than the quus-analogy – in fact a little more must be said, seeing as, on my account, to mean something by a word, even a mathematical function like ‘plus’, is not being construed as following the rule for the meaningful use of that word. Well, for one thing – when we do add together ‘68 + 57’ for the first time – we are surely seeing analogies between this novel addition problem and past addition problems that we have already encountered – namely  $8 + 7$ , and  $6 + 5$  (+1 that we carry, of course). Surely, we cannot overlook the fact that we have performed additions using these very numerals before, making it more than natural for us to see the analogy with the addition function, and eventually come out with the answer ‘125’, rather than ‘5’. We are taught to add the individual numerals, to carry the tens, etc., and in performing this novel addition problem, we are bound to see analogies to our past experiences in these ways. While we have each only performed addition problems involving numbers below a certain size, we are all familiar with each of the individual numerals 0-9, and with combining them to form multi-digit numbers. Surely, the fact that we have each, presumably, added any pair of these individual numerals in the past, means we are inclined to see plus-analogies rather than quus-analogies between these addition problems and a novel problem involving numbers larger than those we have ever added before. In a sense, we are merely recombining addition problems we have previously performed, and so it seems natural for us to see the plus-analogy rather than the quus-analogy.

### The Skeptical Solution

Kripke lays out the ‘skeptical solution’ to the paradox, which he interprets from Wittgenstein. According to this solution, there is no fault in the paradoxical reasoning that led us to the bizarre conclusion that all language is meaningless – and there is no “superlative fact” about someone’s mind that establishes that they really did mean *plus* by ‘plus’, or table by ‘table’, and “determines in advance what (they) should do to accord with this meaning”<sup>141</sup>. According to the skeptical solution, it is a “philosophical misconstrual”<sup>142</sup> that our meaningful use of expressions requires there to be some fact in

---

<sup>141</sup>Kripke(1993),p.65

<sup>142</sup>Kripke(1993),p.65

our mind that establishes that we are following one rule rather than another. While we cannot respond to the skeptic by citing some fact that proves that we meant *plus* rather than *quus* by ‘plus’ in the past, on this solution, our meaningful use of language does not require this kind of justification. However, according to this solution, it is perfectly proper to say “the fact that Jones meant addition by such-and-such a symbol”<sup>143</sup> – it is just that there is not some superlative fact in Jones’s mind that establishes that this was so, and it is only Jones’s participation in a community of speakers that makes such phrases appropriate. One consequence of the skeptical solution is the impossibility of private language: “the skeptical solution does not allow us to speak of a single individual, considered by himself and in isolation, as ever meaning anything.”<sup>144</sup>

According to the skeptical solution, the reason why it is appropriate to say things like ‘Jones means addition by ‘plus’ and ‘Jones means the same thing by ‘plus’ now as she did in the past’, is not that there is some fact about Jones that makes this the case, but just that, in some circumstances such statements are assertable, and asserting them in such circumstances has “a role in our lives”<sup>145</sup> as a community of speakers. The skeptical argument is meant to establish that when we do seem to ‘follow a rule’, like the rule for ‘plus’, or the rule for ‘table’, we “act unhesitatingly but blindly”<sup>146</sup> – insofar as there is not some fact in our mind that our ‘rule-following’ consists in, but we automatically respond confidently in ordinary cases. This is why it is said to make no sense to say of an individual, considered in isolation, that they mean something by their words, insofar as they are just following a brute inclination, and without a community to refer to (to quote Wittgenstein): “whatever is going to seem right to me is right”<sup>147</sup>. It is only by reference to a community of speakers that we have a standard of correctness against which to judge the propriety of someone’s purportedly meaningful use of linguistic expressions. Kripke discusses the example of a teacher checking whether a child is following the rule for addition. The teacher will

---

<sup>143</sup>Kripke(1993),p.69

<sup>144</sup>Kripke(1993),p.68-69

<sup>145</sup>Kripke(1993),p.79

<sup>146</sup>Kripke(1993),p.87

<sup>147</sup>Wittgenstein(1996),p.92

judge that the child is following the rule, that she does mean *plus* by ‘plus’, if she produces the right answer for small addition problems (that is: “the same answer that he [the teacher] himself would give”<sup>148</sup>) and for larger addition problems the child must be judged to be “applying the ‘right’ procedure” (that is – the same procedure the teacher “is inclined to apply”) even if she “comes out with a mistaken result”<sup>149</sup>.

Kripke points out that “something similar is true of adults”<sup>150</sup>. We judge others to mean *plus* by ‘plus’ if their answers to addition problems agree with the answers we ourselves would give, or at least, the procedure they follow (even if they make a mistake) is similar to the procedure we ourselves are inclined to follow. Similarly, we judge others to mean table by ‘table’ if they call the same objects ‘tables’ that we are inclined to call ‘tables’, or at least, when they call something a ‘table’ that we would not be inclined to call a ‘table’, we can see where they’re coming from (maybe it’s a borderline case of a table, that people may have divergent intuitions about). If someone’s inclinations to use a given linguistic expression agree with other members of the linguistic community “in enough concrete cases”<sup>151</sup>, we will judge them to mean the same thing by that linguistic expression that we do. The meaningfulness of our expressions (and the propriety of sentences like ‘Jones means table by ‘table’) is rooted in the agreement of the linguistic inclinations of the members of our community. As Kripke puts it: “The [skeptical] solution turns on the idea that each person who claims to be following a rule can be checked by others”<sup>152</sup>. Kripke points out that if someone’s inclinations to use words does not agree with that of other members of their community at all, then they will be regarded as an “incorrigible deviant” who “cannot participate in the life of the community and in communication”<sup>153</sup>. We judge someone to be a competent speaker when they have the same, or a very similar, brute inclination to apply words to novel instances as other members of their linguistic community, ourselves included (assuming we are competent speakers).

---

<sup>148</sup>Kripke(1993),p.90

<sup>149</sup>Kripke(1993),p.90

<sup>150</sup>Kripke(1993),p.90

<sup>151</sup>Kripke(1993),p.91

<sup>152</sup>Kripke(1993),p.101

<sup>153</sup>Kripke(1993),p.92

I think this is where the relevance of the analogical solution I have proposed to (Kripke's) Wittgenstein's skeptical solution comes in – we act unhesitatingly when applying some word in a new instance, because we make an unconscious analogy to our past experiences of that word. We cannot help but automatically make sense of a novel situation in terms of our past experience, by analogy. While it is perfectly legitimate to talk about our meaning a definite thing by some word, and such talk has a useful role in our lives, all we are doing when we act unhesitatingly but blindly in applying a word in a novel instance is extrapolating from our past experience. To say that someone means *plus* by 'plus', or table by 'table' is to say that we regard them as a competent speaker (at least when it comes to using these expressions) – which consists in them making similar, or analogous, analogies to other members of the linguistic community when it comes to applying linguistic expressions. The analogousness of our analogies is rooted in the members of a linguistic community each having a similar stock of experiences to other members of the community on which to draw, when applying a word in some new instance (we all learnt addition by performing many of the same addition problems, and we've seen a similar bunch of tables, and so on). This is why it doesn't matter that there's not some superlative fact that our meaning a specific thing by a linguistic expression consists in – assuming we are competent speakers, our applications of linguistic expressions agree with those of other members of our linguistic community in enough cases - and this is in part because the analogies we make are rooted in a common set of experiences (or, at least, analogous sets of experiences). We share, as Wittgenstein puts it, a "form of life"<sup>154</sup>. This consists, not in some objective fact that establishes that we follow the same rules, but just in the fact that our inclinations to use linguistic expressions generally agree. I will now discuss what we are to say about people who apply words in bizarre quus-like ways, who call owls 'tables', and respond to '68 + 57' with '5'.

One potential benefit of an account of the meaning of linguistic expressions that advocates the idea that meaning something by a word is to follow the rule for the meaningful use of that word, is that when

---

<sup>154</sup>Wittgenstein(1996)p.11

people use language in bizarre quus-like ways, we can explain what it is that they're doing wrong: they're not following the rule. However, I think the analogical solution can come up with an appropriate explanation of such problem cases, also. It just so happens that the members of a linguistic community do, in general, tend to see similar analogies when it comes to applying linguistic labels to ordinary objects, and this is, in part, because of the kind of creatures we are, and, in part, because we have similar stocks of experiences on which to draw when making analogies by applying terms in the present. But suppose that there was someone who made analogies in applying linguistic expressions to objects in utterly bizarre ways, such that the rest of us could not understand their word-choices - say if they called something the rest of us would call a 'table' a 'shoe', and went on in this bizarre way in all of their linguistic practice. Obviously we would not regard them as a competent speaker, and we would not take their words at face-value - but the challenge for the analogical account is to say what, if anything, they are doing wrong, if we can't appeal to the idea that they're not following the rules. As suggested previously, we are taught to make similar analogies to other speakers in the process of language acquisition, and it seems that, if someone was always to apply words in bizarre quus-like ways, we would not regard them as having acquired language, they would merely be making sounds, imitating words like parrots are known to, but without meaning anything by their words. The analogies that underlie our decisions to apply ordinary words to objects we encounter are public, in the sense that if only one person was able to see an analogy (say a 'dog'-analogy, or a 'plus'-analogy), and others couldn't see at all where they were coming from, we would not see this as a meaningful application of a word. To apply a word meaningfully to some object requires that other members of the community be able to understand your word-choice, even if they don't quite agree with it (hence, we may disagree about whether something is a 'good book', without questioning whether someone else understands what 'good book' means). So to be a competent speaker requires not only that you make some kind of analogy when you apply words, but that other members of the community accept your language usage as meaningful, rather than merely the random imitation of sounds you have heard.

The analogical solution that I have proposed to the rule-following paradox is best understood as an attempt to flesh out the skeptical solution. The skeptical solution tells us that there is no superlative fact that we can cite to establish that we are following a particular rule when we use language meaningfully. The analogical solution claims that when we apply words meaningfully in some newly encountered instance, we do this by drawing an analogy with our past experience of applying such linguistic labels. When someone makes bizarre quus-like analogies, other members of the linguistic community will not be able to understand where they are coming from, it will seem like they are just imitating sounds, rather than making analogies. While what these people are doing wrong cannot be spelled out in terms of their failing to obey a rule, we can simply say that we will not regard them as competent speakers, we will not take their word for it, and it will seem legitimate to say that they do not mean *plus* by ‘plus’, or table by ‘table’, or dog by ‘dog’. Saying such things will have a purpose – namely to say that when these people apply words to newly encountered objects, they are not making anything like an analogy, as far as we can see, and so their usage should not be taken to be meaningful.

### An Analogical account of ‘Rules’

On the skeptical solution outlined by Kripke, while there is no superlative fact we can cite to demonstrate that someone meant, for example, *plus* by ‘plus’, in some circumstances it is assertable to say that someone did mean *plus* by ‘plus’ – and, because the rule-following paradox is premised on the idea that meaning something by a word does consist in following a rule, it may be assertable to say that someone is thus following the plus-rule. This is possible even in the absence of some fact that demonstrates that they really are following the rule. In providing an analogical solution to the rule-following paradox, I have emphasized that meaning something by a word can be best characterised as being governed by analogy, rather than rules. However, as I indicated at the end of the chapter 1 – the analogical account of meaning is universal in its scope – every word that has a use in a language has a meaning in that language, and such meaning is said to be governed by analogy. Thus – even the key terms in the discussion – such as ‘rule’ and ‘analogy’ are said to be governed by analogy, rather than rules. If

the word ‘rule’ is governed by analogy, rather than rules, then this means that there are not necessary and jointly sufficient conditions for what it is for something to be a ‘rule’ – rather, what all ‘rules’ have in common, is a series of analogical resemblances between instances. Therefore, it is possible for an analogical account to make room for the idea that we can speak of meaning-rules, after all, insofar as we may perceive that that which governs meaningful language use, analogically resembles some of those other things we call ‘rules’. In arguing that meaning something by a linguistic expression should be thought of as governed by analogy, rather than rules, I am not ruling out the idea that we may have some cause to speak of ‘rules’ of meaning, but just that, if there are such rules of meaning, these rules are very unlike many of the things we do typically refer to as ‘rules’.

For example, we speak of the ‘rules’ of chess – and these rules determine, precisely and unambiguously, for each point in the game of chess, just which moves are legal, and which moves are illegal. Thus, the rules determine how each piece may be moved, and in some cases, which pieces may be moved - thus if someone’s King is in check, the rules determine that they must either move the King out of check, or move another piece so as take the King out of check, and that no other move is legal. It does not seem that if we speak of ‘rules’ of meaning, that we could be speaking of rules very much like the rules of chess. If there are rules of meaning, it doesn’t seem that these rules precisely and unambiguously determine how expressions may be used ‘legally’ and ‘illegally’, given that, as I indicated in chapter 1, the categories corresponding to at least many of our words lack sharp boundaries. This would be analogous to a game resembling, though very different from, chess, in which, while some moves were clearly legal, and other moves clearly illegal, other moves were neither clearly legal, nor clearly illegal. Another significant feature of the rules of chess, is that if someone does not follow the rules correctly (thus if a piece is moved in an ‘illegal’ way), they cannot be said to be playing chess. Conversely, if there’s a rule governing the word ‘plus’, it’s not the case that if someone does not follow the rules correctly (and so produces an incorrect answer to an addition problem), that they could no longer mean *plus* by ‘plus’ – the rules of meaning are meant to be such that they tell us how we should apply the word,

not how we will in fact apply the word. One can follow a rule of meaning, and make an error, but one can't follow the rules of chess (which govern which moves are legal, and which are illegal) and make an error. So there are significant disanalogies between the rules of chess, and any putative 'rules' of meaning – the rules of chess are precise and unambiguous, and one cannot follow these rules, but make an error, unlike any putative rules of meaning.

Another rule-governed practice with which we can compare meaning something by a linguistic expression, is a sport, such as football. The rules of football govern in just what circumstances, a set piece, or a foul, or a goal, may be awarded, as well as the duration of the match, and the final result. However, unlike the game of chess, a sport such as football requires there to be a referee – somebody who applies and interprets the rules in any specific game, who is meant to be impartial and authoritative. Clearly there is not one single 'referee' who applies and interprets the rules of meaning. However, it does seem to be the case, that for some linguistic expressions, we will regard particular people as more authoritative when it comes to applying any putative rules than others. For example, as Hilary Putnam has observed, while ordinary speakers of our language can use the expressions 'elm tree' and 'beech tree' meaningfully, they may not be able to tell the difference between elms and beeches – and when it comes to making such distinctions, we tend to defer to the experts, those who are specialists and so are better able to distinguish between cases in which a tree should be categorised as an 'elm' tree and cases in which it should be categorised as a 'beech' tree.<sup>155</sup> Thus, one analogy between the rules of a sport, such as football, and any putative rules of meaning, is that we rely on specialists to apply (some of) these rules. On the other hand, in a sport such as football, we rely on the referee to apply the rules all the time, whereas for many of our words, such as 'table', 'green', etc., we will not need to defer to any external authority when applying these rules (although perhaps we are all experts when it comes to applying these rules, as competent speakers). Another significant disanalogy between the rules of a sport, such as football, and any putative rules of meaning, is that in a sport such as football, we expect those who apply

---

<sup>155</sup> Putnam(1975),143

the rules to be able to state the rules. Yet, as was demonstrated in the chapter 1, just because someone can use a word meaningfully, we shouldn't expect them to be able to state some rule, which determines precisely how the word may be used. Thus – there are significant disanalogies between the rules of a sport, such as football, and any putative rules of meaning, even though there is some analogical resemblance between the two, insofar as both practices rely on experts to apply (some) of the rules.

Another body of rules we may wish to consider, are the rules of a society, the laws of the land. These rules determine whether or not somebody's actions should be considered illegal, and if they should be considered illegal, how this person should be punished or sanctioned in light of this. On the face of it, there clearly don't appear to be rules which govern how somebody should be sanctioned if they use a word in a way that breaches a rule. However, if somebody was to consistently use a word in a bizarre quous-like way, we would not regard them as competent speakers, just as, perhaps, people who consistently don't follow the rules of the law are not regarded as law-abiding citizens, but as criminals. However, the rules of law are generally speaking explicitly stated – and as I have indicated, we have a very hard time explicitly stating any putative rules of meaning. There are other significant disanalogies – the putative rules of meaning are supposed to govern each and every meaningful use of a linguistic expression – whereas it doesn't seem that ordinary, legal actions (such as making a sandwich) are governed by legal rules, even if such rules deem such actions to be legal. It is only certain kinds of action, about which the law has anything to say, whereas the rules of meaning are supposed to govern every action within the domain of meaning something by a linguistic expression. Thus, there seem to be significant disanalogies between legal rules and any putative rules of meaning.

However – there is one other useful similarity between legal rules, and their application, and meaning something by a linguistic expression. In case law, particular legal judgements that have been made in the past, are said to establish a precedent, such that, thereafter, courts will rule on cases they judge to be similar, in much the same way. For example, in U.S. law, there was a 1973 ruling referred to as 'Roe vs

Wade', which, in one particular case, ruled that a woman, whose life was not endangered by her pregnancy, could legally have an abortion, which had not been legal before that time. This case established a precedent, such that in future, abortions were, generally speaking, judged to be legal up until a certain point in the pregnancy, even if the woman's life was not in danger. Thus a particular judgment, about a particular woman and her pregnancy, established a precedent, which determined how courts would rule in future. An analogy to the legal precedents that are established in case law, may be useful to the analogical account of meaning. On the analogical account of meaning, objects are referred to meaningfully by an expression, because they are seen to analogically resemble other objects that have in the past been referred to by that expression. Thus our past categorisations (our application of a word to some particular object) may establish a precedent, such that we will categorise future objects in a particular way, based on their perceived similarity to the precedent. Just as a legal ruling about a particular woman, could establish a precedent that would apply to lots of future women when it comes to legal judgements about abortion, so too may our categorisation of a particular object, establish a precedent, such that we will categorise other objects that we perceive to be similar to that object, in much the same way in future. It shows how our particular past judgments, rather than some abstract rule, may govern how we go about judging future cases, both in the domain of law, and in the domain of meaning.

Thus, we have seen that there are some similarities between various kinds of rule-governed practices, and that which governs linguistic meaning. However, there are also significant dissimilarities. If the word 'rule' is governed by analogy, rather than by a rule, however – then the fact there are some similarities between various kinds of rule, and that which governs linguistic meaning, may give us reason to claim, after all, that meaning could be rule-governed, in some significant way. It may be legitimate to speak of 'rules of meaning', because of the family resemblances between various other kinds of rule, and that which governs linguistic meaning. However, the analogies and disanalogies I have considered show that if there are 'rules' that govern meaning, these rules lack many of the properties that we typically expect rules to have. Many rules (such as the rules of chess) are supposed to be precise and unambiguous, and

are supposed to be such that we cannot, while following these rules, make a mistake – and neither of these properties appear to be true of meaning something by a linguistic expression. Many rule-governed practices rely on one single authority figure to apply those rules (such as the referee in a game of football), and we should expect this authority figure to be able to explicitly state the rules. However, there is no such single authority in the case of meaning something by a word, and we can follow the rules of meaning (if there are any), even if we cannot explicitly state these rules. Unlike legal rules, any putative rules of meaning are not stated explicitly, and the ‘rules’ of meaning are supposed to govern each and every meaningful use of a word, whereas it seems strange to say that legal rules govern each and every action that is considered to be legal. However – I did note an important similarity between legal judgements, and using linguistic expressions meaningfully – both may rely on precedents, on particular judgements that have been made in the past, to inform future judgements in cases that we regard as relevantly similar. Thus the comparison between law, and meaning, may demonstrate how our judgements (or categorisations) could be informed by analogy to past judgements, rather than by the application of an abstract rule.

I have argued that, since the analogical account is universal in its scope, even the meaning of the word ‘analogy’ should be accounted for analogically, rather than in terms of a rule. Thus, there are not necessary and jointly sufficient conditions for what it is for something to constitute an analogy – rather, there are just family resemblances between particular instances of those things which we call ‘analogies’. A good example of a usage of a linguistic expression that we should clearly regard as governed by analogy, is a metaphorical use of a linguistic expression. As I noted earlier, the only plausible way to account for metaphorical meaning, if there is such a thing, is in terms of analogy. Thus – a metaphorical usage of an expression relies on the perception of a similarity between distinct objects (specifically, between an object or objects that the expression is literally true of, and some other object that the expression is not literally true of). When I speak of meaning being governed by analogy, I mean analogies much like those that underlie clearly metaphorical uses of linguistic expressions. They rely on the

perception of a family resemblance between distinct objects – and given that every object, action and situation is unique, even when we use linguistic expressions literally, this usage is governed by the perception of family resemblances between objects, even if the kind of family resemblance at issue, may be rather different from the kind of family resemblance on which clearly metaphorical uses of linguistic expressions rely. Often when we think about analogy – we think about very abstract and insightful analogies – such as the analogy between the structure of an atom, and the structure of the solar system, or the analogy between the Iraq war and the Vietnam war (both of which are touched on by Hofstadter and Sander<sup>156</sup>). However, I think that even our day-to-day categorisations of ordinary objects, such as our categorisation of something as a ‘table’, relies on the perception of similarity between that object and objects we have previously encountered, and thus we may appropriately refer to all linguistic meaning as governed by analogy.

### Conclusion

In this section, I have set out the rule-following paradox, that Kripke interprets from Wittgenstein. This paradox casts doubt on the idea that our meaningful use of a word could be governed by rules, by pointing out that, for all we know, any given usage of a linguistic expression could have been governed by any of indefinitely many distinct rules. The paradox set out on the idea that meaning something by a word does consist in following a particular rule, and thus led to the conclusion that, so conceived, there is no such thing as meaning, seeing as we cannot cite some fact that our following a particular rule consists in. However, I have argued that an analogical solution to the rule-following paradox, which has significant points of contact with the skeptical solution laid out by Kripke, can preserve the idea that language can be used meaningfully, in spite of these doubts. By conceiving of meaning as governed by analogy, we can explain how it is that we go about applying a word meaningfully in the present, based on the perception of analogical resemblance to past instances, rather than by following some rule. However, the skeptical solution still states that it is assertable to say that someone meant, for example, *plus* by

---

<sup>156</sup> Hofstadter&Sander(2013)p.17, p.143

‘plus’, or even that they are following the same rule that we are, as such statements have a role in our lives as a community of language users. I explored the idea that perhaps we could speak of ‘rules’ of meaning after all – insofar as the word ‘rule’ is itself governed by analogy, rather than by a rule, and there may be notable similarities between certain things we do call ‘rules’, and that which governs the meaningful use of language. However, I argued that there are also significant dissimilarities, between many kinds of rule, and that which governs meaning, such that meaning something by a word is best characterised, not as being governed by a rule, but as being governed by analogy. This can be the case, even if the word ‘analogy’ too is understood analogically, such that there are not necessary and jointly sufficient conditions for what it is for something to constitute an ‘analogy’.

## Conclusion

In conclusion, meaning something by a linguistic expression should not be thought to consist in following a rule, but should be thought of as governed by analogy. As chapter 1 demonstrated, for many of our words, there seems to be no one thing that every meaningful instance of that word has in common. However, it is undeniable that there are similarities between particular instances of any given word, which may be characterised as ‘family resemblances’. Thus, we apply a particular linguistic expression to a newly encountered object, because we perceive an analogy, or family resemblance, between that object, and objects that we have previously experienced that word being applied to. I indicated that the analogical account is universal in its scope, and thus that it even applies to the key terms in the discussion – ‘analogy’ and ‘rule’. If linguistic expressions are applied on the basis of analogical extrapolation, rather than in accordance with a rule, then the categories corresponding to our linguistic expressions must be defined, not as the classical account has it, by formal membership criteria, stating necessary and jointly sufficient conditions for category-membership, but in some other way.

Chapter 2 explored an alternative account of categories – Eleanor Rosch’s prototype theory, which conceives of categories as being defined and structured, not by each and every instance of the category sharing one defining property in common, but by a network of family resemblance relationships between particular instances. The experiments conducted by Rosch and Mervis demonstrated that, at least many of our categories do seem to be structured in this way. While I noted that it was possible to interpret the evidence differently, I pointed out that it seems to be a very ad hoc response to the experimental data, to propose that our categories could be sharply defined in terms of the network of family resemblance relationships obtaining between them, by insisting that all members of the category do share one thing in common after all – even if it merely an overlapping and criss-crossing series of resemblances, shared by some, but not all category members. Rather, the experimental results seem to lend strong support to the analogical account of meaning – for it really does seem to be the case that there couldn’t be some rule,

stating just what it is that each member of any of the particular categories studied has in common, whereas there are undeniable analogical resemblances between particular instances.

Chapter 3 looked at how classical accounts of meaning may conceive of metaphorical uses of linguistic expressions, in contrast to literal uses. I highlighted two seemingly viable accounts – one which claims that there is no such thing as metaphorical meaning, and that metaphorical uses are distinguished from literal uses pragmatically, rather than by any difference in meaning – and another, according to which there is metaphorical meaning, but it is governed by analogy to literal instances of the expression, rather than by rules. However, I argued that the phenomenon of dead metaphors demonstrates, not only that the first of these accounts is wrong – insofar as they force us to accept that there must be such a thing as metaphorical meaning – but that the latter account is wrong too, since it insists on there being a sharp and principled distinction between using an expression literally and using an expression metaphorically. It seems that metaphors can die slowly – and thus that expressions, once used metaphorically, can, through repeated use, become less and less metaphorical, and more and more literal, over time. If metaphoricality and literality are a matter of degree, rather than a fundamental difference in kind, in this way, then it seems that both metaphorical uses and literal uses alike must be governed by the same principle. I proposed that literal and metaphorical uses alike should thus be thought to be governed by analogy, rather than rules. Both literal and metaphorical uses alike rely on the perception of analogical resemblances between objects, and the literality or metaphoricality of a particular use of an expression, characterises the kind of resemblances and the degree of resemblance between objects, that such uses rely on. I indicated that it is not only when considering the etymological history of dead metaphors that the sharpness of the literal/metaphorical distinction is thrown into doubt. As Stephen Yablo has demonstrated, even for many of the things we say on a day-to-day basis, it's far from clear whether they should be regarded as literal or metaphorical. Thus we should seek a common explanation of the literal and the metaphorical use of an expression, and only analogy can play this role.

In chapter 4, I looked at the rule-following paradox, that Kripke interprets from Wittgenstein. This paradox sets out on the idea that meaning something by a linguistic expression must consist in following a rule, but indicates that we are unable to cite some fact that demonstrates that a particular rule was being followed, for any given usage of a linguistic expression. Thus, the paradox demonstrates, that any particular usage of a linguistic expression could have been governed by any of an indefinite number of distinct rules, or no rule at all, for all we know. If meaning is conceived of as following a rule, this paradox threatens to undermine the very idea that we could ever mean anything by a linguistic expression. I proposed an analogical solution to the rule-following paradox, that explains how our meaningful use of language in the present, is dependent on our perception of analogical resemblances between past instances of a linguistic expression, and the present instance. The solution I proposed has significant points of contact with the skeptical solution that Kripke outlines. I indicated that just as the skeptical solution has it that it is assertable to speak of someone meaning *plus* by ‘plus’, or even of someone following the same rule that we follow, insofar as doing so has a useful role in our lives as a community of language users, there may also be scope to speak of ‘rules’ of meaning on the analogical account. This is because the analogical account, as I indicated in chapter 1, is universal in its scope – thus it even applies to the key terms in the discussion – ‘analogy’ and ‘rule’. Thus – it may be that we may choose to refer to meaning something by a linguistic expression as a ‘rule-governed practise’, after all, insofar as there are certain analogical resemblances between various kinds of rule-governed practise, and that which governs the meaningful use of language. However, I noted that there are also significant dissimilarities between various kinds of rule, and that which governs linguistic meaning, such that the best way to characterise meaning something by an expression, is as governed by analogy, rather than rules. This is possible even though the word ‘analogy’ too is accounted for analogically – even though there are not necessary and jointly sufficient conditions for what it is for something to constitute an ‘analogy’.

## **Bibliography**

- ‘Surfaces and Essences’ – Douglas Hofstadter and Emmanuel Sander, Basic Books, New York, 2013
  
- ‘Little Oxford English Dictionary’, Eighth Edition, Oxford University Press, New York, 2002
  
- ‘Philosophical Investigations’ – Ludwig Wittgenstein, Blackwell Publishers, Oxford, 1996
  
- ‘‘And’ and ‘But’: A Note’ – Saul Kripke, from ‘Thought: A Journal of Philosophy’, Volume 6, Issue 2, June 2017
  
- ‘Knowledge by Acquaintance and Knowledge by Description’ - Bertrand Russell, ‘Proceedings of the Aristotelian Society, New Series, Vol. 11 (1910 – 1911)’, pp. 108-128, Wiley, JSTOR
  
- ‘The Meaning of Ouch and Oops’, David Kaplan, transcribed by Elizabeth Coppock, Howison Lecture in Philosophy delivered at UC Berkeley, <http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf>, 2004
  
- ‘Family Resemblances: Studies in the Internal Structure of Categories’ – Eleanor Rosch and Carolyn B. Mervis, from ‘Cognitive Psychology’, Volume 7, Issue 4, October 1975
  
- ‘Principles of Categorization’ – Eleanor Rosch, from ‘Concepts: Core readings’, edited by Eric Margolis and Stephen Laurence, Bradford Books, 1999
  
- ‘Wittgenstein and Categorization Research in Cognitive Psychology’ – Eleanor Rosch, from ‘Meaning and the Growth of Understanding: Wittgenstein's Significance for Developmental Psychology’, edited by Michael Chapman and Roger A. Dixon, Springer-Verlag Berlin Heidelberg, 1987

- 'The Problems of Philosophy: Vagueness' – Timothy Williamson, Taylor & Francis e-Library, 2001
  
- 'What Metaphors Mean' – Donald Davidson, from 'Critical Inquiry', Volume 5, Number 1, Autumn, 1978
  
- 'The Go-Between' – L. P. Hartley, Penguin Modern Classics, 2004
  
- 'The Problem of Dead Metaphors' – Marga Reimer, from 'Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition', Volume 82, Number 1, April 1996.
  
- 'Does Ontology Rest on a Mistake?' - Stephen Yablo, 'Aristotelian Society Supplementary Volume', Volume 72, Issue 1, June 1998
  
- 'Metaphor and Prop Oriented Make-Believe' – Kendall L. Walton, 'European Journal of Philosophy', Volume 1, Issue 1, April 1993
  
- 'Wittgenstein: On Rules and Private Language' – Saul A. Kripke, Blackwell Publishers, Oxford, 1993
  
- 'Rule-Following and Meaning' – Alexander Miller and Crispin Wright (editors), Acumen Publishing, Chesham, 2002
  - 'Introduction' – Alexander Miller
  
  - 'The Reality of Rule-Following' – Philip Pettit

- 'Ontological Relativity' – W. V. Quine, 'The Journal of Philosophy', Volume 65, Number 7, April 4th 1968

- 'The Meaning of 'Meaning'' – Hilary Putnam, 'Minnesota Studies in Philosophy of Science', Volume 7, 1975