The bottom of things: essences for explanation

Citation: Gibson, Peter (2014) The bottom of things: essences for explanation. [Thesis] (Unpublished)

Birkbeck, University of London

THE BOTTOM OF THINGS
Essences for Explanation

Thesis for the degree of PhD

Department of Philosophy

Peter Gibson

Submitted May 2014

For Sylvia, Eileen and Kate, who have each, in their own way, given wonderful support to this endeavour

Declaration

I confirm that this thesis is entirely my own work, and that all quotations from other sources have been acknowledged in the text and in the bibliography.

ABSTRACT

Central to the philosophy of Aristotle is the belief that the aim of serious enquiry is knowledge of the constitutive essences of a given field. Modern scientific essentialism claims that this still holds good, and this thesis aims to support that approach by elucidating and applying the original concept of essence. Chapter one argues that Aristotle formulated his theory of essences entirely in the context of the theory of explanation expounded in *Posterior Analytics*. The components of that theory are explained, and the implications of Aristotle's view for current debate are considered. Chapter two examines the reasons for the decline of Aristotelian essentialism during the scientific revolution, the metaphysical problems which resulted, and Leibniz's reasons for defending the older view. Chapter three considers the nature of explanation in a modern context, starting with the preconditions for any grasp of reality that are needed to make explanations possible; it is then argued that only essentialist explanation can occupy the role which these preconditions entail. Chapter four surveys the components of that picture of reality that seem explicable, to see how essentialist explanations would actually be formulated. The theoretical discussion concludes with an account of what form essences should take, in order to occupy the explanatory role that has been assigned to them. The final chapter examines the cases of counting physical objects, explaining abstract axiomatic systems, and the discovery of the periodic table of elements, showing how attempts at explanation in these cases all converge on the sorts of essence which have been delineated in the thesis.

Contents

ONE Aristotle on Essence

1. Starting with Aristotle 6
2. Translating Aristotle 6
3. Aristotle's metaphysical project 8
4. Explanation 10
5. Demonstration 13
6. Definition 14
7. Unity 17
8. Essence for explanation 20
9. Aristotle and modern science 24

TWO Crisis for Essentialism

10. The scholastics 27
11. Rejection of Aristotle 30
12. New science 32
13. New scepticism 34
14. Locke's ambivalence 37
15. Leibniz's response 42

THREE Explanation for Essentialists

16. Four examples 51
17. What needs explaining? 52
18. Stable structures 53
19. Foundations 53
20. Levels 54
21. Direction 55
22. Ontology 55
23. Induction 56
24. Typicality 56
25. Coherence 57
26. Induction problems 57
27. Necessities 58
28. Preconditions summary 58
29. Prediction and accommodation 59
30. Covering laws 60
31. Reductive explanation 62
32. Causal explanation 64
33. Models and mechanisms 65
34. Levels of mechanism 67
35. Abstracta 67
36. Convergence and fruitfulness 68

FOUR Essence for Explanations

37. Psychology of essences 69
38. Necessary or essential? 71
39. Metaphysical explanation 73
40. Explaining individuation 74

41. Explaining tracking 78
42. Explaining conceptualisation 79
43. Explaining classification 80
44. Explaining definition 81
45. Explaining extrapolations 82
46. Explaining natural kinds 83
47. Explaining modal reference 86
48. Role of essence 87
49. Essence and concepts 89
50. Essence and classification 90
51. Essence for induction 91
52. Essence and natural kinds 92
53. Essence and possibilities 93
54. Essence and unity 93
55. Essence and change 94
56. Essence and definition 95
57. Explanation aims at essence 95
58. Value of essentialism 99

FIVE Cases of Explanatory Essentialism

59. Testing the theory 101

Case 1: *Unity and Counting*

60. Symbols for counting 101
61. Concepts for counting 102
62. Perception and counting 105
63. Unity and counting 108

Case 2: *Axiomatised Systems*

64. Explaining systems 111
65. The kernel of a system 111
66. Axioms and explanation 114
67. The essence of classical logic 115
68. The essence of set theory 117
69. The essence of arithmetic 118
70. Peano Axioms as the essence of arithmetic 119
71. Logic as the essence of arithmetic 121
72. Set theory as the essence of arithmetic 123

Case 3: *The Periodic Table*

73. The nature of gold 126
74. Scientific essentialism 127
75. Mendeleev 127
76. Explaining the periodic table 128
77. Essence of an element 130
78. Weak and strong conclusions 131

Bibliography 134

ONE

Aristotle on Essence

1. Starting with Aristotle

The concept of an essence seems to be an inescapable feature of all human thought, and arises without prompting in small children. Extensive research by the psychologist Susan Gelman conclusively demonstrates a tendency among even the youngest infants to search for the hidden essence of a thing, which will explain the patterns of the thing's behaviour (Gelman 2003, to be revisited later). A quest for the 'essence' of things runs throughout early Greek philosophy, and Plato's theory of Forms can be seen as making a bold claim about essences, that they must be unchanging, and so cannot be a feature of the visible world, where change is endemic. Perhaps the most significant rebellion of Aristotle against his teacher was on this issue, because it seemed that change can only be understood if there is something unchanging present within reality to support the process, and so he proposed that essences were present in the experienced world precisely to fulfil that role. Thus Aristotle launched the essentialist view of nature which interests us.

Most philosophers of the early modern period could read Greek, had direct acquaintance with the full corpus of Aristotelian texts, and understood him very well. In recent times, however, the picture has been less encouraging. We moderns tend not to read Greek, and are so daunted by the new complexities of analytic philosophy that we are tempted to imbibe our Aristotle through second- or third-hand summaries. In consequence certain misunderstandings have become ossified, and the actual reasoning that led to the original essentialist doctrines has dropped out of view. Serious students of modern essentialism must endeavour to achieve three objectives. The first is to arrive at an understanding of how we should translate a small family of central terms in the ancient Greek. The second is to search out from the Aristotelian texts a wider range of quotations than is customary (two or three friendly sentences often being thought sufficient). The third is to pay as much attention as possible to the Aristotelian scholarship which has flourished in recent times, since a huge effort has been made by a community of experts, resulting in a substantial shift in our understanding of the works, particularly of the way in which texts should be read in the light of the whole corpus, rather than in isolation. The first objective of the present discussion is to assay these three tasks, in an attempt to reach a more accurate picture of what Aristotle was proposing, and the reasons why he was proposing it.

2. Translating Aristotle

Aristotle's discussion of what we call 'essentialism' centres on a small group of Greek terms, most of them taken from normal speech but given a slightly specialised nuance. The translations of Aristotle most readily available in English tend to preserve traditional readings which have drifted significantly from the original, and a return to the original words is a necessity for even relatively unscholarly modern philosophers. The key fact to be recognised is that conventional translations of Greek terms come to us indirectly, via their use in Latin versions of the texts (dating from a time when medieval scholars spoke fluent Latin, but could not read Greek). In our ensuing discussion certain original Greek words will often be supplied in parenthesis, and we must first attempt to clarify what the author seems to have meant by them. Urmson offers a helpful guide to key translations (1990).

Ousia is a cognate of the verb ‘to be’, and literally means ‘being’ or ‘reality of’. It is a broadbrush term used by Aristotle when he is trying to grapple with general problems of existence. The best usual translation in English is ‘being’. Aristotle also has a concept of *prote ousia*, which means ‘primary being’, which is a rather more specific notion of whatever is central to a particular thing’s existence, for which the English word ‘substance’ is an understandable equivalent. When he uses the word *ousia* it is often clear from the context that he means *prote ousia*, so mere ‘being’ may not be sufficient. The important fact about the word *ousia* which needs to be understood is that in Latin the word became ‘essentia’, based on the Latin verb ‘to be’. This, of course, has given us the English word ‘essence’, producing the impression that *ousia* can be translated as ‘essence’, which is usually misleading or wrong. The best account of the situation is that *ousia* is the problem Aristotle set himself, and ‘essence’ is the beginnings of a solution, so the two concepts must be kept separate if we are to read him correctly. At *Met* 1017b13-17 he expounds a considerable range of meanings for *ousia*.

Tode ti has the literal meaning of ‘a this-such’, so that it indicates a particular thing, which can usually be picked out by saying what sort of thing it is, as when we might say ‘that cat’. The phrase involves an ambiguity which is at the heart of modern exegesis, depending on whether the emphasis is on the ‘this’ (the distinct entity) or the ‘such’ (the kind). At *Gen and Corr* 317b21 Aristotle himself wonders whether his enquiry should focus on the ‘this’ or the ‘such’. Witt suggests ‘individual’ as the best translation for *tode ti* (1989:164), but ‘one of those things’ might capture it well. It certainly indicates a distinct entity, and if *ousia* is Aristotle’s initial problem, then analysing the nature of a *tode ti* is his first step towards a solution.

To hupokeimenon literally means ‘that which lies under’. In his analysis of distinct entities, his next step is to postulate some ‘underlying’ aspect of the thing, which unifies the thing, supports change and predication, and might remain after the thing’s superficial attributes are discounted. The closest translation might be ‘the ultimate subject’, but the philosophical term ‘substrate’ seems to capture it well. In Latin ‘that which stands under’ became ‘substantia’, and hence the English word ‘substance’. Unfortunately in modern discussion the word ‘substance’ is used ambiguously, to mean either the whole of some entity which is distinct, unified and separate, or to mean the substrate or even essence of such a thing. The first usage, for the whole thing, is quite useful in modern discussions, but the ambiguity means that the word ‘substance’ is probably better avoided, and ‘substrate’ is preferable.

To ti en einai literally means ‘what it was to be that thing’. In Aristotle’s analysis of the nature of being, he focuses on individual things, postulates something lying beneath the superficial attributes, and terminates his enquiry at ‘what it really is’. *To ti en einai* is the only phrase in the Aristotelian corpus which can legitimately be translated as ‘essence’. In English a reference to the ‘nature’ of a thing comes close to his concept, but if pushed the English will distinguish between a thing’s superficial nature and its essential nature, so ‘essence’ or ‘essential nature’ capture most clearly what he meant.

Kath’ auto is a phrase often used by Aristotle to qualify *to ti en einai* (the ‘essential nature’), and means ‘in itself’. In this case the Latin translation of ‘per se’ captures the concept well, and survives unchanged in English. Frequent use by modern metaphysicians of the word ‘intrinsic’ aims to capture exactly the concept Aristotle had in mind.

Aitia is another term, from less metaphysical areas of the texts, which needs to be understood. It is standardly taken to mean ‘cause’, giving us the philosophers’ term ‘aetiological’ for theories based on causation, but it also means ‘explanation’. Only context, and a principle of charity in translation, can tell us which reading to prefer, and it may often be the case that ‘causal explanation’ best captures what is intended.

One example must suffice to illustrate the sort of translation problem that arises in modern discussions of Aristotle. Brody seeks to defend his view that the essential attributes of a thing are those which are necessary to it over time, by quoting *Post An* 74b6, which he gives as ‘attributes attaching *essentially* to their subjects attach necessarily to them’ (1980:146). However, a check with the original Greek of this passage reveals that the word ‘essentially’ in his quotation is offered as a translation of *kath’ auto*, and Aristotle is apparently telling us that the intrinsic features of a thing are necessary to it. The quotation should not be used to equate a thing’s necessary features with its essence.

3. Aristotle’s metaphysical project

Aristotle take an understanding of ‘being’ to be the highest quest in philosophy, but that this can only be achieved through a study of particular instances of being. The problem of Being seems intractable unless a distinctive line of approach can be developed. Hence Aristotle seeks items of being which can be analysed, so that ‘if the thing has being, it has it in a certain way and, if it does not have being in a certain way, it does not have being at all’ (*Met* 1051b34). This seems to indicate that only something which is ‘a this’ [*tode ti*] is appropriate for ontological study, confirmed by the remark that ‘that which is means a thing with thisness, a quantity or a quality’ (*Met* 1030b12). This raises the question of whether an entity has sufficient unity to qualify as a ‘this’, and how to understand being that lacks unity. The varied role of essence in the Aristotelian approach to this metaphysical question will need to be examined if we are to get the whole picture we are after.

For scholars the initial questions concerning Aristotle’s metaphysics are the chronology of the key works, and the continuity (or otherwise) of his views. Gill says that we must discount chronological considerations, because we lack solid evidence (1989:9-10). This is not quite true, since at *Met* 1042b8 we learn that *Metaphysics Z* postdates *Physics*, and at 1037b9 we learn that it postdates *Posterior Analytics*, which gives some grounds for treating *Metaphysics* as a group of ‘late’ works. The key issue, though, is the relation between *Categories* and

Metaphysics, and here it seems highly plausible to follow Frede (1987:25) and others in treating the former as a considerably earlier work.

The traditional picture of Aristotelian exegesis is that he is taken to have 'changed his mind', but modern scholars are seeing close continuity. The traditional view says that *Categories* committed to particular objects as *prote ousia* – the subjects of predication – with the answer to 'what it is' given separately in generic terms, referred to as 'secondary' *ousia*, a reading of *Categories* which retains general agreement among experts. The traditional account says of *Metaphysics* that the new hylomorphism rejects this account, since in place of the primitive particulars we have matter-plus-form, with the generic element being subsumed within the form, which is the essence, and thus giving us an essentialism which is generic in character. Many modern scholars, however, incline to a more continuous reading in which the particular entity remains central, with its nature now explicated hylomorphically in the later work, so that the generic character of a thing arises from the intrinsic features of its individual form, which is the essence. The only shift of emphasis is that predication now pertains to the form of the particular, rather than to the whole particular, so that a causal explanation of the problematic aspects of a thing can begin to emerge. We either have the genus and kind of a particular thing existing as some external truth about it, or as arising from its intrinsic nature. We will take the latter view to be much more satisfactory as a metaphysical picture, and so we will follow Witt (1989) and Wedin (2000) and Politis (2004) in placing particulars unwaveringly at the centre of Aristotelian metaphysics. The most plausible reading is that in the earlier work the individual thing is almost a primitive, with essence given by category, but in the later work he realises that we must analysis the structure of the individual thing, and seek the essence below the surface. The keen desire of Aristotle to expound the unity of an object, explored by Gill (1989), also gives centrality to the particular, even though she takes a more generic perspective. This is not to deny that plausible cases can be made for both readings (Politis: 251), but our preference here will pay dividends later on. Aristotle certainly thought in terms of particular essences when he remarked that 'the essence for a single thing is the essence for a particular' (*Met* 1054a16). Modern Aristotelian essentialists are divided over the relative priority of the kind and the particular, but the particular looks the better bet, both for a coherent picture of reality, and as an account of what Aristotle probably intended.

A basic understanding of Aristotle's metaphysics nowadays involves reading *Categories* and *Metaphysics*, with *Topics* as an optional supplement. Students of his epistemology, on the other hand, will focus on *Posterior Analytics*. It is only in recent times that the close links between this latter text and his metaphysical project have been explicated fully, and the key role of *Posterior Analytics* will be emphasised in the present attempt to understand Aristotle's metaphysics.

We begin with the first sentence of Metaphysics Α: 'By nature, all men long to understand' (a translation recommended by Annas (2000:15)). This sounds vague, but no philosopher has done more than Aristotle to clarify the requirements of understanding. The second step is to put aside *Metaphysics* and switch to *Posterior Analytics*, which is a careful analysis of the procedures which lead to understanding. The recommended procedure for understanding some entity, event or phenomenon may be summarised as follows: first seek a range of appropriate explanations, then gradually refine definitions of the basic items involved in the explanations, then produce logical demonstrations of the necessary features and relations which result from what has been defined; when we understand each thing's nature, and how reality is necessitated by these natures, we have reached the greatest understanding of reality that is possible for human beings. This sounds easy, but Aristotle gives plenty of detail, which we must now examine. The aim is to track the steps in his account, to see how the concept of 'essence' emerges from a particular context, and thus attain the best possible understanding of an 'Aristotelian essence'.

4. Explanation

The aim of enquiry is to achieve understanding, and we are told that 'the study of the reason why' is the best route to understanding (*PA* 79a24), and that 'we understand something simpliciter when we think we know of the explanation because of which the object holds that it is its explanation, and also that it is not possible for it to be otherwise' (*PA* 71b10). The first step in this enquiry, of outlining a general framework for explanation, immediately throws light on where this will lead.

One of the most familiar ideas in the philosophy of Aristotle is that there are said to be Four Causes, but this proposal is not well understood (Moravscik 1974). The Greek word here is *aitia*, which can translate as 'cause', but mainly refers to explanation; we really have the Four Modes of Explanation - material, efficient, formal and final - all of them causal in character. A second common misunderstanding is a failure to register the significance of the 'formal' cause or explanation. The type of causation discussed in our contemporary literature appears to be Aristotle's 'efficient' cause, but for him that is what we might call the 'trigger' of the event, with the formal cause offering a much more powerful explanation. The 'form' here is the concept involved in what we call his 'hylomorphism', which is his most developed essentialist theory. In brief, the main explanation of a thing is its essence, or the essences involved. Hence right from the start of his enquiry he is aiming to attain understanding by means of an explanation which rests on the essential form. A further misunderstanding concerns the 'final' explanation, which tends to be seen as fitting things into a wide teleological world view, when it can equally mean the simple function of some thing (*Eud Eth* 1219a8: 'each thing's function [*ergon*] is its end [*telos*]'). Offering 'final' explanations is not a passé mode of ancient mystical thought, but the normal explanation we give by saying what purpose is served by something like an ear or a watch.

The other interesting aspect of his starting point is that he tells us that 'the things we seek are equal in number to those we understand: the fact, the reason why, if something is, and what something is' (*PA* 89b24). These four don't quite map onto the Four Modes of Explanation, but it is important to see that the starting point is not how the world is, but what we are capable of understanding. It is only in the context of the way in which the human mind grasps the world that the theory of essentialism is propounded, and we have already mentioned strong evidence that essentialism is deep-rooted in our psyche (Gelman 2003). One other commonly overlooked aspect of his Four Modes (given at *Ph* 194b23-195a26) is that there is a second statement of them (at *PA* 94a21) which omits the 'material' cause (e.g. the stuff of which a statue is made), replacing it with 'an antecedent which necessitates a consequent'. Scholars are unsure how to interpret this, but the appearance of 'necessitates' adds a new dimension to our concept of an explanation. The necessitating power of the fundamental level revealed in an explanation will be a recurrent theme here.

For understanding, then, we seem to need an explanation which will give the material involved, the initiator of the explanandum, the form or forms that are its source, a notion of what it is for, and a sense of how it is necessitated. The next step towards the required explanation, having perceived the situation in these ways, is to formulate a *logos* (a rational 'account', or the 'principles'). This is a vague term, which must be filled out with a procedure for attaining *logos*, and this will involve analysis of the situation, because 'we think we know a thing only when we have grasped its first causes and principles, and have traced it back to its elements' (*Ph* 184a12). The formulation of the *logos* will eventually lead us to a definition of the essence.

Explanations, we will find, are predominantly, but not exclusively, causal. The importance of the formal cause here is seen in two nice illustrations, when he tells us that 'what it is and why it is are the same', as when the account of an eclipse of the moon also shows why it happens, and an account of the ratio between musical notes tells you why they harmonise (*PA* 90a15). Hence knowing 'what it is' becomes the focus of explanation, and hence of our entire understanding of nature. That 'what it is' is the same as 'essence' [*to ti en einai*] is seen in the claim that 'we have knowledge of each thing when we grasp the essence of that thing' (*Met* 1031b8). That other types of explanation will not suffice is confirmed in this: 'it is when we know what a man is or what fire is that we reckon that we know a particular item in the fullest sense, rather than when we merely know its quality, quantity or location' (*Met* 1028a36).

How we should understand 'what it is' will need to be examined, and the first question this raises is whether we are trying to grasp particular things or generalities (as seen in the dual meaning of *tode ti*). There is no simple answer here, and the issue haunts many of Aristotle's own metaphysical puzzles [*aporiai*] explored in *Metaphysics* β. His best account of the dilemma is this: 'If the principles are universal, they will not be primary beings, …but if the principles are not universal but of the nature of particulars, they will not be scientifically knowable; for scientific knowledge of anything is universal' (*Met* 1003a8). If the world were made up of dots, science would be joining the dots. If we can only start from the faculties of human understanding (as noted above), then it seems that we can only work towards a grasp of primary being (our main target) via generalities, because 'reason grasps generalities, while the senses grasp particulars' (*Ph* 189a6*;* also *PA* 86a30). Elsewhere he puts the problem this way: 'what is most universal is furthest away, and the particulars are nearest' (*PA* 72a5). At *PA* 87b29 he tells us that, unfortunately, perception explains nothing, and so only universals can give us the explanations and demonstrations that we are after.

This dilemma over whether the general or the particular has primacy will not go away, but we will adopt a simple solution, which we take to be both the best, and the solution favoured by

Aristotle. We need some grasp of how Aristotle understands universals, and we find that ‘it is from many particulars that the universal becomes plain; universals are valuable because they make the explanation plain’ (*PA* 88a5). ‘Induction is the progress from particulars to universals’, he tells us, and illustrates it by deriving the general concept of a ‘skilled man’ from observing successful charioteers and ships’ pilots (*Top* 105a15; also *PA* 100b4). The picture in the whole corpus is complex, but this remark from *Topics* gives us a simple view. Particulars must take metaphysical priority over general concepts, because generalities depend on particulars. The rival reading says that in *Categories* the particular is basic but our understanding comes from the generic ‘secondary substances’, and in *Metaphysics* forms are prior to particular entities, with forms entirely couched in terms of generalities. Our preferred reading says that particulars are foundational in both texts, but simple in the earlier work, and complex in the later work. In *Categories* it is not clear that there is any genuine essentialism to be seen, since the whole particular is basic and simple, and the ‘secondary substances’ are not true essences, because they are predicated of something, which is contrary to the foundational role of an essence (Frede 1987:26). Aristotelian essentialism is hylomorphism, or it is nothing. The genus, the category, and the ‘sortal’ concept we will meet later, are shorthand summaries of inductive generalisations, crammed with information drawn from many sources. As such, they are tools to achieve a rapid comprehension of what is shared between individual essences, but to speak as if they actually constituted the nature of essence we will regard as a misunderstanding. You can’t disagree with the observation that ‘he who describes ‘man’ as an ‘animal’ indicates his essence better than he who describes him as ‘pedestrian’’ (*Top* 128a24), but this is because inductive generalisation has accumulated far more information about humans than about pedestrians. The target of the whole metaphysics is to understand primary substance, and ‘the substance [*ousia*] of each thing is something that is peculiar to each thing, not pertaining to anything else, whereas the universal is something common’ (*Met* 1038b10). What we really want to know is not the universals involved, but ‘why the primitive term falls under the universal’ (*PA* 99b12).

Witt makes the interesting suggestion that Aristotle’s focus on the particular is driven by an epistemological requirement (just as the interest in essence is also driven). The quest is for an understanding of the actual world, but ‘knowledge is ... a double thing, being both potential and actual; now potentiality is like matter - it is universal and indefinite and it is the potentiality of something that is universal and indefinite. But actuality is definite and of something definite, being a this-such [*tode ti*] of a this-such’ (*Met* 1087a14). In his hylomorphic view, the actuality of the form is imposed on the potentiality of the matter. Thus our knowledge of actuality requires knowledge of the form, but actuality is only found in a *tode ti*, which is an individual. Witt comments that it is precisely the generality of all universals which renders them indeterminate, and hence excludes them from our basic knowledge of actuality (1989:169). Frede somewhat disagrees with her view, suggesting that individual forms ‘in part are constituted by unrealized possibilities’ (1987:90). This modal dimension to the concept of a form or essence will need to be addressed further.

So our quest for understanding starts from human cognitive faculties, seeks four modes of causal explanation, tries to articulate a *logos* for these (focusing particularly on the 'formal' explanation), and moves towards a definition (necessarily expressed in universal terms), which will get as close as we can manage to the essence of each individual thing. Given the proposal that each thing is 'formed matter', his main target becomes 'that by virtue of which the matter is in the state that it is in' (*Met* 1041b8), so an understanding of the causal powers of form is what we seek, and what the explanation must illuminate. The next step is to look at 'demonstration' [*apodeixis*], which is understood to be logical deduction which 'gives the explanation and the reason why' (*PA* 85b24).

5. Demonstration

A demonstration is a form of syllogistic argument, which takes necessary facts about the world as input, and shows the necessities that follow from them. Things are explained because we see that they have to be that way, and nothing more is needed for understanding. The starting point of a demonstration is what exists, the background assumptions, and the various attributes (*PA* 76b12), and the aim is to elucidate the unfamiliar in terms of what is familiar (*PA* 71b22), because if we rest on the simple and familiar 'knowledge will come more quickly, and that is preferable' (*PA* 86a35). Speed of understanding is not a criterion for modern epistemologists, and it again reveals the way in which Aristotle starts his enquiries from what is human, rather than from what is taken to be real. It is speed of understanding which makes the kind to which a thing belongs so important in his approach, as when we identify the species of some approaching animal. As Russell noted, the word 'dog' is 'a condensation of many inductions' (1940:76).

It would be a neat picture if the process of demonstration converged on the desired definitions, but in fact we are told that the definitions must come first, because 'all demonstrations clearly suppose and assume what a thing is' (*PA* 90b30). We might say that definitions reveal the dots, and demonstrations reveal how they are joined, so that the two procedures work in tandem. Since his essentialism asserts that the joins (the necessities) arise from the dots (the forms of particular things), the definitions will give deeper understanding than the demonstrations.

Because syllogistic reasoning is involved, we can see why universals are required in the explanatory process. In the classic syllogistic inference of Socrates's mortality from his humanity, we start from a particular man, but it is only the overlap of generalities (the inclusion of the humans among the mortals) that makes the inference of his mortality possible. Particular demonstrations are also possible, but these are said to 'terminate in perception' rather than in thought (*PA* 86a30). If we consider the inference by a detective of the identity of an individual criminal, this claim doesn't seem quite right, but it shows how Aristotle is trying to unite the role of the universal and the particular in one coherent picture. He remains committed to the target of his project, which is that 'we must advance from generalities to particulars' (*Ph* 184a24), even when broad understanding resides in the generalities. He illustrates his procedure when he notes that if you demonstrate a truth about some particular triangle, you will find that you have demonstrated a truth about all triangles, and demonstrating that one man is an animal reveals that all men are animals (*Met* 1086b36). It is, of course, the necessity detected in the particular which reveals the general truth.

Necessity is ‘what makes it impossible for something to be other than it is’ (*Met* 1015b3). The approach adopted in the present discussion will take seriously the word ‘makes’ here; that is, we should be cautious about any claim of necessity that doesn’t offer some sort of ‘necessity-maker’. In his account of demonstration, Aristotle tries to flesh out this picture, by outlining a process of necessitation. Demonstration is a type of syllogistic deduction, so we need to know where the necessity enters the process. In *Prior Analytics* a deduction is defined as ‘a discourse in which, certain things having been supposed, something different from the things supposed results of necessity because these things are so’ (24b18). Hence, from Aristotle’s perspective, the formal process of deduction proceeds by necessity, but the conclusion could still be contingent, if the premisses were contingent, so what will ensure the required necessary conclusion? Will one necessary input suffice, or must all inputs be necessary? After careful discussion, he concludes that the demonstration we are after must ‘proceed from necessities’ (*PA* 73a24), but also that ‘your demonstration must proceed through a middle term that is necessary’ (*PA* 75a13). In the classic example, it needs to be necessary that Socrates is a man, and also necessary that men are mortal. Of the start of the process he says that the initial necessity comes from ‘whatever holds of an object in itself’ (74b5), which will turn out to be the essence of the thing, specified by its definition. For a fuller picture we should note his distinction that ‘for some things, the cause of their necessity is something other than themselves, whereas for others ….there is no such external cause, but rather they are themselves the necessary cause’ (Met 1015b14), which requires distinctions to be made among our ‘necessity-makers’, but we will focus for now on the way things themselves generate necessities. We now approach the main target of our present enquiry, which is the role of definition in establishing the essences, which are the sources of the necessities which explain.

6. Definition

The aim of an Aristotelian definition [*hurismos*] is best described as a verbal isomorphism with the complex essence of the definiendum: ‘a definition is a formula [*logos*], and every formula has parts; further, as the formula stands to the object, so do the parts of the formula stand to the parts of the object’ (*Met* 1034b20). We must be careful that we do not see definition in terms of modern lexicography, which suggests a short phrase aimed at facilitating usage of the word. An Aristotelian definition is potentially a much more substantial affair, which ‘contains many assertions’ (*Top* 155a3), and should be understood as something like a modern science monograph, particularly if it describes the complex procedure by which the definition is reached. The ‘parts’ of an object referred to may be very fine-grained, but more than the mere physical parts should be involved, since the definition of an essence should show that it unifies the parts.

With that picture of definition before us, we must ask what it is that we are attempting to define. Here we come up against the human limitations of what can be achieved, and we meet again the tension between what is particular and what is general. We might think that almost anything can be defined, but Aristotle’s interests are narrower, because ‘a definition must be of something that is primary' (*Met* 1030a8), and 'only substance admits of definition' (*Met* 1030b34). This seems to be aimed at the individual, but in more than one place he tells us that particulars cannot be defined. This is 'because an account is general' (*Met* 1040b2), but also because 'particular perceptible substances' are said to be 'in a variety of states', which makes them elude definition (*Met* 1039b30). There is, though, no question that he is committed to individual distinctiveness, and that much can be said about it, in this remark: 'even things in the same species have different causes, differing not, evidently, by species but in as much as particular things have different causes. For instance, your matter, form and motive cause are all different from mine' (*Met* 1071a27). We must stand by our earlier reading, and say that the 'variety of states' must refer to the 'accidental' attributes of the thing, and not to its essence (a defining role of which is, as we will see, to remain stable amidst change). The generality of language accounts for his pessimism about arriving at a complete definition of a particular individual. Just as overlapping universals were needed for the syllogisms of demonstration, so universal terms are required in any attempted expression of a *logos*. A definition is certainly not just a generic statement, since he tells us that 'the definition ought to be peculiar to one thing, and not common to many' (*Top* 149a24), and he also makes the further (and stronger) claims that 'there cannot possibly be one definition of two things, or two definitions of one thing' (*Top* 154a11), and that 'everything that is has one single essence [*to ti en einai*]' (*Top* 141a36).

The target, then, of a definition is to get as close as possible to pinpointing each individual, and also to provide the information about its essence which determines the exact kind to which it belongs. We see how this is to be achieved when we examine the method of definition recommended by Aristotle. The normal label given to his approach to definition is the method of 'genus and differentiae', which implies (roughly) that definition is a particular approach to classification, but it is better if we talk of his 'method of division'.

Definition is not a precise activity, but we can say that the Method of Division begins with three steps: first you 'take what is predicated' of something, then you 'order' these items, then you ensure you haven't missed anything (*PA* 97a23). He also suggests starting with a comparison of items similar to the target (what Lipton calls explanatory 'foils'), picking out first the similarities, and then the differences (*PA* 97b7). In modern parlance we might start with the advice to 'make a complete list of the properties' (though we note that Aristotle talks of 'predicates', not of our 'properties'). The next step is that 'the first term to be assigned ought to be the genus' (*Top* 132a12). The 'genus', we have already been told, is 'that which is predicated in the category of essence of several things which differ in kind' (*Top* 102a32). This is not yet a definition, because a definition must be unique to each thing, but we are now in a position to begin the 'division'. In one version he tells us to 'divide a whole into its primitives, then try to get definitions of these. Thus you establish the kind, and then study the attributes through the primitive common items' (*PA* 96b16); in another version he says 'the contents of definition by division are the so-called primary genus (such as 'animal') and the differentiae. ...It should always be prosecuted until the level of non-differentiation is reached, ...and the last differentia will be the substance' (*Met* 1037b30).

He gives further details, and some examples from geometry and biology. The interest in the first version is that we seem to aim at a definition of the 'genus', but we then divide and then define again, so that definition is an evolving and iterative process, and we should pay attention to the terminus of the process, because that is where essence is to be found. The second version, while suggesting that 'substance' is a 'differentia' (a view not supported well in the remaining texts, and contradicted at *Top* 122b17), gives what should be seen as the key to essentialist definition, that division should be pursued right to its limit, and that only there will we hope to find what we seek – an account of 'what it is'. Thus he refers to the question of what the 'last item in the series' is in analysis, and recognises that we might terminate either in matter, or in the substance [*ousia*] we seek (*Met* 1048a32). No guidelines are offered for interpreting this final stage. It may be that we do no more than catalogue the differentiae, since he tells us that 'we usually isolate the appropriate description of the essence of a particular thing by means of the differentiae which are peculiar to it' (*Top* 108b5). Defenders of the generic view of essences will regard this lowest level of the definition as giving the narrowest species of the thing (the 'infima species'), but we take it as established that a grasp of the particular 'in itself' is what should emerge here, not just a precise classification of that particular. It may be that the only perfect grasp of the particular that Aristotle offers is through the inarticulate mode of perception, but the aim of our talk is to bring our reason as close as we can to such a state.

One puzzle about definition with which Aristotle was concerned, but which doesn't bother modern thinkers much, is why a definition is understood to be unified (*Met* 1037b10; *PA* 92a30). If a definition is seen as something like a proposition, then this resembles the problem of the unity of the proposition that gave trouble in the early twentieth century. If a definition culminates in a list of features, then what unifies a mere list is an obvious difficulty for anyone. The complexity of the problem for Aristotle is shown in the interesting remark that 'things are numerically one in matter, formally one in their account, generically one in their pattern of predication, and one by analogy if related to a further one' (Met 1016b30). Hence the unity of the 'account' [*logos*] is only one aspect of the problem of unity. If you view definition as primarily verbal and conceptual, then you will struggle with the problem of the unity of definition (and it may be that unity can only be stipulated); if you focus on what is being defined, and accept that a definition only succeeds when the definiendum is a unified entity, then a successful definition will inherit that unity. Aristotle devotes a great deal of effort to the unity of particular entities, and we will address that question below. The problem of the unity of definition recedes in the texts as a solution to the unity of particulars emerges.

The problem of unity is not simply the problem of the unity of definition. Aristotle writes that 'being one in form is just another way of saying one "in definition"' (*Ph* 190a16), but that does not mean that an essence just *is* a definition. Lowe, for example, defending Kit Fine's focus on definition to renew the Aristotelian conception of an essence, is bothered that if an essence is an entity then it too will have an essence, leading to a regress (2013:23). Hence he says firmly that an essence is 'no entity at all' and tells us that 'all that grasping an essence amounts to is understanding a real definition, that is, understanding a special kind of proposition' (2013:28). Since the essence isn't an entity, and is evidently an abstraction covered by the phrase 'what it is', we seem to have nothing occupying the role of essence apart from the definition. This is a possible view, but it is not Aristotle's view, who tells us plainly that 'if a definition is the recognition of some essence, it is clear that such items are not essences' (*PA* 90b17). There is a real difficulty with a regress of essences, which will need to be addressed later.

Demonstration aims to establish necessary truths about the world, using a process (deduction) which preserved necessity, and which must both 'proceed from necessities', and also involve necessities in the intermediate stages (the 'middle terms'). We also noted that definition must precede the process of demonstration, and so the definitions must contain necessities. We need to examine the relationship between what is essential to a thing and what is necessary, but for now one important remark from Aristotle will give us the required link: 'whatever is predicated in what something is is necessary' (*PA* 96b3). If an essence is expressed by a set of predicates which constitute a definition, those predications will be necessary truths, precisely because they state what the thing is. Aristotle aimed to account for change, and if you change the accidental features the thing remains the same, but if you change a thing's essence, it cease to be that thing. For some thinkers that is all there is to an Aristotelian essence, but we will argue below that such an account is not at all what is required for a fully developed neo-Aristotelian metaphysics.

7. Unity

Particular unified entities are the basis upon which Aristotle constructs the general truths of science that emerge from explanations, demonstrations and definitions. We have also seen that he takes the particulars to be the units which give rise to counting (and 'they say that the unit [*monada*] is the starting point of number' (*Top* 108b30)). The remarkable unity of the whole of Aristotle's philosophy seems to rest on the slippery issue of what makes a thing *one* thing. In *Categories* (taken to be an early work) the issue did not arise, because individuals as a whole are treated as primary substances, and the question of what unifies an individual need not be addressed, presumably because this was self-evident in the biological organism which were always his paradigm cases. In that work, everything worth saying about the individual falls within the 'secondary substance', but the reliance on universals in that respect is precisely the reason (according to Frede, 1987:50) why Aristotle abandoned secondary substance, and offered instead a structural account of the object, locating the inescapable universals within a linguistic definition or a deductive argument. Hence it is in *Metaphysics*, where entities have a hylomorphic structure, that the problem arises of whether the unity of a thing derives from the form, or from the matter, or from the composite of the two (and Frede also notes that a complete particular includes accidents, as well as form and matter (1987:74)).

It seemed clear that Aristotle's concept of the unity of a thing does not involve its accidental features when he told us that we are not seeking the 'quality, quantity or location' of a man, or of fire (*Met* 1028a36), but this shows us that we are not pursuing the unity of a complete particular item at some given moment (e.g. of a man when he is sitting), principally because the unity sought will persist through change, and through contradictory predications (such as 'standing'). Hence to grasp his theory of unity (if he has a full 'theory') we must attend to form and to matter.

Gill offers an illuminating discussion of this question by identifying a potential paradox in Aristotle's view, one which can only be dispelled by a careful account of Aristotle's understanding of matter (1989).

The difficulty is a possible conflict between the unity of an entity over time and the unity of a thing at an instant (what she calls 'horizontal' and 'vertical' unity). It appears that unity over time is provided by the matter, which can pre-exist the coming-to-be of the thing, and can survive its passing-away, but this implies that the matter has a nature which is distinct from the form of the thing (since the form comes to be and then passes away). However, when we consider the timeless ('vertical') unity of the thing, Aristotle seems to tell us that the role of the form is precisely to produce a single indivisible entity through its operation on the matter (Gill 1989:145). Hence unity for change appears to undermine unity for predication, and the resulting lack of unity prevents the entity from being a primary substance, since one part of it can be predicated of the other. The solution offered by Aristotle, and expounded by Gill, is summarised in his remark that 'the problem of unity disappears if our account is adopted. We allow a matter component and a shape/form [*morphe*] component, one existing potentially the other in actuality. …The account is of a unity because one component is material, the other shape/form' (Met 1045a24). Roughly, to see a paradox here would be to fall into a category mistake.

The introduction of the distinction between what is potential [*dunamis*] and what is actual [*energeia*] introduces a new dimension into theories about the unity of an entity. Modern translators struggle to find appropriate English equivalents for these terms, but Beere, in a book entirely focused on *Metaphysics Book Θ*, which deals with these two concepts, translates *dunamis* as 'capacity', while leaving *energeia* untranslated, but glossed as 'the exercise of a capacity', or as 'activity', or as 'actuality' (2009:3-5). Gill also offers 'power' as a good translation of *dunamis* (1989:173). The exegetical debate here is complex, but we can see that Aristotle has a concept of the unity of a thing which involves both of what modern discussion calls the 'dispositional' and the 'categorical' properties of a thing, though the Aristotelian language seems more dynamic in character than our words. In simplest terms we might say that a house is made of bricks, and while the bricks have the potential to become a wall, the house can only ever be a house. Aristotle seems to be offering two modes of existence for the form and the matter (as opposed to the modern claim that a statue and its clay are two existing 'objects' with properties that can be compared). Of the form, Aristotle writes that 'the what-it-was-to-be-that-thing [*to ti en einai*] is a unity of a kind straight off, just as it is a being of a kind; and that is why none of these things has some other cause of their being a unity' (*Met* 1045b4). Hence the form and the matter have two modes of existence, but it is the role of the form to incorporate the matter into one unified entity (without destroying the 'capacity' for other activities which characterises the matter).

We have already established that the form of a thing is its essence, and so the essence is what bestows unity on a thing. On the whole the texts seem to support the view that if a thing has an essence, then it must be unified, and if it lacks an essence, it is thereby bereft of unity. Essence is both necessary and sufficient for unity. Furthermore, since it seems clear that the concept of essence has only arisen in the context of the quest for explanations which will deliver understanding, we see that there is some sort of close connection between explanation and the unity of the objects involved in the explanation. If one's view of explanation is a highly pragmatic and contextual one, then this may threaten to make the concept of the 'unity' of an object wholly conventional, and even spurious. That is not Aristotle's view, since he has fought hard to explain the real unity of things, and his concept of an explanation rests on the search for the foundational features of each thing, which give the full and objective revelation of what that thing is. As we will see, subsequent challenges to Aristotle's hylomorphism have not only raised doubts about the concept of essence, but they have also given rise to alternative views of explanation, and also given modern metaphysicians major difficulties when they attempt to ascribe unity to anything.

It is interesting to see what sorts of things turn out to have unity for Aristotle, such as the unmoved mover, planetary bodies, elements, matter in general, artefacts, and living things. The unmoved mover is certainly a unity, but hylomorphism is not invoked in its account because 'there exists an eternal unmoved substance separate from sensible things. It can have no magnitude, and is without parts and indivisible. As the source of movement for infinite time, it must itself be infinite' (*Met* 1073a05; also *Ph* 267b19). Here we have an absolute unity which is primitive, and foundational for his metaphysics. This unity is also extended to the planetary bodies, which initiate their own movement, and hence also are intrinsically 'without magnitude' (*Met* 1073a36). The elements (earth, air, fire and water) are interesting cases, because on the one hand he tells us that 'none of them is a unity; rather they resemble a heap until such time as, by subjecting them to concoction, something that is a unity is produced from them' (*Met* 1040a8), but on the other he is prepared to consider two separate drafts of water from a well as distinct entities that belong to the water species (*Top* 103a20), and he is tempted to refer to the elements as 'objects' (*Ph* 192b9). Earth is a borderline case, because 'although there is a sense in which mud breaks down into lumps of mud, there is also a sense in which it does not' (*Ph* 188a14). It seems that unity can come in degrees, and that drawing a draft of water or breaking mud into lumps imposes some sort of form on the raw matter. Matter in general, as we have seen, exists only in potentiality in its undifferentiated state, and so lacks unity because it lacks any sort of form.

We are told that while an artefact can be considered as a 'whole', the unity is of a lesser degree because 'you make something a unified whole by gluing it, banging nails into it or tying it up'; which means that it does not 'contain in itself the cause of its being continuous' (*Met* 1052a 22-24). Aristotle is much more impressed by an entity which has a natural unity, but that is because it 'contains within itself a source of change and of stability' (*Ph* 192b14). Hence it comes as no surprise that Aristotle regards living entities as by far the best candidates for genuine unity. We may see that as predictable, given his initiation of the whole science of biology, but it is equally plausible to say that his interest in the metaphysics of unity is what drew him to the study of living things.

The unity of non-living terrestrial objects such as pebbles seems not to figure in his thinking. Anything which counts as earth, air, fire or water will usually be excluded as mere potential, and so his actual world contains very little that is even a candidate for true unity, apart from living things. The tradition that ascribes true terrestrial unity only to living things will be found again in Locke and Van Inwagen, and we will return to it. The problem which will then be faced is that the modern exploration of matter has focused on ‘objects’ such as molecules and atoms which are necessarily treated as having a high degree of unity in the theories of modern science, and this raises metaphysical questions which did not bother Aristotle.

8. Essence for explanation

We now see how Aristotle intends to achieve general understanding, and so we can ask how much clarity he aimed for and achieved in the concept of *to-ti-en-einai* [essence, what-it-was-to-be a ‘this’]. Aristotle’s direct account of form gives us little more than the two words ‘*eidos*’ and ‘*morphe*’, which, with the word ‘*hyle*’ (‘matter’), gives us our term ‘hylomorphism’. *Morphe* is the English word ‘shape’, and *eidos* is a vague and broad word covering ‘form’. The word *arché*, a ‘first principle’, also crops up in the discussions (e.g. *Met* 1041b31). The fact that in *De Anima* we learn that the soul [*psuché*] should be understood as the ‘form’ of the body shows that we are a long way from simple modern ways of understanding these words (*De An* 412a20). If this concept is meant to explain reality to us, it is rather thin and elusive, and more fruitful is to ask what role Aristotle takes ‘form’ to play in his mature metaphysics.

In general, the role of an essence is to ground a successful explanation. The five aspects of things that require explanation are the ability to support a range of necessary and contingent predicates, the phenomenon of a thing remaining what it is through the vicissitudes of normal activity, the fact that a diversity of parts can exhibit unity, the range of causal powers which each thing exhibits, and the ability of things to transmit necessities. There is no prospect here of a very specific account of essence, because the concept of essence is a generic one which covers a range of features and structures which play the explanatory role. We can only work our way towards the concept of whatever it is about a particular thing which could offer explanations of all five of these puzzling phenomena. What is it that supports predicates, survives through change, unifies a thing, supports its powers, and necessitates behaviour? Do we need five separate explanations, or will the five explanations converge? It is only in his biological studies that Aristotle himself made a comprehensive effort to grasp essence. In the absence of microscopy and of the experimental method, he made huge progress in that respect, mostly through the method of classification.

If an essence fulfils the role of unifying an entity, what is it that is unified? We can either approach this matter mereologically, by citing its parts, and looking to see what holds them together, or we can approach the matter through attributes, asking what can connect and support a ‘bundle’ of such things. If we ask what maintains the nature of a thing through change, we see the essence as not only ‘holding together’ the parts or attributes, but also maintaining their order and causal powers (as when skin heals after a small wound). Further, if we take the view that individuals fall into categories such as species as a result of inductive generalisations about their evident features and powers, then the essence of each creature will be very similar, as seen in the very similar features and powers which each individual exhibits (despite their unique variations), and this essence will be responsible for maintaining that particular set of features. The word 'maintain' recurs in these characterisations of essence, and if we add the requirement that necessities are transmitted through these activities, as Aristotle emphasises in his account of demonstration, then the 'maintenance' achieved by essence must be characterised in a very strong way.

Of the words that might capture this strong sense of 'maintenance' concerning essence, *arché* seems the most illuminating, because there is some sort of fixed principle that is maintained while the configuration of the relevant matter varies. At this point we might invoke controlling 'laws', or even the 'occasionalist' intervention of a divinity, to underpin such a strong principle, but Aristotle obviously sees the 'principle' as intrinsic and embodied. There is a choice here over whether to understand the essences of physical objects as themselves physical, or as abstractions. If essences are abstract, then 'principles' seems the best word for them, but if they are physical then the word that best seems to capture the purposeful structures which they have to be is 'mechanisms'. If we wish to examine essence in both physical and abstract realms, we may wish to retain both words.

Sceptics will say that mere speculation about something to which words like 'principle', 'mechanism' and 'form' are attached throws no light on anything, since (in Locke's words) it might equally be labelled the 'I-know-not-what'. However, if we respond by saying that Aristotle's proposal should be understood not as a theory, but as a research project, we ought to be more sympathetic. Consider a famous parallel: Molière created a nice joke when a character explains why someone falls asleep after taking opium by citing the 'dormative power' of the opium, as if this were a scientific theory. But if we had four theories of why the person fell asleep (such as illness, exhaustion, hypnotism and opium), it might be very astute to spot that it was the dormative power of the opium. The scientists could then investigate exactly what that power is (and by now they have probably managed it). That, I take it, is the way in which we should understand hylomorphism. A quest for understanding should study the intrinsic foundational principles and/or mechanisms of the object of study.

But has Aristotle hit on the correct research programme for science? The scientific revolution is normally seen as deriving its main impetus from a rejection of Aristotle. For example, experiments under controlled conditions showed that Aristotle was wrong in his claim that life spontaneously generates in decaying meat. The corpuscularian approach seemed to generate new discoveries in a way not possible for hylomorphism, and eventually Newton's triumph suggested that phenomena like gravity emerged from the behaviour of mathematically describable laws, and not from the teleological characterisation of the intrinsic natures of earthy objects as having a 'downward' purpose (*Ph* 254b21). Since then, philosophers and scientists have been much more interested in laws (perhaps as axioms of experience on the lines of the Mill-Ramsey-Lewis account), and explanations have tended to invoke laws rather than 'natures'.

The present discussion will attach itself to the Aristotelian view of science. In brief, there is a growing realisation that the view of science launched by Galileo's claim that the book of nature is written in mathematics, and consolidated by the wonderful equations of mathematical physics, does not fit the science of biology at all. Biologists sometimes pursue the spirit of the Galilean view, with a role found both for mathematics and for tentative 'laws', but the triumphant discoveries of biology are not mathematical relationships, but the revelations of hidden mechanisms. This might have split science into two separate activities, until the possibility emerges that even physics might actually be homing in on mechanisms, with the famous equations understood not as mere abstractions, but as very precise descriptions of the mechanisms (with '$e=mc^2$', for example, pinpointing a universal conversion mechanism). Is the triumph of quantum mechanics its equations, or its precise explanations of how small particles combine to generate our world? Is gravity best understood as an equation, or as the universal behaviour of matter? If we see the equations as 'principles', do they arise from the intrinsic nature of the matter and objects involved? Answering such questions is not on the immediate agenda, but such questions summarise the present approach. Given the Aristotelian view of modern science, can that approach be equally successful in metaphysics? The modern perspective on his work, of understanding the metaphysics through its role in the scientific method of *Posterior Analytics*, offers the prospect of integrating an old scheme of thought into a recent picture of nature.

The key thought of the present discussion of Aristotle is that he did not surmise that (ontologically speaking) there are essences out there, and we are on a mission to understand them. Rather, we start from the mission to understand, and the process of explanation required for that has to be essentialist. In a Kantian spirit, it is a precondition of our understanding the world that our thinking is essentialist. The first of many questions here is 'might that mean that essences are just necessary fictions?' Having already defended Aristotle's robust realism, we must deny that. The thought is simple: essences are thoroughly real features of the world – but they are features which are only picked out in the context of real explanations. Children essentialise about objects when they are confused by them, and are struggling to understand, and adults do the same. If I stare at a tree and ask for its 'essence', no useful answer is forthcoming. If I try to understand its dominance in a forest, its vulnerability to some disease, its utility for humans, or just its beautiful shape, then an enquiry after the explanatory foundations of its nature will pick out what can plausibly be called its 'essence'. The view is 'essentialist' because such questions seem to converge on certain restricted features of the tree. The hallmark of the type of Aristotelian essentialism being defended here is that explanations of the many puzzling aspects of any distinct entity will tend to converge on a single account.

Lowe is sympathetic to essentialism, but rejects this characterisation of essences as explanatory. We have seen his preference for the central role of definition rather than of explanation, and he rejects the explanatory approach because explanation is 'multifaceted', falling into a number of species, of which essentialist explanation is just one type (2013:20). He suggests that we cannot offer a clear picture of explanation, and then infer from it our concept of essence. Hence his approach is to try to formulate a concept of essence, which can then be used as a mode of explanation. We have already seen, though, that he seems to retreat from the reality of essence, offering instead a purely abstract and conceptual account in its place. The present approach returns us to the reality of essence (at least for physical objects), but

Lowe is right in that an extremely diverse, pragmatic and conventionalist view of explanation will open the floodgates to an equally diverse (and useless) view of essence.

Oderberg also defends essentialism, while rejecting its basis in explanation, saying rather that ‘the role of essence is not explanatory but constitutive’ (2007:47). He also favours the generic rather than the particular view of essences, and regards essences much more as features of daily life than as objects of scientific research. Achieving the Aristotelian aim of knowing ‘what it is’ is largely regarded as common sense. It is possible to read the Aristotelian corpus in this way, especially if it is filtered through the scholastic interpretation which Oderberg favours (and which favoured *Categories* as an authority), but the reading that places Aristotle’s essentialism in the context of individuals which have tangible causal power, and in the perspective of the *Posterior Analytics* views of explanation, demonstration and definition that amount to scientific enquiry, points strongly to explanation (largely of what is hidden) as the main motivator of Aristotle’s theory.

To meet the challenge that the concept of explanation is too diverse, we will draw attention to the causal nature of the explanations that led Aristotle to essentialism. His clearest and most basic pronouncements are these: ‘real enquiries stand revealed as causal enquiries (and the cause is the ‘what-it-was-to-be-that-thing’ [*to ti en einai*])’ (*Met* 1041a28), and ‘the substance [*ousia*] of each thing….is the primary cause [*aition*] of being for it’ (*Met* 1041b27). Thus we find in Witt’s summary that ‘the primary role of essences in Aristotle's theory of substance is causal, rather than classificatory’ (1989:179), Wedin concludes that an essence can perform its role ‘only if it is a cause’ (2000:416), and Gill says that ‘the solution to the problem of unity will finally depend upon Aristotle’s doctrine of form as an active cause’ (1989:9). Aristotle connects the causal character of essence with his aim of explaining change and predication in this remark: ‘the nature of a thing is a certain principle and cause of change and stability in the thing’ (*Ph* 192b20), and even goes so far as to say that ‘the fundamental duty of a philosopher is to gain possession of the principles and causes of substance’ (*Met* 1003a19). To simply say that the explanations involved are causal will not be sufficient to fully meet Lowe’s challenge concerning the hopeless diversity of explanation, but it narrows it down considerably, and if we emphasise that the explanations are intended to be ‘real’ and ‘objective’, this adds a further constraint. We will examine the concept of an ‘explanation’ more closely below.

It is also not possible to simply say that all explanations are causal in character, since explanations occur in the world of abstracta, and in the physical world we cite an absence, or the mere presence of things, or their structure, as well as active causes. This is no problem in the first instance, since the claim is only that the concept of essence arises for Aristotle out of a need to explain the basic entities of the physical world, and its value for us is also explanatory. The further (and much bolder claim) that all explanations (even of abstracta) are essentialist is also worth examining.

Aristotle himself found the ontology of mathematical abstracta perplexing, and he doesn’t generally discuss it in terms of explanation (though mathematics contributes proofs, which are the basis of demonstration (*PA* 79a3)). He takes a very naturalistic view of arithmetic, which he sees as abstracted from the particulars of the physical world (*Met* 1061a30 and 1078a22). His account of particulars tries to account for their unity, and once that is achieved each particular can be treated as a unit for counting purposes, so that 'it makes no difference whether we speak of the particular or the one in number; for by the one in number we mean the particular' (*Met* 999b33). Counting is what connects abstracted arithmetic to the natural world, and we can assume that explanations of arithmetic ultimately terminate in the physical world. He does, however, see mathematical objects in essentialist terms. 'Mathematics is concerned with forms' (*PA* 79a7), and we have seen that forms are essences. The link between the essences of abstracta and the essences of physical entities is to be found in the definitions. He tells us that 'something holds of an item in itself if it holds of it in what it is - e.g., line of triangles and point of lines (their essence comes from these items, which inhere in the account which says what they are)' (*PA* 73a35). That is, lines are intrinsic to triangles because they must feature in the definition of their essence. A common modern view of mathematics is that since the entire subject seems to consist of some sort of interconnected necessary truths, the subject is a modal level playing field in which nothing like essences can be distinguished. An older tradition, though, identifies with Aristotle's essentialist view of triangles and lines, and if essences are understood by their role in explanations, this older tradition looks worth revisiting, and will be addressed below. If explanation has a role to play in the study of abstract systems, then we should expect essences to make an appearance.

At this point we can venture a summary of the concept of essence which is to be found in Aristotle's writings, and which will give a basis to our subsequent discussions:

> **Aristotelian Essence**: a nexus of causal powers within an ordered or principled structure, which is identified as the single foundational level for a variety of satisfactory explanations of some entity. It not only persists through change, but enables that persistence, and is understood as what unifies it, supports its predicates, and fixes its generic kind. It explains why a thing has its surface features, and why it behaves in certain ways. The essential nature of such a thing is given optimal expression in an accurate, comprehensive and unique definition.

9. Aristotle and modern science

We have seen that Aristotle takes the identification of necessities to be revealed by an understanding of an essence, and that these necessities are preserved in the process of demonstration. We have also seen how explanation is largely understood in causal terms, and that the role of the 'form' in *Metaphysics* is almost entirely causal in character. It is these aspects of Aristotle's metaphysics which provide a strong connection to the modern philosophical movement known as 'scientific essentialism'. Gelman, a psychologist, observes that 'essentialism encourages a "scientific" mindset in thinking about the natural world, a belief that intensive study of a natural domain will yield ever more underlying properties' (2003:296), and this scientific essentialism has coalesced around the combination of a renewed interest in mechanisms as explanatory, an interest in the proposal that necessities may be empirically discoverable, and a growing realisation that certain modern scientific discoveries actually seem to coincide with the ‘forms’ which Aristotle proposed as the original target of scientific investigation.

The main concept of ‘scientific essentialism’ is very simple: the achievements of science, and its continued aims, are best understood as a revelation of the essential natures of the things that constitute the physical world. A further commitment of this approach is that ‘laws of nature’ are not imposed on passive matter, but actually arise from the nature of that matter, and that these laws will be necessitated by the matter. The orthodox (‘Humean’) approach to the laws, routinely claiming that the stuff of the world could remain the same, but the laws be quite different, is taken to be incoherent, since the stuff is the sole source of the laws. In general, scientific essentialists claim that science reveals a system of interconnected necessities in nature, which both entail that the possibilities in nature are far more precise and restricted than is normally thought, and also reveal exactly what those true possibilities are. The proposal is that the possibilities in nature should not be identified by means of what our imaginations can or cannot conceive (the prevalent approach to natural modality among empirically-inclined philosophers since the seventeenth century), but should be identified by scientific research. The result will mostly be disappointing for speculative thinkers, since fancied possibilities are regularly ruled out by the revealed facts. A simple example is that everyone can imagine a bonfire of wood burning on the surface of the moon, but no one can imagine wood combining with oxygen if there is no oxygen present. A real example offered to illustrate scientific essentialism is the claim that the essence of gold has actually been discovered, and consists of the nature of its atomic nucleus, and the structure of its electron shells. These give us the ‘nature’ of gold, and enable us to explain and predict its surface properties and interactive behaviour. We not only understand gold as a result, but we also see why the alchemists’ dream of transforming base metals into gold was doomed. A prize exhibit for defenders of scientific essentialism is the periodic table of elements (in which gold has its place), and we will examine its role in scientific explanation below.

Whether this meets Aristotle’s aspiration to find the ‘forms’ of things is open to discussion. He was certainly in tune with the modern approach, when he observed that ‘it would be strange for a natural scientist to know what the sun and the moon are, but to be completely ignorant about their necessary attributes’ (*Ph* 193b7), and ‘it is not possible for fire to be cold or snow black’ (*Cat* 12b1). Insofar as the forms are causal [*aitia*] and structural [*morphé*], the modern findings seem to fit the aspirations of Aristotle perfectly. Insofar as the forms are understood as principles [*arché*] that generate unified entities, modern science seems to offer a regress of principles which hesitate before an obscurity at the lowest level we can attain. Here we are reminded of the regress of essences that bothered Lowe, and some account of it must be given.

Recent scientific essentialists have tended to derive their views from post-1970 philosophy, and have paid lip-service to Aristotle, but not given him close attention. For example, Ellis (a leading champion), writes that ‘scientific essentialism is less concerned with questions of identity, and more with questions of explanation, than is the essentialism of Aristotle or of Kripke’ (2001:55), which is entirely contrary to the view outlined above. This modern view is understandable, but the quest for a closer view of Aristotle’s actual account is very illuminating, and a grasp of the way in which essentialism arises out of an explanatory project should be foundational for the scientific essentialist movement.

Earlier generations of philosophers not only understood Aristotle much better, but were also thoroughly sympathetic to the attitude to science which he promoted. It is only really with the advent of the positivist movement, with Comte and then the Vienna Circle, that the Humean approach has become dominant. Locke showed sympathy with scientific essentialism when he wrote that ‘which ever hypothesis be clearest and truest, ...our knowledge concerning corporeal substances, will be very little advanced.. , till we are made to see, what qualities and powers of bodies have a necessary connection or repugnancy one with another’ (*Essay* 4.3.16). Hegel, from a very different perspective, offered the view that ‘the movement of pure essences constitutes the nature of scientific method in general’, and that ‘scientific cognition demands surrender to the life of the object, or, what amounts to the same thing, confronting and expressing its inner necessity’ (1807: Pref 34 and 53). Nietzsche produced a similar remark, that ‘one must understand all motion, all 'appearances', all 'laws', as mere symptoms of inner events’ (*Notebooks* 36[31]). Twentieth century progenitors of scientific essentialism include C.I. Lewis (who writes in 1923 that ‘the scientific search is for such classification as will make it possible to correlate appearance and behaviour, to discover law, to penetrate to the "essential nature" of things in order that behaviour may become predictable’ (Thayer (ed):368)), and the splendid Irving Copi not only saw that ‘modern science seeks to know the real essences of things, and its increasing successes seem to be bringing it progressively nearer to that goal’ (1954:715), but also saw that Aristotle’s explanatory account of essence should be the basis of such a view.

The doctrine of scientific essentialism is not a novelty thrown up by recent accounts of reference in modal logic, but a thread which can be traced throughout the history of philosophy. More recent writers have, however, looked at the doctrine much more closely, and have brought a more sophisticated understanding of science and of modality to the discussion, even if they seem unaware that many of their disputes (e.g. concerning the foundational roles of dispositions or of categorical properties) are revisiting Aristotelian and scholastic discussions, with a modified vocabulary. Exponents of this modern view include Harré and Madden (1975), Ellis (2001), Molnar (2003), Mumford (1998, 2004), Heil (2003), Bird (2007), and Martin (2008).

The aim of the present enquiry is not to evaluate scientific essentialism, but to explore how an authentically Aristotelian metaphysics might offer the most coherent framework within which to think fruitfully about such realms of study. The aim of an ideal philosophical conceptual scheme should be to accommodate five interconnected areas: ordinary thought and talk about the world, precise and surprising theorising among scientists, the formalism found in mathematics and logic, the way in which we understand the semantics of various modes of language, and the psychological capacities of human beings. If the Aristotelian approach is seen as growing from an aspiration to explain things, and as a theory that this is achieved by studying the convergence of various modes of explanation on the essential nature of each thing, then it is a scheme of thought well suited to the role.

TWO

Crisis for Essentialism

10. The scholastics

The status of Aristotle in the later ancient world was such that there developed a huge industry of commentaries, and a Peripatetic school to follow his teachings, but they seem to have adhered fairly closely to the original thought, and their tradition was swept away when their schools in Athens and Alexandria finally closed. A dynamic movement of scholastic philosophy began around 1250, largely ended by the Church of Rome in 1347. Serious philosophy resumed with Bacon and Descartes around 1610. The present study is not concerned with the history of ideas, but it makes the assumption that tracking the history of an idea is the best way to understand it, with the first proposals showing most clearly what is involved, and the first objections throwing the best light on the problems. For a good understanding of genuine Aristotelian essentialism, we have examined what Aristotle seems to have actually said, and we will now look at the period of first disillusion with his views, to see what the key issues were. The period of initial disillusion is the beginning of the scientific age in which we still find ourselves, so the place of Aristotle in modern thought is highlighted particularly well in this period. As a guide to the flourishing and decline of scholastic philosophy we will make particular use of Pasnau's panoramic guide (2011), and we will then see how philosophers of the seventeenth century coped with the apparent collapse of Aristotelian doctrines. Once we have highlighted the central issues in this historical way, we will move on to an assessment of the place of Aristotelian essentialism in current thinking.

We will start by noting a few significant aspects of the scholastic movement. For the scholastic philosophers Aristotle had an exceptional authority, and was referred to as 'The Philosopher'. However, the religious context of their thought introduces a much more immanent concept of God, rather than Aristotle's remote Unmoved Mover, and their metaphysics was particularly influenced by a need to accommodate the doctrine of Transubstantiation, which said that the bread and wine used in the Eucharist were literally transformed into the body and blood of Christ. After consecration, it was held that the substances of the bread and wine no longer existed, but that their attributes survived (Pasnau:185). Hence Duns Scotus writes that 'accidents are principles of acting ….but it is ridiculous to say that something is a principle of acting … and yet does not have any formal being' (*Ordinatio* 4.12.1 – Pasnau:196). Thus the accidental features of a thing acquired their own status in ontology, and there were consequent difficulties for the doctrine of hylomorphism. The most important consequence of this for our purposes was that 'originally you count substances for ontology; once there is the doctrine of real accidents (in the fourteenth century) the list of ten categories begins to look like an inventory of the kinds of things there are, and *Categories* looks like the fundamental text' (Pasnau:222). In the thirteenth century Aristotle's *Categories* was viewed as a beginners' book, preparatory to reading *Metaphysics*, but this shift in emphasis seems to be the source of the later treatment of Aristotle (especially among Catholic thinkers) as what we would now call a 'sortal essentialist', whereas we have argued that his mature doctrine is that ontology rests on particular substances, understood in hylomorphic terms. The notion that definition needs little more than the establishment of genus and species (rather than tracking the differentiae as close to the particular as possible) followed on from this later scholastic view.

In addition to the adaptation of Aristotle to the needs of theology, the scholastics also struggled to achieve a complete account of his philosophy, and two areas gave them especial difficulty. The first was the question of whether an apparently unified entity contains a single form, or whether there might be further forms contained within it, and (if so) what the status of such things might be. Aristotle mentions the parts of animals as constituting natural objects (*Ph* 192b9), but the authority of Aquinas endorsed the view that a true entity could only have one form: 'if Socrates were animal and rational by different forms, then to be united they would need something to make them one' (*Q. De Anima* 11c, Pasnau:578). The rival view derives from Duns Scotus, and his formulation said that the subsidiary forms of a thing were allowable because they did not constitute an 'ens per se' (an entity in itself), because that would make them true particulars (*In Praed* 15.1, Pasnau:607). Thus we have less than complete forms within a single mastering form. This may sound bizarre until we encounter a later example from Suárez, that a tree presumably has a substantial form, but also a leaf or a fruit from the tree seem to have their own forms. Suárez's formulation talks of these as 'partial forms', which are 'apt to be united …to compose one complete form of the whole' (*Disp Met* 15.10.30, Pasnau:631). If we are to take hylomorphism seriously, this is an interesting puzzle. As so often, William of Ockham pursued the issue to more drastic conclusions, which threaten the whole hylomorphic picture, because he spotted a tricky case: 'when a piece of wood is divided in two halves, no new substance is generated; but there are now two substances, or the accidents of the two halves would be without a subject; they existed before hand, and were one piece of wood, but not in the same place' (*Seven Qs* 4.19, Pasnau:611). This is the kind of awkward question that was emerging, prior to the suppression of liberal university teaching in 1347. The interest here is the difficulty for hylomorphism if the phenomenon of predication is treated as its basis. Not only can each tiny sliver from a block of wood support its own private predicates (of shape, for example), but we quickly see that thoroughly disunited aggregates such as piles of bricks can support predicates (such as being chaotic). Aristotle's desire for a subject of predication throughout his metaphysics should be treated cautiously.

The second notable area of difficulty in the scholastic reading of Aristotle is the question of 'prime matter'. Sadly, Wedin takes a key Aristotelian passage in support of prime matter to have been misread (2000:190), and Gill's account concludes that 'prime matter has no place in Aristotle's elemental theory; ...references to prime matter are found in Aristotle's work because his theory was thought to need the doctrine; if I am right, these passages will all admit of another interpretation' (1989:252). Gill argues that the elements, rather than 'prime matter', are fundamental in Aristotle (42). Nevertheless, scholastics wrestled with the problem of what matter could be when it was devoid of form, and found themselves driven towards quasi-mystical language for its ontological status, since the role of form was to bestow true actuality.

For example, Aquinas quotes Avicenna (Abu Ibn Sina) as saying that ‘the ultimate material of things has the unity of total formlessness’ (Aquinas 1993:97), and Averroes (Ibn Rushd) says that ‘prime matter falls halfway, as it were, between complete non-existence and actual existence’ (*on Phys* 1.7, Pasnau:38). Peter Auriol tells us that prime matter is ‘indeterminately and indistinctly a material thing’ (*Sent* 2.12.1.1, Pasnau:39). Those early views set the agenda, and the late scholastic Eustachio a Sancto Paulo concluded that ‘everyone says that prime matter, considered in itself, is free of all forms and at the same time is open to all forms’ (*Summa* 3.1.1.2.3, Pasnau:35). The significance of these obscure views for our purposes is that when Aristotelian doctrines came under fierce scrutiny in the early seventeenth century, it was the scholastic account of matter which appeared to be the source of the problem. Rejection of this pseudo-Aristotelian view of matter was central to the scientific revolution, and to the critique of hylomorphism that went with it.

Apart from such problems, there was also a shift in emphasis in the understanding of substantial forms, away from the ‘principles’ that Aristotle was partly concerned with, and more towards the causal character for a form. There is a deep issue over the whole doctrine of Aristotelian essentialism which arises at this point, and will recur in subsequent discussions. Leibniz claimed that substantial forms were dubious in physics, but indispensable to metaphysics (in *Discourse* §10), and Pasnau follows him in this view, portraying scholastics as having misguidedly portrayed hylomorphism as a theory of physics (when Aristotle actually presented his theory as metaphysics), and thus opening it to the criticisms of the new experimental physics (p.538). We have argued above, however, with good support from modern scholars, that Aristotle was very much concerned with causal issues in his theory of forms. The best reading of Aristotle seems to be that hylomorphism is a theory of physics *and* metaphysics, precisely because there is not taken to be a sharp division between the two (a division assumed by Leibniz and Pasnau). Nevertheless, Pasnau has picked out a difficulty with scholasticism, which brought trouble for the Aristotelian approach. Albert the Great wrote that ‘there is no reason why the matter in any natural thing should be stable in its nature, if it is not completed by a substantial form; but we see that silver is stable, and tin and other metals; therefore they will seem to be perfected by substantial form’ (*On Minerals* 3.1.7, Pasnau:561). This remains a persuasive claim (that natural kinds need some sort of ‘form’ to support their striking stability), but less persuasive is the example given by William of Ockham, that ‘it is clear to the senses that hot water, if left to its own nature, reverts to coldness; this coldness cannot be caused by anything other than the substantial form of the water’ (*Seven Qs* 3.6, Pasnau:561). In general, Pasnau sees a steady slide during the scholastic period away from emphasis on the ‘formal’ cause/explanation in Aristotle to an emphasis on the ‘efficient’ cause/explanation, and a tendency to treat forms as much more substantial than Aristotle had intended (p.549). The problematic result of this move is that hylomorphism is presented as a straightforward research programme for science, rather than as an overview of our grasp of nature, and in that former guise it is not up to the job, and was rejected as soon as the new experimental philosophy got under way.

We should not presume that among scholastics there was a slavish adherence to the perceived doctrines of Aristotle. In the early fourteenth century there was radical criticism that anticipated the corpuscularian approach of the early seventeenth century, notably in the bold independence of Nicholas of Autrecourt, whose books were burned in 1347. One interesting question that arose among the critics was whether unity might come in degrees, rather than the absolute unity implied by hylomorphism, and the Coimbran commentators (in Portugal in the 1590s) summed up an emerging possibility by offering five degrees of unity: by aggregation (stones, for example), by order (an army), per accidens (inherence), per se composite unity (connected), and per se unity of simple things (Pasnau:556). We will find Leibniz making a stand against the implications of this graduated view of unity, and it remains a central issue for any metaphysical discussion of objects, whether or not the spirit is essentialist. Duns Scotus said he believed that ' "unity" is one of the more difficult words in philosophy, for there are in things many hidden (*occultae*) unities that are obscure to us' (*Lect* 1.17.2.4, Pasnau:208), so there was no complacency that Aristotle had solved that problem. Similarly, the hope that substantial forms somehow solved problems of causal explanation is a long way from this gloomy remark of Roger Bacon's: 'no one is so wise regarding the natural world as to know with certainty all the truths that concern the nature and properties of a single fly, or to know the proper causes of its color and why it has so many feet, neither more nor less' (*Opus Maius* 1.10, Pasnau:543). That the essence of animals was fixed by membership of a species was orthodoxy for several centuries, but Francis of Marchia wrote 'let all accidents be removed from a lion and a horse; nothing remains in the intellect to distinguish them; we distinguish a lion and a horse only by analogy to the accidents proper to each; the intellect does not have an essential concept of either one' (*Sent* 1.3.1, Pasnau:127). Francis predates Locke by three and half centuries.

When the wholesale attack on Aristotle finally arrived, the scholastics offered some easy targets, and their uneasy interpretations had shown where the problems lay, but they also offered sharp insights into how essentialism might be reformulated, and were quite realistic about the gulf between aspiration and reality when it came to substantial forms.

11. Rejection of Aristotle

The rush of criticism that accompanied the emergence of experimental sciences focused especially on three aspects of Aristotelian metaphysics – prime matter, substantial forms, and teleology. In addition there was growing doubt about Aristotle's detailed theories of the physical world. For example, the idea (mentioned earlier) that life spontaneously generates in rotting meat was demonstrated to be false, and Arnauld and Nicole cite rejection of his view that nerves centre on the heart (which modern anatomy had disproved), and his view that 'the speed of heavy things increases proportionally to their weight' (which had famously been disproved by Galileo) (1662:20).

The formulations of the concept of prime matter which we saw above made that an easy target, and Francis Bacon swept it aside with the remark that 'stripped and passive matter seems nothing more than an invention of the human mind' (*Phil Studs* 1611-19:206, Pasnau:123). Hobbes similarly reduced prime matter to an abstraction from bodies (*De Corp* 8.24, Pasnau:72). We could ignore this mere dismissal, if it were not that the nature of matter was of the utmost interest to the new thinkers. The heart of the new doctrine was Corpuscularianism, which is the view that almost everything physical will be explained by the ways in which the parts of matter combine amongst themselves by means of 'force', without help from some additional entity like a substantial form. This view is behind Vanini's 1615 remark that 'the whole of prime matter, considered as prime matter, is nothing other than its parts', which comes from a very late scholastic philosopher, and not a new scientist, showing the collapse of the tradition (*Amph* ex 5, Pasnau:40). The really revealing remark is another from Bacon: 'prime, common matter seems to be a kind of accessory and to stand as a substratum, whereas any kind of action seems to be a mere emanation of form; so it is that forms are given all the leading parts' (op. cit.). The proposal of the new thinking is simple: scholasticism offers obscure matter, and supposedly illuminating forms, when in truth the forms are utterly obscure, but there is a real prospect of understanding matter. In other words, Aristotelianism is rejected as a research programme.

There are three closely related ideas at the heart of Aristotle's metaphysics – substance, essence, and substantial forms – and it is important for this period to understand that it is only the third of these which came in for vigorous criticism. The idea that there were 'substances', unified entities which constitute the world, and that there are 'essences', hidden natures which characterise the substances, remained perfectly respectable concepts, employed in generalised contexts throughout the seventeenth century. The specific target of hostility was the idea that there is some very specific entity called a 'form', which might deliver understanding if it were investigated closely. Bacon became dubious about substantial forms, but it was the brusque dismissal by the great Descartes which was most influential; for example, he wrote to Regius in 1642 that 'clearly no explanation can be given by these substantial forms for any natural action, since their defenders admit that they are occult and that they do not understand them themselves, ...so they explain nothing' (quoted by Oderberg 2007:267). We are defending the concept of an essence as the focus of successful explanation, but for Descartes that is exactly what they failed to do, at least if they are understood as Aristotelian intrinsic forms. There are two modern responses to that, for the modern essentialist: either that Descartes has been proved wrong, and that science has been successfully investigating forms without quite realising it, or that explanation by Cartesian routes (which ignore 'forms') has turned out to be successful, but that the resulting explananda are still exactly what Aristotle meant by an essence. Subsequent chapters will examine both explanation, and how we should now view essences, in quest of the right response to that question. Boyle also turned the new community of scientists away from substantial forms. He offered the more considered criticism that 'If it be demanded why rhubarb purges choler, snow dazzles the eyes rather than grass etc., that these effects are performed by substantial forms of the respective bodies is at best but to tell me what is the agent, not how the effect is wrought' (1666:68). This at least supports the point made earlier about the dormative powers of opium, and recognises hylomorphism as a possible research project, but the failure to tell us how it is done was precisely Descartes's complaint. Boyle's response to this impasse gives us the strategy developed for the new age, of concentrating on the matter, instead of the form: ‘the form of a natural body being, according to us, but an essential modification and, as it were, the *stamp* of its matter, or such a convention of the bigness, shape, motion (or rest), situation, and contexture (together with the thence-resulting qualities) of the small parts that compose the body’ (1666:69). From Boyle’s time onwards, substantial forms are expelled from science, and Hume’s contemptuous aside gives the standard view – that the Aristotelian system in this respect is ‘entirely incomprehensible’ (1739:1.4.3).

The difficulty for the Aristotelian teleological approach to nature was that Aristotle’s own examples were often implausible, such as the proposal that rain falls in order to make crops grow (*Ph* 198b16). Once you begin to study air pressure and so on, mechanistic explanations for rain begin to emerge, and rain is seen as pushed by causation, not pulled by purpose. This neglects the point that Aristotle’s ‘final’ causes also cover the concept of a functional explanation. While such things are not much invoked in modern physics, they certainly seem meaningful in biology, and the teleological approach seems to have received over-harsh treatment in the seventeenth century. Bacon rejected final causes on the grounds that ‘to say 'leaves are for protecting of fruit', or that 'clouds are for watering the earth', is well inquired and collected in metaphysic, but in physic they are impertinent. They are hindrances, and the search of the physical causes hath been neglected’ (1605:113). Descartes is his usual brusque self (‘we shall entirely banish from our philosophy the search for final causes’ (*Princs* 1646:1.28)), and for Spinoza final causes were ‘nothing but human fictions …for that which is in truth the cause it considers as the effect, and vice versa’ (1677:I App). Leibniz, as we will see, had a rare good word to say for final causes, but the mocking attitude to teleology greatly accelerated the decline of the Aristotelian account of science.

12. New science

A full survey of the new science would take us too far afield, but certain aspects of it are important for the present enquiry. If nature is not to be explained by the hidden essential natures of the things in the world, we need to see how the new explanations worked, and the extent to which they could replace the older view. We can summarise the new approach very concisely: explanations will henceforth be by means of physical mechanisms, mathematically expressed relationships, and laws of nature. The means to achieve the explanations will be empirical observation, controlled conditions experiments, and the tracking of nature’s causal pathways.

The resort to ‘mechanisms’, and the new ‘mechanistic philosophy’, rested on the corpuscular view of matter mentioned above. Pasnau’s summary is that ‘according to strict corpuscularianism the only real constituents of a substance are its integral parts’ (606), and the core of the corpuscularian approach is seen in Newton’s view that ‘the attractions of the bodies must be reckoned by assigning proper forces to their individual particles and then taking the sums of those forces’ (1687:I.II.Schol). In Hobbes’s case the corpuscularian approach to matter was supplemented with wholesale materialism, so that nature consists entirely of corpuscles (as in ‘the world is corporeal, that is to say, body...and every part of the universe is body, and that which is not body is no part of the universe' (1651:4.46)), but most thinkers felt an imperative to resist such an implication. The main point is clear – that if an account can be given of how the interior parts of anything relate together by forces, and can then map these parts and forces, explanations may well drop into our lap.

The 'map' which would produce these explanations would be expressed in mathematics. Galileo famously claimed that this was the language of the book of nature, Descartes confirmed the approach when he wrote that 'I do not accept or desire any other principle in physics than in geometry or abstract mathematics, because all the phenomena of nature may be explained by their means, and sure demonstrations can be given of them' (1646:164), and Newton (the greatest practitioner of the new approach) wrote that 'the moderns - rejecting substantial forms and occult qualities - have undertaken to reduce the phenomena of nature to mathematical laws' (1687:Pref). It is the series of mathematical equations (about pendulums, gas pressures and gravity) which convinced the neutral that the mathematical route was the one to take, and there seemed no prospect of giving a mathematical account of hylomorphism, so at this point the game seemed to be up for Aristotle.

The most interesting aspect of the new approach, though, is the appeal to the idea of a 'law'. It is not clear where this idea developed, but the concept of natural moral law appeared long before this period. Aristotle invokes 'natural' justice (*Eth* 1134b18), and Annas says that the Stoics are the main source of social and political 'natural law', characterised as universal right reason (1995:302). In an isolated remark, Lucretius said that 'nothing has power to break the binding laws of eternity', but this is probably a metaphorical invocation of universal necessity (c.60 BCE:5.56). Lange, in a book on the laws of nature, offers Hooker (in 1593) as the beginnings of the new approach to laws of nature, though his supporting quotation seems more like hylomorphism than the new revolution, since natural things obey laws only 'as long as they keep those forms that give them their being' (2009:6). The best historical account seems to be the one most generally accepted – that the concept of 'laws of nature' came to dominate all of subsequent science because of the work of Descartes, and we find Newton saying that 'the (active) principles I consider not as occult qualities, supposed to result from the specific forms of things, but as general laws of nature, by which the things themselves are formed' (*Qs on Optics*:q31, Pasnau:544). The obvious question, for anyone with metaphysical inclinations, concerns the nature and groundings of such laws, and it is not surprising that in the seventeenth century the contribution of God was invoked in this context, so that Descartes introduces his commitment to laws in just such terms: 'I have noticed certain laws that God has so established in nature, and of which he has implanted such notions in our souls, that …we cannot doubt that they are exactly observed in everything that exists or occurs in the world' (*Disc* 1637:§5). Newton is quite explicit about the religious underpinnings of the laws he describes, when he writes that 'this most elegant system of the sun, planets, and comets could not have arisen without the design and dominion of an intelligent and powerful being' (1687:3.Gen Schol). Effectively, the question of what we should take these new 'laws' to be is pushed to the margin at the beginning, but only by invoking forces outside of nature in a way that Aristotle had avoided. We should note that Newton invokes not only the 'design', but also the 'dominion' of a supreme being, and this implies the view that matter is passive, and the activity of nature is driven by the laws. Ellis cites Euler in the 1760s as saying that ‘the powers necessary for the maintenance of the changing universe would turn out to be just the passive ones of inertia and impenetrability; there are no active powers, he urged, other than those of God and living beings’ (2002:62). If the laws are seen to be transcendent in origin, then of course the precise mathematical statement of a law of nature is a startlingly illuminating explanation, beyond which we could never hope to advance. As we will see below, though, modern theorists have attempted to apply laws in explanations without the traditional supernatural support, and that reopens the questions about explanation that are addressed here.

13. New scepticism

The scientific age was launched by the infectious optimism of Bacon, and experimentalists like Boyle (a devoutly religious man) worked away at their apparatus without too many worries about background theory, since the substitution of corpuscularianism for hylomorphism seemed to be all that was required. For theoreticians away from the empirical front line, though, doubts and problems began to surface. Accompanying the new scientific attitude was an overt empirical philosophy, seen in Bacon, Gassendi and Hobbes, but a refusal to transcend fairly immediate and accessible experiences closes many of the routes to traditional understanding. In addition, a vein of sceptical thought from the ancient world had resurfaced, in the discovery of the writings of Sextus Empiricus, and Montaigne was voicing a great deal of scepticism about metaphysical questions before the scientists entered the stage. Much the most articulate sceptic was Hume, whose great work appeared in 1739, but many of the sceptical themes in Hume had already been voiced. The famous doubts about induction, for example, appeared in the new translations of Sextus, who wrote that ‘induction cannot establish the universal by means of the particular, since limited particulars may omit crucial examples which disprove the universal, and infinite particulars are impossible to know’ (c.180 CE:II.204). Hume clarified the circularity of any attempt to justify induction from direct experience, and showed that the reliance on a mere increasing repetition of experiences suggested that the basis of induction was psychological, rather than logical. Since Newton had written that ‘in experimental philosophy, propositions are deduced from the phenomena and are made general by induction’ (1687:3 Gen Schol), this seemed to leave physical sciences without a decent conceptual foundation.

Hume’s well-known doubts about the supposed necessity in causation had likewise been anticipated by Hobbes (‘in knowing the meaning of 'causing', men can only observe and remember what they have seen to precede the like effect at some other time, without seeing between the antecedent and subsequent event any dependence or connexion at all’ (1651:I.12)), since the question is implicit in any attempt to give an empirical account of our study of nature. The questions about induction and causation are focused most vividly in sceptical empirical questions about the status of the so-called ‘laws’. Berkeley, for example, saw that empirical evidence set limitations on our ability to infer absolute laws of nature, when he wrote that ‘the set rules or established methods wherein the Mind we depend on excites in us the ideas of sense, are called the 'laws of nature'; and these we learn by experience, which teaches us that such and such ideas are attended with certain other ideas' (1710:§33). He adds that the exceptional regularity of the ideas points to a Divine authority, and it is unlikely to be coincidental that the sceptical question was best articulated by Hume, because his atheist tendencies left him puzzled as to what sort of 'dominion' over nature the laws could offer. An application of his empirical tests to the matter pointed to the view, now widely held, that laws are nothing more than descriptions of regularities among our experiences.

Not only was the presence of necessity in induction, causation and laws thrown into doubt, but the very distinction between contingency and necessity seemed insupportable in a corpuscularian or empirical context. Hence we find Hobbes saying that the apparent contingency of a traveller being caught in the rain is mere ignorance of the separate causes bringing the two together (1654:95), and both he and Spinoza became notorious for their flat rejection of all contingency, and of free will with it. Locke, too, cannot find empirical evidence for free will, since 'a man is not at liberty to will or not to will, because he cannot forbear willing' (*Essay*:2.21.24). The culminating view, as usual, is expressed by Hume: 'necessity …is nothing but an internal impression of the mind' (1739:1.3.16). The rivalry of rationalist and empiricist approaches is particularly sharp over the question of the source of necessity. Rationalists take necessity to be the product of a priori understanding, with its very hallmark being that it is amenable to knowledge by pure reason, as in Leibniz's remark that 'there are two kinds of truths: of reasoning and of facts; truths of reasoning are necessary and their opposites impossible; …a necessary truth is known by analysis' (*Monadology*: 33). The source of necessity for the great rationalist philosophers will always be God, with God-given reason our means of grasping it. The empiricist foundation for necessary truths is the imagination, as in Hume's remark that 'whatever the mind clearly conceives includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible' (1739:1.2.2). In the empiricist case, the external source of modality is generally unknown, and necessity rests simply on the falsehood being unimaginable. The Aristotelian approach is to seek the grounding (and explanation) of modality in the natural world, rather than in a divinity or in the human mind (either as conceivability, or as convention), so this issue becomes of great interest to us.

Hylomorphism had offered a highly integrated picture of the mind-body relationship, since the mind is the form of the body, inheriting the broad concept of 'mind' from the Greek *psuché*, which is possessed even by plants, and seems to include its 'life' as well as any consciousness and reason. The Christian doctrine of soul was kept separate from this Aristotelian idea. Once 'forms' were disallowed, the picture based on *psuché* had to be dropped, and corpuscularian philosophy then threatened to sweep the field with a highly materialist view of the mind (a materialism fearlessly embraced by Hobbes). This scepticism about the mind was quickly countered by Descartes's famous defence of mind-body dualism, and the next century saw something of a standoff between two radical approaches to the mind, which included Malebranche's Occasionalism, Leibniz's Parallelism, Spinoza's Dual-Aspect Monism, and La Mettrie's thoroughgoing Physicalism. An interesting consequence of this rift across the fairly unified account of creation which the medieval mind had developed was Descartes's startling thought that animals might be entirely mechanical entities (as when he remarked that beasts 'have no reason, and perhaps no thought at all' (*Pass* 1649:I.50)).

A second line of defence for the status of humanity in the face of a nature made of nothing but 'corpuscles' was Locke's proposal that 'person' was a distinct ontological category (*Essay* 2.27.9), which he seems to have offered because he was unable to defend the Cartesian dualism which others found adequate to keep humanity away from mere soulless matter. This strategy also met with scepticism from Hume, who offered his famous view of the self as nothing more than a 'bundle' of experiences, amongst which no unifying principle could be discerned (1739:1.4.6).

An interesting final scepticism from the period is a surprising one. At the dawn of the golden age of science we find both Locke and Hume highly pessimistic about the prospects for the future of science. Locke wrote that 'as to a perfect science of natural bodies (not to mention spiritual beings) we are, I think, so far from being capable of any such thing, that I conclude it lost labour to seek after it' (4.3.28), and saw no prospect at all of our ever predicting all of the properties of gold from the ones we are able to observe (4.3.14). Hume saw no prospect of ever discovering why bread nourishes us, and wrote that 'the ultimate springs and principles are totally shut up from human enquiry; elasticity, gravity, cohesion of parts, communication of motion by impulse; these are probably the ultimate causes and principles which we shall ever discover in nature' (1748:4.1.26). We might attribute such pessimism simply to their temperaments, but actually belief that science could make little further progress seems to be implicit in the wave of sceptical questions that arose about the scientific endeavour. Since physical science has made progress beyond the wildest dreams of even the optimists from the early period of science, this will give us grounds for suggesting that they got the story wrong. We must investigate further.

Science was arriving in triumph, but the philosophers of the period were gripped by sceptical puzzles about the new approach, and the consequent fate of metaphysics since that period has been much less happy. Since the time of Kant the subject has either survived in a cautiously minimalist form, or else been consigned to oblivion. In the late sixteenth century (the period of Suárez) it seemed that metaphysics was simply a matter of fine-tuning the Christianised Aristotelian account, and that serious thinkers possessed a thoroughly comprehensive and satisfying picture of the underpinnings of reality. A hundred years later we are facing what Pasnau describes as 'a metaphysical train wreck' (2011:632). What he particularly has in mind is that the older worldview is built entirely around the behaviour of objects, and these have an intrinsic unity which not only gives them individuation and persistence conditions (p.654), but generates the causal powers which we use to explain natural behaviour, and gives us the basis for the natural kinds which are the heart of our relationship to the natural world. The 'wreck' is the result of pulling the mat out from under this foundation, by challenging the very idea of the unity of an object, and the causal essence which is at its heart. We no longer seem to have a criterion for either the boundaries of an entity, or for its unity. The corpuscular account is seen in Newton's observation that 'the particles of bodies attract one another at very small distances and cohere when they become contiguous' (1687:3 Gen Schol), which seems to offer no distinction between a leaf and a pool of mud.

In addition to our new inability to hang on to the distinct existence of each component of our experience, we have also sketched a drift towards a thoroughly unsatisfactory picture of the world, in which matter has become the passive servant of laws of nature, but those laws of nature have been reduced to mere regularities in our impressions. Only the imposition of divine command can offer a driving force for such a cosmos, but a view of nature as the continual subject of divine intervention (the 'occasionalist' doctrine) was quite opposed to the direction in which most theologians wished to move, as well as being anathema to those (such as Hobbes and Hume) with creeping new doubts about the very status of religion. Philosophy found itself in a state best described as 'crisis', in which no one could offer a persuasive framework to explain the physical behaviour of the world, and find a plausible place for mankind within it.

We must not, though, adopt a caricatured account of what the new mathematical physics was achieving, as if it were just some neat mathematics that fitted the flickering patterns in a vast mass of corpuscles. It is obvious that a mathematical relationship in nature will facilitate exciting new predictions, such as the arrival of comets, but the important question facing us is whether the new findings helped to *explain* the world. That it does just that can be seen in the 1693 letter that Leibniz wrote to Newton, saying that 'you have made the astonishing discovery that Kepler's ellipses result simply from the conception of attraction or gravitation and passage in a planet' (Newton 2004:106). More famously, Newton's equation showed a connection between falling apples and planetary orbits, but what we should attend to is not that there is concise mathematics involved, but that the mathematics has revealed a *connection* in nature – and one which the contemplation of the contrasting essences of planets and of apples was unlikely to have ever revealed. In this way, the new science does indeed offer explanations which the old approach could not match. Even if we accept that science is merely the study of regularities, if the mathematical approach reveals hidden regularities we would never otherwise have imagined, there is a huge leap forward in understanding. We will consider below how this story has unfolded, but for now let us return to the 'train wreck', and consider the reactions of two philosophers who made prodigious efforts to find a metaphysics for the new world view that was emerging – Locke and Leibniz. For a thoroughly sceptical view we can study Hume, but Locke was an admirer of Boyle and was keen to create a system that acknowledged the ancient framework while embracing the new dogma of empiricism. Leibniz fought a fascinating rearguard action against the way things seemed to be moving, and throws considerable light on how we should view the Aristotelian project within a changing scientific picture.

14. Locke's ambivalence

Commentators on Locke routinely find it very difficult to pin down the whole of his philosophical system with any precision, because there are frequent shifts of emphasis, and even apparent contradictions, within the *Essay*. This should probably be seen as a reflection of the quandary he found himself in at a difficult moment for theoretical philosophy. Locke learned his philosophy at the latter end of an Aristotelian era of thought, but was greatly influenced by the new science, and worked in Oxford, where Boyle and Hooke performed their early researches. Since Locke formulated his ideas before Newton achieved fame, he was more interested in experimental attempts to understand matter than in the grand sweep of mathematical cosmology. He also quickly fell in with the strongly empiricist assumptions that had been championed by Bacon, Hobbes and Gassendi (rather than the more dominant rationalism championed by the famous Descartes). What is distinctive about Locke's empiricism is that he is a thoroughgoing realist about the external world, with no interest in sceptical arguments on that topic, and no interest in the idealism found in Berkeley, or the phenomenalism found in Hume. This meant that for Locke explanations (the focus of our discussion) were not to be expressed entirely in terms of experiences and ideas, but were always felt to connect with the real world where Aristotelian essences were said to reside. Thus, in a rare comment on the idea of laws of nature, he expressed dissatisfaction with the regularities of modern Humean orthodoxy, because explanations should go deeper: 'the things that, as far as observation reaches, we constantly find to proceed regularly, do act by a law set them; but yet by a law that we know not; ..their connections and dependencies being not discoverable in our ideas, we need experimental knowledge' (4.3.29). The Aristotelian concept of a substantial form had by this date suffered irreparable damage, but the idea that things had essences still flourished, and so finding a place for such things within an empirical system had the highest priority for Locke.

From among the multifarious roles played by essences in Aristotle's account, the three that most interested Locke were their contribution to the unity of objects, their causal underpinning of surface properties, and their role in categorising the objects. Locke was an unabashed nominalist, in that he committed to the central nominalist dogma, that 'all things that exist are particulars' (3.3.1). He never wavers from this view, so throughout his discussions of kinds, categories, species and sortal concepts it must be remembered that he agrees with the approach to Aristotle we argued for above, that the individual is primary, and the category is secondary (no matter how much the latter may dominate our ideas and speech). Locke has an equally strong commitment to the unity of his particulars, but empirical grounds for asserting such unity proved hard to articulate. A survey of his remarks on the subject show a remarkable diversity, since he claims in various places that an object is unified 1) by having a unique origin (2.27.1), 2) by having a spatiotemporal location (2.27.1, 2.27.3, 4.7.5), 3) by being structurally unified (2.23.17), 4) by having a unity in the idea it produces (2.16.1), 5) by the act of perception (2.27.1), 6) by a necessary precondition of categorisation (3.6.28), 6) by being mereologically unique (2.27.3), and 7) by an arbitrary imposition of the mind (2.24.3). The last view is worth quoting ('there are no things so remote, nor so contrary, which the mind cannot, by its art of composition, bring into one idea, as is visible in that signified by the name 'Universe'' (2.24.3)), to show the despair over the problem that occasionally struck him. The remaining offerings fall into two groups in the way distinctive of empiricist theories (such as Hume's definitions of causation), with some finding the unity in our mode of thought, and others finding unity in the presumed objects of the thoughts. We should not presume hopeless confusion from the variety of proposals here, since Locke, if challenged, might be at liberty to defend all of them, under an umbrella account which speaks of self-evident unities which are primitive in all our experiences of reality. We have to say that the idea of unity is 'primitive' for him, since he has no theory of real unity, other than a cautious commitment to groupings of corpuscles. The one thing which does not appear (as far as I can discover) in any of his direct accounts of unity is the concept of real essence, presumably because such things had failed in scientific explanations, and were not accessible to empirical investigation.

Unexpectedly (given Locke's refusal to cite them as explanations of unity), Locke believed in real individual essences. When it came to what he considered the most important aim of the new research into the hidden nature of matter, the prediction of properties was his key test; ideally, by knowing four main properties of gold, we should be able to predict a fifth property (4.3.14). This prediction would arise if we could discover intrinsic dependence relations within a physical thing, which would lead to the essence, defined as 'the real internal …constitution of things, whereon their discoverable qualities depend' (3.3.15). Such an aspiration looks like a belief that the discredited substantial forms are what are needed for the job. We might surmise that while scientific developments in mathematical cosmology swept substantial forms away, the developments in the understanding of matter were not so hostile. As we will see below, the idea remains implicit in many developments in chemistry. For Locke, though, with very limited information about the structure of matter, the prospects for explanatory essences looked highly desirable but hopeless. We have seen his pessimism about the future for scientific investigation of the essence of matter, so we can assume that we must look elsewhere for our understanding of a thing's properties and categories. Pasnau, however, identifies an argument in Locke which runs much deeper, and is highly relevant to the present discussion, since he claims that Locke has thoroughly undermined any possibility of arriving at a concept of essence as 'explanatory' (2011:27.7). The argument against explanatory essences which Pasnau finds in Locke is a much more significant challenge than the brief complaints of Lowe and Oderberg that we met with earlier, so it must be considered carefully. According to Pasnau, Locke demonstrated that the discovery of explanatory individual essences is not only unlikely to happen, but is actually impossible.

Pasnau concedes that he is presenting a new slant in Locke scholarship, but we can take the argument on its merits. The gist of the argument is that Lockean essences are individual, and to define an individual essence one must analyse 'all the way down', to capture all the nuances of the thing that make it that individual, in the manner that we indicated for Aristotelian definition. On the other hand, to decide the kind or the species of an individual thing, we only need to identify the more general features which all things of that kind have in common. In the case of the thing's kind, it is obvious which features are the 'essential' features and which the 'non-essential', since the former are the features which are never absent from any member of the kind. The difficulty is with the individual case, since you find yourself specifying every single feature of the thing, including its accidental properties, and there is then no possible criterion for deciding which features are the essential ones. The key passage in Locke is his statement that 'particular beings, considered barely in themselves, will be found to have all their qualities equally essential, and everything, in each individual, will be essential to it, or, which is more true, nothing at all' (3.6.5). Pasnau takes this argument to be a conclusive refutation of the explanatory account of essences. The conclusion of the argument is that kind essences are the only possible basis for essentialism, but this type of essence only offers categorisation, and not explanation.

In Locke's case, this places him on a slippery slope. He had taken the distinction between real and nominal definitions (expounded in Arnauld and Nicole 1662:1.12), and applied it to distinguish between real and nominal essences. His aim was to explain how our system of categorisation works, given that the real essences upon which categories were traditionally based are too obscure to ever be known. Categorising by means of real shared inner features is rejected using the analogy of watches, which seem to be a single species, and yet vary greatly in their inner workings (3.6.39). The nominal essence is the ideas which we have in our minds of the main defining features of a thing, rather than the elusive real features of the thing, and according to Locke 'our ranking, and distinguishing natural substances into species consists in the nominal essences the mind makes, and not in the real essences to be found in things themselves' (3.6.11). Thus he gives up on the essences of the particular things to which he is committed, and he then gives up on categorising things by the real features that they have in common, which was his only hope for essence. Having abandoned real particular essences, he then gives up on real kind essences, and is left with unreal nominal essences. We recall his most pessimistic remark about unity (that the mind can unify anything it likes), and a similar pessimism about real categories also underpinned his thinking: 'in the visible corporeal world we see no chasms or gaps. All quite down from us the descent is by easy steps and a continued series of things, that in each remove differ very little from the other. There are fish that have wings, and birds inhabit water' (3.6.12). Categorising by nominal essence is, on the whole, done by convention, and so the upshot is that Locke has largely rejected essences, despite an initial commitment to real essences in individuals. Real essences play the role of backdrop to his metaphysics, rather as the concept of the 'noumenon' functions for Kant. It is notable that when Locke added material on identity for his 1694 edition, in introducing his discussion of the identity of persons, the concept of an essence is not mentioned (though Pasnau wonders whether Locke's famous definition of a person (2.27.9) might qualify very well as the definition of a real essence! 2011:725).

The problem here for Pasnau's claim is that what is presented as an interesting argument in favour of kind essences and against individual essences may well lead into a rejection of every sort of real essence, so that explanatory essences are rejected simply because all essences are rejected. It is certainly the case that individual essences offer causal powers, and connection with the actual fabric of reality, and hence offer explanatory possibilities which mere mention of the shared properties of some species will not achieve. 'It's ferocious because it's a tiger' is not much of an explanation, even if you add 'and all tigers have the feature of ferocity'. The key difficulty of the Locke/Pasnau argument remains, though, which is that if the essential features of the individual cannot be distinguished, then no explanation which arises from those features can qualify as 'essentialist'. In subsequent chapters we will address the topics of explanation, and of the best way to understand essence, and this difficulty must be confronted there.

A few further points about Locke's discussion are worth mentioning. He is responsible for introducing the word 'sortal' into the philosophical vocabulary, simply as a term which specifies what 'sort' of thing some individual entity is (3.3.15). Sortal terms figure in our expressions of a nominal essence, and we can detect a faint worry about circularity in the fact that, as we saw, the ranking of things into species is done by the nominal essence, but the expressed nominal essence contains sortal terms which intrinsically rank things. The sort of question which regularly bothered Plato, and should bother us, is how you can decide which sortal terms to apply to something if you haven't already sorted it, and how could you manage the sorting without the sortal terms? David Wiggins has placed the concept of a sortal term at the centre of his modern account of essences, so we will examine such questions when we try to find the right concept of essence for this project.

For Aristotle, living things were the paradigm cases of unified entities, and we suggested that for Aristotle there simply didn't seem to be many rival candidates. With the rise of corpuscularianism, followed by the triumph of atomism and the emergence of modern chemistry, questions of obvious unity began to extend into the world of matter, but there seemed nothing to say about it (for example, the 41 occurrences of 'corpuscle' in Boyle 1666 all take the concept of a corpuscle as a small unity of matter for granted, without addressing what unifies it). The status of life as a special sort of unity, which was a presupposition of Aristotle's discussion, began to need a defence, and the history of the concept of 'life' is a particularly interesting story. It is therefore significant that in his 1694 additions to the *Essay*, Locke chose to denote a 'life' as a primitive unity, which was distinct from the unity of inorganic objects (which was now expressed mereologically) and the perfect unity of persons and God. His treatment of 'life' as a primitive concept is slightly modified by reference to 'organisation' which distributes nourishment (2.27.4), and to 'the motion coming from within' (2.27.5), which makes a life contrast with most machines, but thereafter a reference to the same continued life is sufficient to bestow a unity of a different order from mere collections of particles. We will look at the essentialist explanatory approach to the unity of lives, and Locke's discussion is an important landmark in the debate.

Finally we should note a remark about abstract entities which presents his views on essence in a different light, since he writes that 'the essence of a triangle lies in a very little compass, consists in a very few ideas; three lines including a space make up that essence' (2.32.24). This voices the standard view of his age, that explanatory essences (or even 'substantial forms') are perfectly acceptable in the world of abstracta. It also shows us that Locke was a thoroughgoing essentialist whenever the real essence was directly apprehensible, which Locke took to be the case here in the world of ideas rather than of sense experiences. There were still disagreements about the essences of triangles and circles, but for simple entities there are no secrets. We can take it that the complex properties of triangles are all to be understood in terms of their simple nature, which is self-evident. Presumably, also, we grasp the essence of any individual triangle, as well as of the whole genus. Later empiricists, such as Berkeley and Hume, were adamant that only the individual triangles were to be apprehended, and there was no such thing as a 'general triangle'.

15. Leibniz’s response

Locke famously described himself as an ‘under-labourer’ in the great new scientific endeavour (1694:Epist), which implies that the metaphysics he was discussing is continuous with the physical sciences. Leibniz took the opposite view, and treated metaphysics as a separate realm from physics. Thus he wrote to Arnauld that ‘one must always explain nature along mathematical and mechanical lines, provided one knows that the very principles or laws of mechanics or of force do not depend upon mathematical extension alone but upon certain metaphysical reasons’ (1686b:4/14.7.1686). If we are to treat essences as deeply entwined with the activity of explanation, we must side with Locke on this one. On the whole this makes little difference, since the question of whether (say) the law of gravity is a divine command or an intrinsic emanation of the physical world will be an appropriate question for either approach. It does matter, though, if we are tempted to assert some truth as being bad science but good metaphysics. Pasnau defends the Leibniz approach, and suggests that the metaphysical side of the divide withered while the new scientific side flourished. Leibniz saw that science was flourishing (especially after Newton published his great work), and fought hard to build a metaphysics which could run in parallel with it. In this way he felt able to make the very unfashionable pronouncement that ‘the consideration of forms serves no purpose in the details of physics and must not be used to explain particular phenomena. …but their misuse must not lead us to reject something which is so useful to metaphysics’ (1686a:§10). This division thus enabled him to mount a very interesting critical defence of the Aristotelian views we outlined earlier, and he is one of the most fruitful thinkers on the topic we are studying, and worthy of careful examination. The works of Leibniz are vast and somewhat fragmented, and scholars differ over whether his view changed much over the course of his career, so we will aim to provide a date for each quotation. The significant influences on Leibniz are Aristotle in his earliest studies, then Descartes around 1670 (and the view that matter is just extension), then Newton after 1687 (and the view that nature is governed by sweeping external laws), and Locke around 1710 (and his challenging view that essences could not really survive empirical scrutiny).

Leibniz accepted the Aristotelian framework for explanation (which gave us four modes of explanation, by matter, by causal initiator, by structural form, and by purpose or function), and the centrality of explanation for him is shown in his view that we should accept the Copernican account of the solar system simply because it is the best explanation (1689; Arlew/Garber:92). We have seen that teleological explanation had been firmly dismissed by Descartes and others, but Leibniz saw that this was too quick. In a simple example, he observes that ‘a house would be badly explained if we were to describe only the arrangement of its parts, but not its use’ (1702; Arlew/Garber:255), and he even suggests that final causes had a role to play in physics (1698; Arlew/Garber:157), though physicists seem unlikely to embrace the suggestion. He suggested that the world could be explained just as well entirely by final causes as it can be entirely by efficient causes (1678; Garber:258). The most significant aspect of his approach, though, is his reluctance to accept the new mathematical ‘laws’ as offering adequate explanations. We have seen him praise Newton for making an explanatory connection, rather than discovering an equation, and when he writes (as a critique of Descartes's view of objects as essentially mere extension) that 'even if we grant impenetrability is added to extension, nothing complete is brought about, nothing from which a reason for motion, and especially the laws of motion, can be given' (1704/5; Arlew/Garber:183), it is striking to hear someone of that period demanding that the new laws of motion be explained, rather than merely admired and used.

Leibniz is happy to talk of the 'laws of nature', and he was as committed as Newton to the divine source of those laws. The big difference is that Leibniz put a very Aristotelian spin on the new approach, by insisting that the laws are not impositions of an active divine will (the occasionalist approach), but are created by God as intrinsic to the natures of the entities in the world. In other words, the new laws are to be found in the old essences. His point is that 'it isn't sufficient to say that God has made a general law, for in addition to the decree there has also to be a natural way of carrying it out. It is necessary, that is, that what happens should be explicable in terms of the God-given nature of things' (1698; Woolhouse/Francks:205). Newton had written to Bentley in 1692 'that gravity should be innate, inherent and essential to matter ...is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it' (Newton:102), and yet seven years later Leibniz drafted a letter in which he wrote 'I believe that both gravity and elasticity are in matter only because of the structure of the system and can be explained mechanically or through impulsion' (1699; Arlew/Garber:289). In some way, Leibniz took it that these huge generalities about the behaviour of the universe were written into the intrinsic structures and natures of natural objects. Wiggins quotes Leibniz as writing that 'nothing is permanent in a substance except the law itself which determines the continuous succession of its states and accords within the individual substance with the laws of nature that govern the whole world' (Wiggins 1980:76).

Aristotle understood the process of demonstration as revealing necessities which depended on the essences which were captured in definitions. Leibniz connects this approach to the new physics, and seeks a way to formulate the idea that the natures of things contain 'laws'. The claims that he made for such immanent laws were a little extravagant for modern tastes, since he implies that every event that ever happens to some entity is contained in its essence, but it is probably best to say that his proposal is as close as seventeenth century thought came to fundamental powers and dispositions. Pasnau's view is that thinkers of the earlier seventeenth century never accepted our idea of a 'disposition', despite vocabulary which might suggest it (2011:519). Leibniz, however, took the source of all activity and change to reside within 'substances'. He wrote to Burnett that 'I consider the notion of substance to be one of the keys to the true philosophy' (1703; Arlew/Garber:286), and this is because it is the terminus of natural explanation. The most striking aspect of his account of substances is his lack of interest in them as merely passive supporters of properties and predicates, and such 'substrates' are dismissed as mere metaphors (1710:217). He repeatedly emphasises that substances or essences have to be active, in order to fulfil the role which requires their postulation: 'I maintain that substances (material or immaterial) cannot be conceived in their bare essence devoid of activity; that activity is of the essence of substance in general' (1710:65). Once this has been said, though, what further can be added? Leibniz offers two thoughts. The first is that the character of the intrinsic initiator of activity must be in some way lawlike, and he uses the phrase 'law of the series' for his view, since Leibniz says 'the essence of substance consists in ...the law of the sequence of changes, as in the nature of the series in numbers' (Cover/O'Leary-Hawthorne:220). In terms of Aristotle's original account of essence, this emphasises the abstract guiding principle rather than a causal mechanism.

His second thought is rather more significant for the development of the Aristotelian account of nature, since he connects the intrinsic laws within things to the new concept of 'force', which brings out the causal side of essence. Garber's summary of this idea observes that 'a standard criticism of the scholastic notions of matter and form is that they are obscure and unintelligible; but in Leibniz's system they are connected directly with notions of active and passive force that play an intelligible roles in his physics' (2009:128). The concept of force was controversial in the seventeenth century, and only achieved respectability with the publication of Newton's three laws of motion in 1687, which are couched in terms of forces, expressed as mathematical quantities. In physics it was felt that an explanation was needed not only for active forces such as gravity and magnetism, but also for passive forces such as inertia and impenetrability. Hence Leibniz attributes these two aspects to his fundamental intrinsic forces, and writes that 'the dynamicon or power [*potentia*] in bodies is twofold, passive and active; passive force [*vis*] constitutes matter or mass [*massa*], and active force constitutes entelechy or form' (1702; Arlew/Garber:252). This maps hylomorphism onto the new physics, and means that form or essence is wholly active in character (perhaps suggesting the modern word 'energy'). We can plausibly take this thought of Leibniz's as initiating modern scientific essentialism, although the most fundamental aspects of physical things, which Leibniz is attempting to characterise, almost certainly resist all the enquiries of science. No matter how deep the physicist digs, be it as far as atoms, or protons, or quarks, or fields, or strings, the puzzle of what 'drives' the whole system looks thoroughly elusive. Hence enquiry will always focus on structures and emergent complex powers ('derivative forces' in Leibniz), rather than raw fundamental powers. In that context, we can take Leibniz to have shown that Aristotle should sit at the head of the table.

Leibniz's most famous idea emerges when he looks for the best way to characterise the active forces of essences. He took these active forces to exist because the law of God must 'leave some vestige of him expressed in things' (1698; Arlew/Garber:158), which means we are approaching something divine in studying them, and the search for some analogous phenomenon which generated its own principled activity led straight to one place: 'the clearest idea of active power comes to us from the mind; so active power occurs only in things which are analogous to minds, that is, in entelechies; for strictly matter exhibits only passive power' (1710:172). Thus we arrive at the concept of 'monads'. This idea has been the subject of misunderstanding, caricature and ridicule, and has found almost no support from other thinkers, but we can give monads a serious hearing as long as we read Leibniz attentively and sympathetically. The key word to notice is 'analogous', which occurs in the passage above, and in almost every context where Leibniz discusses monads. If we take minds to roughly constitute consciousness, reason, appetite and sensation, it is only the last two of these which Leibniz attributes to a monad, and monads are never considered to be fully conscious or rational, or to exhibit acts of will. Hence they are not minds, but are analogous to minds. If Newton's second law of motion says that accelerations in the world result from forces, an explanation of these forces needs a fundamental drive which has to be analogous to human appetites. If Newton's third law of motion says that actions cause reactions, the explanation of this needs some detection of the action prior to the reaction, and this must be analogous to human sensations. This doesn't mean that monads are hungry, or are watching you, but that our best hope of understanding the foundations of nature is to attend to such faculties within ourselves. When he writes to Johann Bernoulli that 'I don't say that bodies like flint, which are commonly called inanimate, have perceptions and appetition; rather they have something of that sort in them, like worms are in cheese' (1698; Arlew/Garber:169), the reference to worms probably didn't help his case, but the choice of flint (as inanimate an example as he could think of) is there to show that he is not remotely talking of something full of little minds, but is simply trying to achieve an imaginative grasp of the nature of flint, given that flint is much more active than its inert image might suggest.

Leibniz was a thoroughgoing Aristotelian essentialist. This is worth saying, because in recent discussions he has acquired a rather different reputation, and is referred to as a 'super-essentialist'. Penelope Mackie says that Aristotle 'makes all an individual's properties essential to it' and that this 'should be regarded as an extreme version of essentialism', and she calls this the 'standard view' of Leibniz (2006:1). Wiggins, in contrast, writes that 'Leibniz was not an essentialist' (2001:109). This is puzzling, given how easy it is to find in Leibniz remarks such as the following: 'powers which are not essential to substance, and which include not merely an aptitude but also a certain endeavour, are exactly what are or should be meant by 'real qualities'' (1710:226), which distinguishes between the essential and non-essential powers. Modern writers tend to treat essential features of things as nothing other than features which are necessary for their existence, and part of the interpretative problem resides there, but Leibniz distanced himself from that view when he wrote to Queen Charlotte 'that which is necessary for something does not constitute its essence. Air is necessary for our life, but our life is something other than air' (1702; Arlew/Garber:191). The reason for this misunderstanding of Leibniz seems to arise from what he says about the 'concept' of a thing, so it is worth correcting, because Leibniz is an important figure for modern essentialism.

In 1690 he wrote that 'of the essence of a particular thing is what pertains to it necessarily and perpetually; of the concept of an individual thing on the other hand is what pertains to it contingently or per accidens' (Cover/O'Leary-Hawthorne:127). This is a long way from 'super-essentialism', but it also establishes that a 'concept' of something refers to its contingent features. Another remark from the same year tells us that 'in this complete concept of Peter are contained not only essential or necessary things, …but also existential things, or contingent items included there, because the nature of an individual substance is to have a perfect or complete concept' (Cover/O'Leary-Hawthorne:126). Thus a thing consists of an essence and of accidents; the accidents are referred to as the 'concept', and the combination of essence and accidents is referred to as the 'complete concept'. The complete concept appears to be the ‘super-essence’ of modern discussions (such as Blumenfeld 1982), but it is clearly not the essence. The misunderstanding seems to have arisen because Leibniz believes that it is in principle possible to deduce everything which will happen to something, if only complete knowledge of the thing can be achieved. To Arnauld he expressed this unlikely possibility in this way: ‘apart from those that depend on others, one must only consider together all the basic predicates in order to form the complete concept of Adam adequate to deduce from it everything that is ever to happen to him, as much as is necessary to account for it’ (1686b:48) The error in interpretation appears to arise from taking a knowledge of the complete concept to be required in order to make this prognostication, since if the future of Adam arises from every one of his features, this seems to make them all necessary, if Adam is to be explained. However, this is to ignore the phrase ‘as much as is necessary to account for it’ in this key quotation, because that explicitly tells us that not all of the complete concept is involved. Presumably the unlikely knowledge of Adam’s complete career arises from a knowledge of his essence and of some of his relevant accidents, and hence there is no ground here for equating his complete concept with his essence. It seems clear that Leibniz was not a super-essentialist.

No philosopher (not even Aristotle) cared about the unity of objects as passionately as Leibniz. The motivation for this seems to be expressed in the remark in a letter of 1704 that ‘there is no reality in anything except the reality of unities’ (Garber:363), and yet he cheerfully addresses the reality of disunited aggregates such as a wall made of bricks. Perhaps the real motivation is seen better in the remark to Arnauld that ‘nothing should be taken as certain without foundations’ (1686b:71), which expresses Leibniz’s foundationalist temperament as much as it does an epistemological insight. Aristotle felt that his hylomorphism thoroughly solved the problem of unity, especially for the most obvious case of the guiding form which is in the life of a plant or animal, but Leibniz does not embrace that view, and had a much greater interest in the inanimate (such as flint) than Aristotle. We can take that as symptomatic of his age, when the structure of matter suddenly seemed worthy of study, and Cartesians had downgraded the status of animals. Given that he was fully committed to Aristotelian essences, wherein resided the terminus of natural explanation (even of the new ‘laws of nature’), one might think that the problem of unity was solved for him, since unification is a prime characteristic of the traditional Aristotelian essence. Leibniz understood such essences in terms of the ‘law of the series’ and in terms of ‘forces’, rather as Aristotle had understood them in terms of *eidos*, *morphé*, and *arché.* The problem for Leibniz seemed to be that the ideas of a law and a force are complex, and unity had to be an utter simplicity. Behind his idea of the ‘monad’ seems to be Descartes’ view that a mind cannot have parts, because ‘it is one and the same mind that wills, senses and understands’ (*Med* 6). Apart from the concept of God, the concept of a mind seemed to be the only concept of a perfect unity available. Hence monads must be ‘analogous’ to minds, and yet the complexity of what had to be explained meant that one totally unified monad in each entity was not sufficient. Something like the reasoning that led Democritus to a multitude of atoms seems to have led Leibniz to a multitude of monads.

When Leibniz wasn’t focusing on ‘perfect’ unities, his attitude to unity was thoroughly pessimistic. Thus he writes that ‘without soul or form of some kind, a body would have no being, because no part of it can be designated which does not in turn consist of more parts; thus nothing could be designated in a body which could be called 'this thing', or a unity' (1678; Garber:51). To Arnauld he wrote that 'there are degrees of accidental unity, and an ordered society has more unity than a chaotic mob, and an organic body or a machine has more unity than a society' (1686b:126). It is striking to see 'an organic body' lumped in with a machine, and clearly they only have a fairly high degree of unity, without achieving the real thing. His view is pessimistic because he sees no prospect at all of any unity in the corpuscular approach, as when he writes to Arnauld that 'one will never find a body of which it may be said that it is truly one substance, ...because entities made up by aggregation have only as much reality as exists in the constituent parts. Hence the substance of a body must be indivisible' (1686b:88). In 1712 he offers an account of how we arrive at our normal understanding of everyday unity, given in terms of mechanical connections between parts, and cites 'duration', 'position', 'interaction', and direct 'connection' as the features which legitimate truths about entities, despite our ignorance of the monads (1712; Arlew/Garber:199). This is the best account he can find of unity, in the absence of some absolute underlying unifier, but the things still only 'seem' to be one.

For Aristotle, there was not only the unification imposed on matter by form, but there was also the unity expressed by an 'account' [*logos*] which was a successful definition. Leibniz, though, is not so confident that definitions can fulfil this role. We suggested that Aristotle aspired to an ideal of defining individuals, even if the use of universals seems to stand in the way, but Leibniz is quite firm that individual definitions are beyond us (1710:289). The difficulty with achieving a unique definition is that it requires a 'perfect idea' of the definiendum, whereas the best we can usually manage is a 'distinct' idea, as when we know enough to pick gold out, but never enough to understand it (1710:267). The consequence for definition of having an imperfect idea of it is that 'the same subject admits of several mutually independent definitions: we shall sometimes be unable to derive one from another, or see in advance that they must belong to a single subject' (1710:267). So far, Aristotle would recognise this difficulty, but for Leibniz 'although a thing has only one essence, this can be expressed by several definitions' (1710:294). An Aristotelian essence was held to correspond to a unique definition, and arrival at the unique definition was the hallmark of an essence. This sceptical thought of Leibniz's threatens to break the tie between essence and definition, so this is another problem which needs to be addressed.

Without a theory of monads, Leibniz is what we now call a 'nihilist' about unified objects, the view expounded by Unger (1979), and it might be best to understand his account in terms of the physics/metaphysics divide, with only apparent unity possible in the world of physics, and monads providing true foundational unity for the metaphysics. It is unfortunate, then, that the theory of monads has little appeal to modern thinkers, unless they are sympathetic to idealism or to panpsychism, and it is revealing that Leibniz himself betrayed unease about the success of his metaphysical theory, late in life, in letters responding to challenges from Des Bosses. In those letters he tells us (to our considerable surprise) that 'monads do not constitute a complete composite substance, since they make up, not something one per se, but only a mere aggregate, unless some substantial chain is added' (1712; Arlew/Garber:201). The discussion is too complex to analyse here, but Leibniz first considers his concept of a ‘dominant’ monad to unify creatures and persons, but seems to prefer the ‘substantial chain’ as unifier, and in the last year of the correspondence he writes that ‘the realising thing must bring it about that composite substance contains something substantial besides monads, otherwise …composites will be mere phenomena. In this I think I am absolutely of the same opinion as the scholastics, and, in fact, I think that their primary matter and substantial form, namely the primitive active and passive power of the composite, and the complete thing resulting from these, are really that substantial bond that I am urging’ (1716; Garber:379). Since the scholastic powers have been closely linked by Leibniz to the forces of modern physics, it appears that the need for a substantial chain is leading him back from the metaphysics to the physics, to solve the problem of unity. Garber argues that, contrary to the views of many scholars, Leibniz was in a state of constant metaphysical development, and so the issue is left unresolved at his death. It seems fairly obvious that monads cannot offer a solution to the unity problem, despite their supposed perfect unities, because they occur in multitudes, which then require further unification. Leibniz explored the problem of explaining unity to a remarkable depth, but if there is a solution to this supposed ‘problem’, he does not seem to have found the answer.

With respect to modality, there is a tension in Leibniz’s thought which is worth noting for the present enquiry. Leibniz is the embodiment of the rationalist approach to such things, which is that necessary truths are identical to those which are known a priori, and he writes that ‘the fundamental proof of necessary truths comes from the understanding alone, and other truths come from experience or from observations of the senses. Our mind is capable of knowing truths of both sorts, but it is the source of the former’ (1710:80), in which it is noteworthy that he explicitly gives the understanding as the ‘source’ of at least the proof of necessary truths. This is not, of course, to deny objectivity to such truths, but necessities seem to be identical to truths apprehended by reason (which invites the question of their status if no understandings existed, which Leibniz would consider impossible). The tension is with a number of remarks which connect essences with modal facts, as when he says that ‘essence is fundamentally nothing but the possibility of the thing under consideration; something which is thought possible is expressed by a definition’ (1710:293), and that ‘one mark of a perfect idea is that it shows conclusively that the object is possible’ (1710:268). In the account of his work given by Cover and O’Leary-Hawthorne, they conclude that ‘in Leibniz's view, the essence of a thing is fundamentally the real possibilities of that thing’ (1999:169), which constitutes what we would now call its ‘modal profile’. The essentialist approach offers a possible account of the sources of possibility and necessity, an approach which has been explored recently, starting with work by Fine (1994). It is also worth noting that Leibniz does not equate (in the Humean manner) the possible with the conceivable, and he writes that ‘it does not follow that what we can't imagine does not exist’ (1698; Arlew/Garber:168).

In his occasional remarks on the essences of abstracta, Leibniz endorses the standard view (for the time) that we found in Locke – that the essences of simple geometrical figures are self-evident, and explain the complex truths that are derived from them. For him ‘the essence of a circle consists in the equality of all lines drawn from its centre to its circumference’ (1669;

Cover/O'Leary-Hawthorne:24), a view which Spinoza rejected, writing in *Improvement of the Understanding* that 'no one fails to see that such a definition does not at all explain the essence of circle, but only a property of it', adding that 'the properties of a thing are not understood so long as their essences are not known'. Spinoza thinks that the equality of the radii can be explained by something that goes deeper, and defines the circle as 'the figure described by any line whereof one end is fixed and the other free; this definition clearly comprehends the proximate cause' (p.35). The equality of the radii are to be explained by the fixity of the moving line. Whoever is right here, we should observe that the giants of late seventeenth century philosophy had no trouble in debating the essences of abstract entities, well after 'substantial forms' had fallen from view in physics. The explanatory role of essence is also particularly clear in such examples. A further symptom of Leibniz's foundationalism (and perhaps essentialism) about abstracta is his interest in the axioms of Euclid, of which he says not only that 'to reduce the number of axioms is always something gained' (1710:407), but also that 'I want to see an attempt to demonstrate even Euclid's axioms, as some of the ancients tried to do' (1710:101). The objective of his metaphysical enterprise was to push every area of study back until you hit bedrock. Essence, we are claiming, is precisely this bedrock of enquiry (though this is unlikely to take us right to the 'foundation' of the enquiry).

Finally, we should note an aspect of Leibniz's thought which distances him from Aristotle and brings him much closer to our own times, and that is his emphasis on the considerable degree of subjectivity, convention and pragmatism in our modes of understanding. For Aristotle we have seen that the starting point is to achieve understanding, and the possibilities of explanation are limited by our cognitive mental faculties, but the four modes of explanation still aim to pick out objective features of the external world. The thought of Leibniz often places him historically as midway between Descartes' shift of focus to the mind of the thinker, and Kant's dramatic proposal that the mind imposes a framework on our entire scheme of understanding. That (in the absence of monads) unity would consist entirely of mere 'phenomena' is a symptom of this un-Aristotelian approach, and in 1710 Leibniz went further, writing that 'fluidity is the fundamental condition, and the division into bodies is carried out - there being no obstacle to it - according to our need' (1710:151). The theory that there are underlying monads is not a mere division 'according to our need', and presumably the late commitment to the 'substantial chain' is intended to state an objective truth, but these proposals seem to be driven by the recognition that without them our metaphysics of nature collapses into subjective anarchy.

The philosophy of Leibniz seems to be driven as much by a very human need to understand (which was how we understood Aristotle), as it is by a mere objective attempt at describing how things are. Thus when he asserts the importance of 'force', he writes that 'I believe that our thought is completed and terminated more in the notion of the dynamic [i.e. force] than in that of extension' (1706; Garber:164), where he is clearly asserting the reality of forces, but the aim is to 'complete' and 'terminate' our thinking. In the remark quoted above asserting the importance of 'activity', he not only said that activity is essential to substance, but also said that substances, whether material or immaterial, 'cannot be conceived in their bare essence without any activity' (1710:65), which not only asserts the reality of activity as essential, but also asserts the conditions which are imposed on our grasp of essence by our conceptual abilities. When he talks of our concept of pluralities and unities, he writes that ‘a plurality of things can neither be understood nor can exist unless one first understands the thing that is one, that to which the multitude necessarily reduces’ (1690; Arlew/Garber:103), which tells us that unity comes from the nature of our intellects, as much as from reality. A very general remark confirms this impression, when he writes that ‘one should choose the more intelligible hypothesis, and the truth is nothing but its intelligibility’ (1689; Arlew/Garber:91), where the actual equation of truth with intelligibility expresses this aspect of his metaphysics much more boldly than the rest of his writings seem to indicate. There is sufficient here to see the general attitude in Leibniz’s thought which is sympathetic to the current discussion – that our grasp of essences, and the contribution they make to our metaphysical framework, arises at least in part from our own minds as from an objective description of nature. He is not, of course, espousing cultural, conceptual or linguistic relativism about such things, but his view is that the dictates of our reason compel us to see the world in certain ways, which is a significant attitude distinguishing his essentialism from that of Aristotle.

THREE

Explanation for Essentialists

16. Four examples

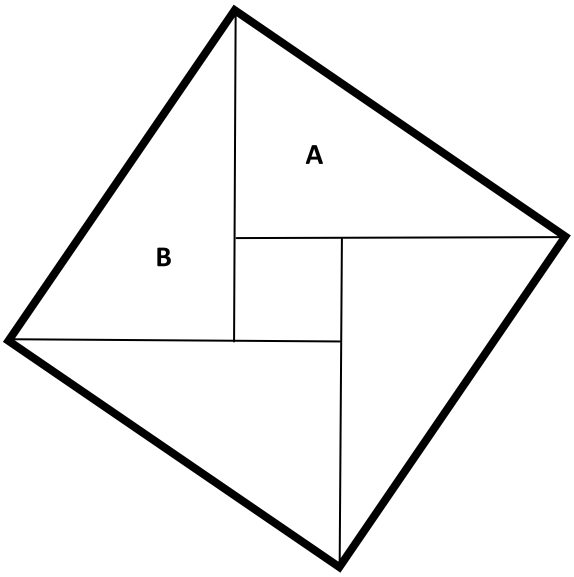
The concept of essence may only exist because we seek explanations, and we must now approach such a thesis from the direction of explanation. We will start with four examples, each illustrating an aspect of the explanatory issues that concern us.

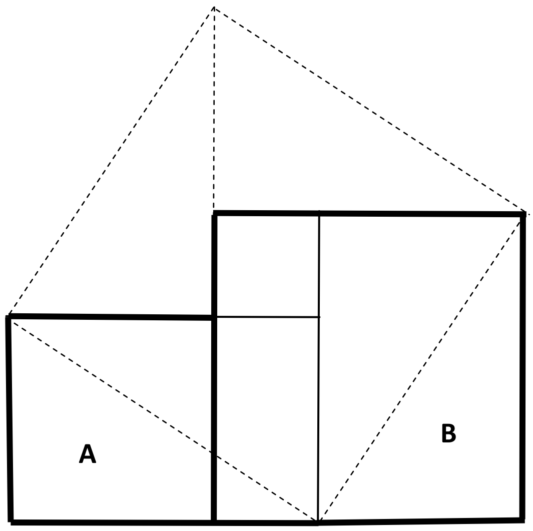
The vast majority of **lung cancer** is caused, it seems, by the smoking of cigarettes. The fact of the causation was established in 1950 by Doll and Bradford Hill, by means of statistical epidemiology, giving an undeniable correlation. However, the correlation did not explain the link between the cigarettes and the cancer. We now know that cigarettes contain a large number of carcinogens, and research seems to be closing in on the correct explanation. As one example, a link has been found between benzo[a]pyrene in cigarettes and gene p53 in lung-wall cells (*trdrp.org*). The story is complex, but presumably a series of links will emerge from research.

A **tsunami** is a huge wave which hits the seashore, often causing extensive damage and loss of life. The explanation of the phenomenon was, we may presume, originally shrouded in superstition and guesswork. Modern explanations pointed to the geology, but this still left unexplained the fact that there were regions where tsunamis did and did not occur. The theory of plate tectonics now gives us the explanation we wanted, because plate boundaries have been observed, and correlated with earthquake and tsunami occurrences. The recent tsunami in north-east Japan was the result of the Pacific plate thrusting under the extremity of the North American plate.

Newton's famous **gravity** equation ($F = m_1.m_2 / d^2$) used the concepts of force, mass and distance to express a universal truth, which concerns all magnitudes, from the vast to the tiny, for all three concepts. When we add that all objects possess mass, and that the mass generally behaves as if it were concentrated at a single point (the 'centre' of gravity), the equation seems to offer powerful predictions as well as universal scope. As Leibniz observed, it shows that a single force can explain all planetary orbits, and it achieves this with precision.

The fourth example is an explanation of **Pythagoras's Theorem**. In Euclid this is given an algebraic proof, suggestion that the explanation is connected to his foundational axioms. Since then a diagrammatic demonstration of the Theorem has emerged, which may reveal the original insight which established its truth. An online animation of the proof shows its workings ('Pythagoras-2a.gif', created by 'Alvesgaspar'), but we can see it in a pair of diagrams:





The left diagram shows the 'square on the hypotenuse' of a right-angled triangle, composed of four versions of the triangle plus a small central square. The second diagram is a rearrangement of the first, with triangle A moved to the bottom left, and triangle B moved to the bottom right. Hence the two thick-bordered squares are also composed of four versions of the triangle plus the small central square. But these are the squares on the other two sides of the triangle, so the truth of the Theorem is obvious. The famous Theorem has been explained, by revealing that it is just a rearrangement of shapes.

The lung cancer explanation reveals a regularity implying a frequent connection; the tsunami example gives the underlying mechanism; the gravity case gives a precise and universal connection; and in the Pythagoras case what was a puzzle has now become obvious.

17. What needs explaining?

With a few examples to get us started, the next consideration is the preconditions which an explanation appears to require. In a world with no conscious minds, a description of the intrinsic features of such a world does not need to mention explanations. If there existed one mind, but it was omniscient, it would still not seem to require explanations, since even the most complex and remote connections and causes would be self-evident. So explanations are only required by minds which are puzzled. In the case of the tsunami, for example, the interaction of plates is overwhelmingly the best explanation, but only for creatures living on dry land. If a fish saw the two plates moving then that part would be self-evident, and a curious fish would want to know why plates move. In each instance the explanation is 'real' enough, but the real features picked out have to be relative to a mind in a state of puzzlement. Putnam observes that aliens visiting Earth might explain forest fires mainly by the presence of oxygen (1981:214). There have been several proposals for the general nature of puzzles which demand explanation. Harré and Madden say that 'only changes require explanation' (1975:163), but that doesn't seem right, since a lack of change might puzzle us and demand explanation (such as the Moon only ever showing one side to us). Robinson says that an explanation presupposes 'something which is improbable unless explained' (2001:216), but that doesn't seem right either, since in Japan tsunamis were always horribly probable, but still cried out for explanation. We explain when we desire to understand, and Strevens may even be right that explanation is a necessary condition for scientific understanding (2011). We certainly desire to understand when we are puzzled, and explaining a puzzle is more than solving a puzzle (such as a sudoku). In the case of the tsunami, the big step towards understanding was to focus the puzzle on the geographical locations of tsunamis, rather than on their mere existence. Puzzlement leads to understanding, but the drive for understanding generates each particular puzzle. Lewis says that the fact that a good explanation aims at understanding 'adds nothing to our understanding of explanation' (1986c:228), but it is clearly informative to know what an activity aims at. It certainly tells us that subjective, pragmatic and contextual factors are inescapable in any decent account of explanation, since puzzles and understanding will be relative to the nature and circumstances of the enquirers.

18. Stable structures

If we label something as 'chaos', we mean that there can be no explanation of what happens in the chaos (though we might explain what caused the chaos); hence explanations require some sort of order. As Ruben correctly observes, 'objects or events in the world must really stand in some appropriate 'structural' relation before explanation is possible' (1990:210). Ladyman and Ross write that 'philosophers sometimes invoke natural kinds as if they explain the possibility of explanation. This is characteristically neo-scholastic. That anything can be explained, and that properties cluster together, express one fact: reality is relatively stable' (2007:292). This seems wrong, since there are two facts involved, not one; reality is relatively stable, and it also has detectable structure. If reality were stable but homogeneous, there would still be no possibility of explanations. If reality were replete with structures, but the structures were hopelessly unstable, that too would prohibit explanations. There are no explanations if there is nothing consistent to pick out, and that implies clustering, which implies something like natural kinds. Picking out might also imply unities, to which we will return. Our explanation of lung cancer needs stability among the constituents to produce the statistical regularity, and molecular structure to lead us to the underlying mechanisms.

19. Foundations

Given that a theory of explanation must be committed to a reasonably stable and structured reality, an attractive thought is that successful explanation reveals the foundations of the structure, so that the puzzling phenomenon is merely the expression of basic entities which are clearly understood. Boyle wrote that 'explications be most satisfactory that show how the effect is produced by the more primitive affects of matter …. but are not to be despised that deduce them from more familiar qualities' (1672; Pasnau:530). However, whether reality and matter possess foundations seems to be a very open question, which is unlikely to be decided by philosophers. Ladyman and Ross are prepared to venture that 'there is no fundamental level (and) the real patterns criterion of reality is the last word in ontology' (2007:178), but few others would be so bold. Mumford offers the thought that structure may run out before the lowest level is reached, which would mean that the explanations also necessarily run out, before we reached the bottom (1998:133). For our purposes, we should probably say that even if the foundations were established, and known, we still comprehend explanations which move ‘sideways’ (or even ‘up’) in the hierarchy, and that explanations can be thoroughly satisfactory without making reference to anything foundational. If quarks and leptons are the foundation, a catalogue of which of them composed a hedgehog would hardly explain the hedgehog, and the explanations of tsunamis and lung cancer don’t appear to involve particle physics.

20. Levels

If we neither know of nor need the foundations of reality for explanation, the concept of there being ‘levels’ may still be required. The idea that there are levels of degree in reality (with some things being less real than others), is not fashionable nowadays, although Plato expounded that view in *Republic*, suggesting that to leave the Cave is to approach what is more real (515d), and scholastic philosophers often speak of more than one mode of existence. Heil has argued persuasively that ‘we should accept levels of organisation, levels of complexity, levels of description, and levels of explanation, but not levels of reality’ (2003:10). In the hierarchy of science that rises from physics to chemistry to biology and beyond, the metaphor of ‘levels’ seems inescapable. It makes sense to us that there could be a world with physics but no chemistry, and chemistry but no biology, but biology without chemistry and physics seems inconceivable to the modern mind. In that sense, there is a self-evident dependence relation. Although reality is a fairly seamless whole, distinguished more by its continuities than by its ‘joints’, it is a key aspect of our vision of reality as having ‘levels’ that these are not merely arbitrary lines drawn across a vertical continuum, but that there seem to be real jumps from one level to another. Although explanations will be possible in a reality without levels, provided that it contains suitable structural components, it is in the explanation of a layered reality that Aristotelian essences seems most appropriate.

The denial of such levels is a brand of anti-realism about the way we understand nature, and a lot more than essences would then have to be left out of the account. Alternative views of the structure of reality might be defended, such as gradations from simplicity to complexity, but while that can offer discernible structure, it does not offer clearly demarcated levels, since mere variations in complexity will be too fine-grained. Schaffer recognises, rightly, that the ‘grounding’ relation is central to there being an explicable order in reality, but his rather formal proposal that ‘by treating grounding as transitive (and irreflexive), one generates a strict partial ordering that induces metaphysical structure’ (2012:122) will offer some sort of territory in which explanations can operate, but offers no natural landmarks on which to pin them. The clear and significant steps that can give us the levels we need in the structure can only arise from empirical observation (of where physics steps up into chemistry, for example).

If we consider the levels of physics and chemistry and biology, the feature that generates the self-evident boundaries is our perception of modularity in the system. By a ‘module’ is meant a standardised and replicated component, such as bricks or tiles used in house-building. If we wished to understand a roof, we would first study the tiles, and then study their structure. The uniformity of the tiles means that understanding of one tile will flood through the whole collection (and as we will see, natural kinds have such a role in the sciences), but the study of the whole roof must then take us to a different 'level'. Further levels will be revealed if we examine the minerals which compose the roof tiles, or the villages which are composed by the houses. Exactly this modularity is exhibited in the standard model of physics (where fundamental particles are the modules for the atoms), and in chemistry (where the elements are the modules for the molecules), and in biology (where molecules such as proteins are the modules for cells). Such modularity is found in numerous other areas of study (such as a language, with its morphemes, phonemes, words and grammatical markers). It is hard to think of any subject which does not either exhibit natural modular structure, or consist of modules that we have created (because our minds are in tune with nature). Hence the landscape of explanation seems to be irredeemably layered in its structure. Whether such layers are universal and uniform or piecemeal and disjointed is a problem facing enquirers, rather than a presupposition.

21. Direction

If there are levels in the structure of reality, and dependence relations between the levels, then there is a 'direction' in that dependence, seen in the priority of chemistry over biology, and in any modularity relationship. Direction is the most important feature of reality for explanation. A classic example is Bromberger's flagpole, which is used to challenge any notion of explanation being a purely deductive affair. A flagpole, it is said, explains its shadow, but the shadow does not explain the flagpole, and this is because there is a direction to explanations, presumably commonly connected to the direction of time's arrow and of causation (and maybe directionality is even the primitive intuition about nature that makes us embrace those two contentious concepts). One could plausibly respond to the example that if I built a wall to shade myself from the sun, the shadow would explain the wall at least as much as the wall explained the shadow, but the two explanations are not in conflict. If the context is an explanation of the existence of the shadow, then the explanation goes from wall to shadow; if the context is the explanation of my building the wall, the explanation goes from shadow to wall. Each explanation has its own single direction. It does not matter whether this directionality that we presume in reality rests on time, on causation, on dependence, or on some Kantian category of thought. Any perception of direction allows the explanatory instincts to get a purchase, and in a non-directional environment the prospects for explanation seem to dwindle.

22. Ontology

To explain things, one needs an appropriate ontology. If, for example, one's ontology included aether, ghosts and non-existent objects, one's explanations would probably veer off into the peculiar. In modern ontologies, the most austere ontology involves just objects (to which Quine adds sets), and a more liberal ontology might add properties and relations. More lavish ontologies might add minds, persons, and universals. Entirely different ontologies can be built from processes, from powers, from events, from tropes, or from sense-data. Any explanation will draw on its background ontology. Rather than demand an ontology prior to embarking on explanations, the more accurate picture suggests that the merits of the emerging explanations are the best test of the virtues of a system of ontology. We should certainly recognise that explanation plays a vital role in metaphysics, and is not just a minor branch of epistemology. As Lipton puts it, 'one of the points of our obsessive search for explanations is that this is a peculiarly effective way of discovering the structure of the world' (2004:66). An even stronger claim from Ladyman and Ross is that 'we reject any grounds other than explanatory and predictive utility for admitting something into our ontology' (2007:179), which is an interesting proposal to entirely reverse the normal account. Some sort of ontology may be a precondition for explanation, but might successful explanations be the precondition for commitment to an ontology?

23. Induction

There is a close relation between the procedures of induction and of explanation, and Harman roughly identified the two, writing that 'we might think of enumerative induction as inference to the best explanation, taking the generalization to explain its instances' (1995:34), and that 'induction is an attempt to increase the explanatory coherence of our view, making it more complete, less ad hoc, more plausible' (1973:15). We have already rejected the idea that a generalisation can directly explain its instances (since 'all tigers are ferocious' offers no explanation of their ferocity), and the direction of explanation here should obviously be from the instances to the generality. Various philosophers have offered to define induction as moving from particular to universal, generalising about phenomena, finding laws to fit experience, animal expectations, and moving from observed to unobserved. Three suggested presuppositions for induction are that events are linked by causation, that nature has hidden necessities, and that the future will be like the past. We will assume the best starting point to be that induction is just learning from experience (which was Hume's view), since there seems to be no discontinuity between the inductive practices of humans and of other animals.

24. Typicality

So what counts as 'learning from experience'? We can learn from a single experience, if we suspect that the single experience is in some way 'typical' of the situation (if, say, I find a new species of creature, and the first one bites me). Clearly, though, deduction from one experience is suspect, if there is considerable variety in this corner of nature, but we often discover that our inferences about a first instance tell us everything about similar items (electrons, perhaps). If we demand many instances in confirmation, we open ourselves to well-known objections, such as the remark of Sextus quoted earlier, that many observations may still miss out the crucial disconfirmation, or that if we propose that 'all natural numbers are less than 1,000,001' we can offer a million confirmations. Our reliance on experience seems to greatly depend on what is deemed 'typical'. Sextus's problem case (of missing the crucial disconfirmation) should be extremely rare, since we don't merely 'enumerate' the instances we examine, but we also try to extrapolate to the unobserved by assessing the intrinsic character of what has been observed, drawing on a great deal of prior knowledge. The million numbers case is absurd because we can extrapolate the whole picture right from the start, without bothering to observe. If we study tsunamis or cigarette carcinogens, we pick typical cases to study. In proving Pythagoras's Theorem we know that all right-angled triangles are typical in the relevant respects.

25. Coherence

Learning from experience is not a narrow process of cataloguing, but the broadest possible process of fitting new observations into our current web of belief, for which the best word seems to be 'coherence'. Smart, in a nice defence of the coherentist approach, commits to the view that explanation simply *is* coherence, writing that 'I want to characterise explanation of some fact as a matter of fitting belief in this fact into a system of beliefs' (1990:2), but this seems too strong (inviting the initial objection that a large array of fictions or falsehoods can be coherent).

Critics of coherentism say the concept is too vague, but Thagard identifies plausible ingredients of a coherent theory which give the whole approach more substance (2000). A degree of rigour is brought to bear on coherence, not by the usual recourse to formal logic or to probability theory, but by attempting to find coherence algorithms for artificial intelligence. In brief, we break a puzzle down into sets of positive and negative 'restraints', and then design a maximum satisfaction for these restraints. There are then five main kinds of epistemic coherence: explanatory, deductive, conceptual, analogical and perceptual, with scientific theories centring on explanatory coherence (2012:43). Informal content is given to the concept of 'constraint satisfaction' by offering seven principles: symmetry (that coherence is mutual), explanation (that good economic explanation *is* coherence), analogy (similar explanations of similar puzzles cohere), data priority (good descriptions should be accepted), non-contradiction, competition (rival explanations are incoherent with one another), and acceptance (an acceptable proposition coheres with other propositions) (2012:44).

This offers some clarification of the aim of coherence, but neither Smart's nor Thagard's coherentist schemes offer explanations on their own. Any coherent explanation needs to be relevant and non-trivial, and we would expect causal and determinative factors to play a prominent role. That such factors achieve coherence can be taken as the prime meta-criterion of explanation, and broad coherence must be one of the best criteria of success in explanation, since the coherence suggests that it is true. We can agree with Thagard that 'a scientific theory is progressively approximating the truth if it increases its explanatory coherence by broadening to more phenomena, and deepening by investigating layers of mechanisms' (2012:46).

26. Induction problems

If sensible induction aims at coherence, the next question is what preconditions are required for such an activity. Most commentators (such as Lipton 2004) take Hume's main problem of induction to have no solution – where that problem is either using induction to demonstrate that induction is reliable, or trying to demonstrate that the future will be like the past. The first challenge seems unreasonable, since we are familiar with the problems of circular confirmation in logical attempts to validate deductive logic, and it is against common sense to abandon trying to learn from experience. Attempting to confirm that the future will be like the past seems to be the really intractable problem, since plenty of entertaining examples show that anyone ignorant of the bigger picture can go badly wrong in induction. If we agree that (for all we know) the entire universe may terminate ten seconds from now, then we have no certainties about the future. This is, however, also no reason to abandon learning from experience. The best strategy seems to be to attach a 'so far so good' footnote to the whole procedure, which places induction closer to the realm of pragmatics than to logic. We assume that the future will indeed be like the past, but that does not tell us the *respects* in which it will be similar. Other well-known problems of induction, such as how to formulate the predicates employed, or how to handle the logical implications of a hypothesis, we will leave to specialists.

27. Necessities

Aristotle took the aim of demonstration to be the revelation of necessities, but we saw that this required necessities in both premisses of his syllogism, in order to achieve the required necessary conclusion. There may exist absolute necessities in reality, and there may exist a priori means by which rational minds can grasp them, but we will adopt the Aristotelian approach to necessities, which prefers to start from a 'necessity-maker', rather than a necessity 'insight'. If an explanation is only fully successful if it reveals a necessity, then the existence of necessity-makers will be a further precondition for the practices of explanation. In syllogistic demonstration we seemed to discover logical necessity, but the necessities of implication are evident even if one of the premisses is contingent, or even false. We might reasonably presuppose the existence of logical necessity, and we may reveal such necessity in order to explain logical inference, even if the foundations of logic are controversial. That would explain the transmission of necessity to the conclusion of a demonstration, but it would not explain necessity in the either of the premisses (such as the possible necessities that Socrates is a man, or that all men are mortal). These would require other modes of necessity, dependent on varieties of necessity-maker. Fine has pioneered modern studies of this approach to modality. He writes that 'I am inclined to the view that ....each basic modality should be associated with its 'own' explanatory relation' (2012:40), and offers a taxonomy in which 'the three sources of necessity - the identity of things, the natural order, and the normative order - have their own peculiar forms of necessity. The three main areas of human enquiry - metaphysics, science and ethics - each has its own necessity' (2002:260). Following this proposal, we may surmise that there are necessities residing in every area of human enquiry, and a clear grasp of such necessities might constitute the acme of explanation, but it doesn't follow that all explanations, even the very good ones, must reveal necessity. If there is contingency in chance events, there seems no reason why such events should not wholly explain some puzzle, such as a chance meeting with my debtor explaining why I came home somewhat richer. If we spot a regular causal link in nature, or a revealing mechanism, these may give excellent explanations, despite our uncertainty about whether what we have found is 'natural necessary', or merely a universal fact.

28. Preconditions summary

The detour into induction was worth making, because it brought the notions of 'coherence' and of 'typicality' to the fore, and these must play an important role in our account of explanation.

So far our list of preconditions for explanation to be a worthwhile activity tells us that reality must consist of a stable stratified structure (presumably across both space and time), that we sense a direction of dependence within this structure, and that we have a plausible underlying ontology to draw on for the explanation's building blocks. We must take an interest in possible sources for all types of necessity, even if it is unclear how deep the necessitation runs. We can also add that we must be committed to coherence in our own belief systems, and that we have some implicit grasp of what that means. Useful criteria for coherence can be articulated, and this may be the most important task facing specialists in epistemology.

The question of what makes something 'typical', which is the key to a view of induction that goes beyond mere statistical enumeration of instances, remains central and problematic for the schema in which good explanations can occur. For now, we can start with the question 'typical of *what*?', to which we can simply say that 'clusters of features' (rather than natural kinds) register in our experience, and that something at the centre of the cluster begins to look 'typical'. A typical penguin is a fairly distinct concept, but a typical bird is much more elusive, and yet not entirely meaningless. The dissimilarities among birds inclines us to subdivision, the aim of which is to converge on a narrower and clearer criterion of typicality. Quine asks excellent sceptical questions about the relation between similarity and natural kinds (1969), so we will return to this matter later in the discussion.

With an account in place of the necessary preconditions for an explanation, we can attempt to say what an explanation might consist of, with a declared interest of seeing whether essences are indispensable to such an account. One option is that there is no answer here, because the activity of explaining is too diverse. That is too pessimistic, but we may have to accept several distinct and unconnected modes of explanation; for example, we might say that purely physical events are explained in one mode, human activities in another, and abstract phenomena in a third (perhaps by citing causes for the first, desires and beliefs for the second, and structures for the third). Explaining how or why Richard III became king certainly seems very different from explaining why tsunamis occur, or from why Pythagoras's Theorem is true, but we may be able to subsume them within a single account. We can revert to the pluralist view if that fails.

29. Prediction and accommodation

One interesting disagreement that runs through the literature concerns the question of whether explanations are more successful if they accommodate known data, or if they make successful predictions. Both would be nice, of course, but which one gives the stronger explanations? Lipton defends predictions, giving as an example that 'we are more impressed by the fact that the special theory of relativity was used to predict the shift in the perihelion of Mercury than we would have been if we knew that the theory was constructed in order to account for that effect' (2004:172), and he also claims that accommodations, unlike predictions, can be suspected of fudging the results (2004:184). We are discussing correct explanations here, so we may presume that a successful accommodation has to result also in correct predictions (should the situation recur, which it may not), and presumably a fudged explanation might thus be revealed as incorrect. The key to the matter is that explanations lead to increased or successful understanding, and we then see that successful prediction may not achieve this. As Salmon puts it, ‘various kinds of correlations exist that provide excellent bases for prediction, but because no suitable causal relations exist (or are known), these correlations do not furnish explanation’ (1989:49). Clearly, I might predict that something was about to happen, from sheer regularity in past observations, and yet I may be utterly ignorant of the explanation. Accommodation to the data may be subject to the same criticism – that we might find a rule or an equation that fitted some regularity, without knowing why it fitted – but accommodations have to pay attention to the complete data set, where a prediction may derive from a few astute selections among the data. Scerri, as we will see, argues that accommodations were more important than predictions in the development of the periodic table (2007:124). We may conclude that accommodations have priority, but take the lesson that understanding may require more – that some ‘depth’ is needed in the account, rather than attention to surface pattern.

30. Covering laws

The principal modern theories of explanation are law-based, reductive, causal, or mechanistic. Less satisfactory proposals include denial of all explanations (by instrumentalists), entirely pragmatic and subjective accounts (which make any nonsense a good explanation if it ‘fobs off’ the enquirers), and probabilistic accounts (which encounter high probabilities which don’t explain, and low probabilities which do). If there is a unified account of explanation to be had, it will probably include ingredients of the four main theories, but there are real conflicts between them.

The ‘covering law’ approach to explanation usually arises from a background commitment to empiricism. The nub of the theory was given by Comte: ‘In positivism the explanation of facts consists only in the connection established between different particular phenomena and some general facts’ (1830:2). The modern covering law theory of explanation (associated particularly with the work of Hempel) proposes (in its main ‘deductive-nomological’ form) that for any event we specify the event and the initial conditions of its occurrence, and explanation then consists of identifying the law under which the event falls, and from which the occurrence of the event could be logically deduced. The attraction of the theory is that it gives a coherence to our grasp of the event, by slotting it into a conceptual scheme which embodies our best understanding of the world. If the overview of science is our best explanatory framework, then connection to that framework might be the only satisfactory mode of explanation.

The empirical background to the theory is what we now refer to as ‘Humean’. If, in Lewis’s terms, the basis of all our accounts of reality is a ‘mosaic’ of simple ‘qualities’ (or Hume’s ‘impressions’), and the laws are our best account of the patterns that emerge from the mosaic, then placing some given event within the pattern by connecting it to some law seems to throw the required light on the background requirements, causes and consequences of the explanandum, which is taken to be an event. Hence doubts about the covering law approach often lead to doubts about the ‘Mill-Ramsey-Lewis’ account of laws (which are just an optimum axiomatisation of the truths about the qualitative mosaic), and these doubts are precisely what drive a desire for some explanation of the ‘mosaic’, rather than a mere description of it. It is not a coincidence that Comte himself was pessimistic about such explanations, when he wrote that ‘in the positive state, the human mind, recognizing the impossibility of obtaining absolute truth, gives up the search for hidden and final causes’ (1830:2), given that if laws merely describe patterns of regularity then they seem deficient in the explanatory power that is needed.

Once the covering-law model of scientific explanation had been formalised by Hempel, objections began to emerge (see Armstrong 1983 for a sustained attack, and Salmon 1989 for an overview). A principal target is the claim that there is a logical deduction involved (of the event, from the background conditions and the covering law). Since logical connections (even implication) are held to be timeless, they do not have a causal direction, which gave rise to the famous example of the flagpole. The problem is that given the shadow and certain geometrical laws, we can infer the height of the flagpole, but it seems a gross error (in most contexts) to think that the shadow explains the pole, when it is clearly the other way around. Similar well-known examples observe that we can infer but not explain storms from barometers, and infer but not explain why a man fails to become pregnant when he takes female birth control pills. All of these examples show that much more is needed to generate an explanation, and the concepts of relevance and causation (the latter discarded by Hempel) are inescapable.

A general objection to the whole programme of lawlike explanation arose in the discussion of Aristotle, in the observation that generalisations must depend on particulars, so there can be no meaningful dependence in the other direction. If particular events cannot depend on laws, then laws are incapable of explaining them, since we have concluded that the direction of dependence is foundational to the process of explanation. Armstrong sees mere circularity in lawlike explanation when he writes that ‘given the Regularity theory, the explanatory element seems to vanish. For to say that all the observed Fs are Gs because all the Fs are Gs involves explaining the observations in terms of themselves’ (1983:102), and Mumford agrees when he writes that ‘laws, qua true generalities, if they exist at all, are ontologically parasitic upon the capacities of particulars, rather than the other way round’ (1998:230). If there are such things as laws, they must be explained by particulars, so they can’t do the job required of them. In his subsequent book, Mumford argued that the laws do not exist, which would certainly terminate this account of explanation (2004).

Nancy Cartwright has also offered an interesting attack on the role of laws in explanations (1983). The key thought is that we are trying to explain the real world, but laws are idealisations, so there is a mismatch. The simple laws of physics are discovered in controlled conditions experiments which filter out the mess of the world, and then they are further purified by smoothing out small discrepancies in laboratory readings, thus arriving at a neat law. In the meantime the world is a tangle of intersecting laws, and every event occurs at a confluence of these generalities. Thus she concludes that ‘when different kinds of causes compose, we want to explain what happens in the intersection of different domains. But the laws we use are designed only to tell truly what happens in each domain separately’ (1983:12). Her headline claim is that ‘the laws of physics lie’, but that seems not to be true if they are actually correct when the situation approaches that of the ideal experiment, though it is plausible to say that ‘the laws of physics don't explain', if nothing in the world even remotely resembles these ideal situations.

Philosophers are, in fact, queuing up to attack the lawlike mode of explanation. Wittgenstein was dismissive of the whole approach when he wrote that 'the whole modern conception of the world is founded on the illusion that the so-called laws of nature are the explanations of natural phenomena' (1921:6.371). Lipton points out that it is easy to imply the facts in combination with a law, simply by inferring them from the conjunction of the two – the elliptical orbits of planets are implied by the elliptical orbits of planets plus a law of economics (2004:27). The reliance on deduction in the modern theory invited easy demolition, but the very principle that suggests we explain things by subsuming them within some known larger story seems to be misguided. It seems a limitation of empiricism that things can only be explained by the regularities that have already been observed, when finding the best explanation of something usually needs us to venture beyond experience. Everything we have observed about explanation so far suggests that we should follow the 'direction' which we find in nature, and that this is picked out by a dependence relation. The future depends on the past, effects depend on their causes, shadows depend on flagpoles, laws depend on their instances, and our main focus of enquiry should be to discover on what the behaviour of instances might depend. If that is beyond us then we may have to retreat to the next stage, but not without a struggle. Bird remarks that 'it is not a necessary condition on A's explaining B that we have an explanation for A also' (2007:59), and it is no objection to lawlike explanation that we have not yet managed to explain many of the laws. One option is to treat the laws as primitive, since no metaphysical scheme can escape adopting primitives of some sort. Maudlin takes the view that 'the laws of nature stand in no need of "philosophical analysis"; they ought to be posited as ontological bedrock'(2007:1), though that seems to leave them as primitives which are deeply puzzling. If the puzzle drives us to seek explanations of the laws, though, the covering-law model of explanation leaves us nowhere in nature to turn to, on pain of circularity, since the laws have done all the natural explaining. The only options are to investigate the supernatural, or to side with Maudlin.

31. Reductive explanation

We will reject subsumption under a law as central to scientific explanation (without denying that it may be interesting and revealing to learn that the fall of an apple fits into a universal pattern of behaviour, governed by a single force). Since we have offered some sort of 'hierarchy' in nature as a key to explanation, the natural next thought is that reducing an event to a lower level in the hierarchy may be what explanations require. We have already suggested that the 'foundations' of a hedgehog in particle physics do not offer much explanation of the hedgehog (and Lucretius spotted that 'one can laugh without being composed of laughing particles', II.988), but Bird's remark that we don't have to explain the explanation relieves us of an obligation to explain all the way down. Further explanations are always welcome, of course, but no one wants an infinite regress (of supporting turtles, perhaps), and we have already concluded that the foundations tend to be too remote from the explanandum to be relevant. If our explanations had to reach right to the bottom of the matter on every occasion, our ignorance of these depths would seem to mean that we currently possess no explanations.

On the other hand, giving an account of what lies lower down in the ‘levels’ of the natural structure is widely seen as very successful explanation. The shape and solidity of an object, for example, are well understood if we show how they arise from electro-magnetic forces within matter, and the ‘modules’ that constitute it. That reduction also involves causation, but we could also say that because a statue is made of bronze, it must thereby have all the characteristics of bronze, without reference to causation. Hanna observes that ‘explanatory reduction is the strongest sort of reduction; ...ontological reduction can still have an “explanatory gap”’ (2006:6), and it is important to see that giving a reductive account of something may still not be an explanation, particularly if the reduction arrives at something less well-known than what was being reduced. Lycan quotes Heisenberg as writing that ‘it is impossible to explain the manifest qualities of ordinary middle-sized objects except by tracing these back to the behaviour of entities which themselves no longer possess these qualities’ (1937; Lycan 1995:111), so that a reductive explanation of the redness of a tomato would be no good if it reduced the tomato to red particles. We therefore seek explanatory reduction, rather than ontological reduction.

However, there is another aspect of reduction which suggests that it cannot play the primary role in explanation (even if the reduction is ‘explanatory’), and that is the fact that reduction may be inherently eliminative. A perfect definition is eliminative, in the sense that the definiens can thereafter be substituted for the definiendum (or vice versa) without loss, and a perfect physical reduction has similar consequences, which lead Merricks, for example, to say that if the atoms of a baseball break a window, then it is ‘overdetermination’ to add that the baseball also broke the window (2003:56). It is not merely that the particles will not explain the hedgehog, but that if you give the account entirely on the level of particles, then it may become supererogatory to even mention the hedgehog. This is not to say that an explanation need not descend to the level of particles or components, but that much more must be added to turn an eliminative reduction into an explanatory one. We can agree with Paul Audi, who writes that ‘I deny that when p grounds q, q thereby reduces to p, and I deny that if q reduces to p, then p grounds q. ...On my view, reduction is nothing other than identity, so p is the same fact as q’ (2012:110). Rather than reduction, we need the connecting pathways from the atoms to the baseball, hedgehog or ear. An explanation must tell us ‘how’, and not just ‘what’, and this must involve structures and causes.

Fine suggests that the general concept of what is required here is ‘grounding’ rather than ‘reduction’, since in his view ‘the relationship of ground is a form of explanation, ..explaining what makes a proposition true, which needs simplicity, breadth, coherence, non-circularity and strength’ (2001:22), and he subsequently writes that ‘it is only by embracing the concept of a ground as a metaphysical form of explanation in its own right that one can adequately explain how a reduction of the reality of one thing to another should be understood’ (2012:41). The drive of explanation is always to show ‘how’ things happen, and not just ‘that’ they happen (or even that they happen very frequently). An attraction of the concept of grounding is that it not only brings out what is missing from mere reduction, but it also shifts the emphasis away from merely causal links within and between levels in reality. Ruben drew attention to the fact that a ‘determinative relation’ can subsume causal relations as one of a number of other types of ‘determination’ (1990:231), and thus Audi can point out that the disposition of a ball to roll can be determined by its sphericity, without rolling actually being caused (2012:104), and he also suggests that normative, aesthetic and semantic facts can have a grounding which determines them, and which is clearly not causal (2012:106). There is a disadvantage for the naturalist in departing from mere causation, which might fit neatly into the physics, and instead employing a relation of ‘determination’ which is considerably more vague, and may turn out to be a family of concepts, but something being determined is a well understood concept in ordinary life, and it offers a focus for the enquiries of metaphysicians.

32. Causal explanation

If we are rejecting the covering-law and the simple reduction models of explanation for physical reality, we can still enquire whether causation and mechanism, which remain from our list of the four best candidates for theories of explanation, are the actual keystone of explanation, perhaps as the most significant mode of ‘determination’. Lewis is noted for his commitment to causation as giving the entire character of physical explanation, and he tells us that ‘here is my main thesis: to explain an event is to provide some information about its causal history’ (1986c:217). For anyone with strongly naturalistic inclinations this must have an appeal (as must a causal theory of numerous other concepts), even if we label the central concept as ‘determination’. An obvious response would be that the ‘causal history’ of anything is a cone of events which widens out indefinitely into the past, and to mention the movement of a soup spoon in the year 1421 is quite irrelevant to explanations of World War One, despite all the causal links. Lewis’s proposal must be seen in the context of his thesis of ‘Humean supervenience’, with everything determined by the universal ‘mosaic’ of qualities. All explanations will merely invoke patterns in the mosaic, with Lewis’s sophisticated set-theoretic and mereological apparatus used to cross-reference and narrow down the information that matters. Lewis has an unlikely ally in Nietzsche, who wrote that ‘showing the succession of things ever more clearly is what's named “explanation”: no more than that!’ (2003:35[52]), where we should note the phrase ‘ever more clearly’. The ideal for these two thinkers seems to be that a full explanation is something like an omniscient mind’s awareness of the totality of the prior causal cone, which can then be perused at leisure.

A number of thinkers have enriched the causal account by adding procedures for homing in on the explanatory causal facts, among which we can mention three: Mill’s focus on when several effects share a single cause (‘agreement’) and when only a solitary cause can trigger some effect (‘difference’) (1843:3.7), the strategy of contrastive explanations, by selecting an appropriate ‘foil’ to illuminate the target of the explanation (e.g. Lipton 2004:33), and Woodward’s view that causal explanations are revealed by our manipulations, as when moving the flagpole moves the shadow, but not vice versa (2009:6.2). The latter approach, which we will not pursue, leads to accounts of causation in terms of equations and graphs which plot correlations found in tests of cause against effect. An example of the contrastive approach is offered by Schaffer’s nice example of the three invitations to explain why ‘*Adam* ate the apple’, or ‘Adam *ate* the apple’, or ‘Adam ate *the apple*’, proposing that this requires attention to Eve eating the apple, or Adam throwing the apple, or Adam eating a pear (2012:131), but actually the emphasis just seems to invite attention to the category of items mentioned in the subject or verb or object of the sentence, rather than any specific group of foils (useful though such things may be). The three invitations concern persons, actions and types of food, rather than any specific foils.

The obvious critique of the causal approach would be to find physical explanations that didn't involve causes, and causes that failed to explain. Of the first, Lipton observes that 'it is often easier to say what a factor would explain than it is to say what it would cause' (2004:137). Thus we are confident that cigarettes explain most lung cancers, long before we have any decent knowledge of the causes. Of the second Lipton says that 'we may think about causes without thinking especially about explanations' (2004:132), which would presumably involve thinking about features in the backward causal cone that had little to do with explanation, such as the creation of the heavy elements in remote stars when considering the rocks that cause a tsunami. These examples will not suffice to refute the causal theory, but in most cases we can see that an explanation requires more than a catalogue of causes, just as it required more than a citation of laws, or an inventory of reductive foundations. Aristotle requires four types of explanation rather than one, and he wrote that 'we think we know a thing only when we have grasped its first causes and principles and have traced it back to its elements' (*Ph* 184a12), which demands to know the causes, but also the 'elements' (which seems to be the constitution), and also the 'principles', which implies a more abstract grasp of what is operating to produce an event (and which might imply something like laws); both 'elements' and 'principles' fit well with the picture of grounding and determination that we have now encountered. There is not going to be a simple account of explanation, even for simple instances - which is why we were led to the concept of 'coherence' in considering the preconditions.

33. Models and mechanisms

In their discussion of Leibniz, Cover and O'Leary-Hawthorne make the suggestive remark that 'the philosopher comfortable with an 'order of being' has richer resources to make sense of the 'in virtue of' relation than that provided only by causal relations between states of affairs, positing in addition other sorts of explanatory relationships' (1999:18). This reminds us of the structural precondition which we proposed earlier, and implies that when we identify a causal nexus leading to some event we are unlikely to understand this nexus in a 'flat' way, like wires converging on a box, but will see a complex of interwoven strands of causation arrayed across our preconceived structure of the world. If we give the causes of a car accident, they will involve weather, drivers, car components, laws, road configurations, chance encounters, and so on. We slot each of these causes into some 'level' in our structure of the world and produce a multi-dimensional model of the incident, built from determination relations, and we are barely conscious that we do this. This model, built from the tools of our preconceptions, is much closer than any mere recital of causes to what we mean by an explanation. By 'model' we mean here some kind of spatio-temporal representation that is isomorphic to the explanandum (what Portides 2008:386 calls the 'semantic' view of models), rather than the 'received' view of models developed formally by Tarski in the language of mathematical logic. While a model might be wholly specified by a set of indicative sentences, we don't want to rule out a very illuminating model which is made of matchsticks. There may be a concept of 'model' between the two, consisting of conceptualised structures built on abstracted inputs and outputs, as sketched by Strevens, though he also comments that in the sciences 'almost any formal construct may serve as a model' (2008:15-16). We will look further at the concept of a model in the next chapter.

If we have eliminated three theories of explanation, does this mean that the fourth (mechanism) must be the correct one? This seems promising if we note that the sketch of a 'model' just given is very close to what we might mean by a 'mechanism'. Salmon, a leading student of scientific explanation, eventually came to champion this approach, and wrote that 'causal processes, causal interactions, and causal laws provide the mechanisms by which the world works; to understand why certain things happen, we need to see how they are produced by these mechanisms' (1984; Ruben:211). Most defences of mechanistic explanation, such as Machamer, Darden and Craver's influential paper, offer fairly literal accounts of what is meant by a mechanism. Those authors emphasise that a mechanism is not just a 'push-pull' system, but give as a definition that 'mechanisms are entities and activities organized such that they are productive of regular change from start or set-up to finish or termination conditions' (2000:3). The main motivation behind the shift from explanation by laws to explanation by mechanisms is the realisation that biology is as much a science as physics is, and it is clearly right that a fully satisfactory explanation of the lung cancer is a complete account of the mechanisms involved, because they show 'how' it comes about. However the world is full of events which can be explained causally without invoking anything like the mechanisms defined in the paper. In particular, there may be good explanatory sequences which are neither regular in their action, nor clear-cut in their starting or terminating conditions. Their paper is concerned with 'scientific' explanations, just as Hempel was similarly concerned, and this raises the question of whether scientific explanations are a distinct species from the explanations of ordinary people. We noted that explanations of historical events, and other human activities, seem quite different in character from explanations by law or by mechanism.

At this point we should recall that the theoretical structure of the natural world within which most of us interpret our experiences has 'levels', and it is plausible to think the levels are comprehensively interconnected. Many thinkers defend a severe dislocation of level in any model at the point where conscious or rational minds enter the picture, but we will not engage with that problem. If we greatly relax the definition of a mechanism, and compare it with the 'model' of the car accident which was suggested, it is a reasonable metaphorical use of the word 'mechanism' to ask what mechanisms led to the accident. We might even stretch the word 'mechanism' rather a lot, and speak of the understanding the mechanisms (or interactions, or processes) that led to the accession of Richard III. Such mechanisms can be specified on many levels, from 'above' and 'below' the accident or accession, and in this sense the mechanistic view of explanations seems the most promising, though the 'model theory of explanation' might be a better label for it. If, for each puzzle facing us, we were offered a 'model' of how it worked, many of us would feel that explanation had been achieved.

34. Levels of mechanism

Machamer, Darden and Craver see the world as a hierarchy of nested mechanisms, and propose that explanations can ‘bottom out’ at the point where mechanisms dissolve into more fundamental ingredients, and they write that ‘there are four bottom-out kinds of activities: geometrico-mechanical, electro-chemical, electro-magnetic and energetic; these are abstract means of production that can be fruitfully applied in particular cases to explain phenomena’ (2000:22). This is a satisfying picture, for those inclined to physicalism, but it seems to imply that all explanations are incomplete if they fail to touch bottom in some way – yet this seems quite wrong for explanations at the highest level. To explain the tsunami fully we might want to push on from the clash of two plates, and explain why the Earth has these plates, but we certainly don’t need to quote the laws of electromagnetism. The fact that the laws of electromagnetism reveal the behaviour of all the matter in the universe, and not just the behaviour of rock in tectonic plates, indicates that the explanation has (for most purposes) moved to a lower level than is normally required. We do not explain a tsunami by explaining all of the matter in the universe.

It seems obvious that an explanation is felt to be complete when it reaches the bottom of the level where the puzzle is located, and not when it reaches the bottom of the whole system. That is a thought we will seize on, because it not only seems right, but it gives a framework in which a fairly comprehensive account of explanation can find a place. If we consider specific events like tsunamis, general causal connections like smoking and cancer, and universal connections found in any law of gravity, their explanations must have natural limits (as distinct from pragmatic limits set by our capacities or curiosity), and the concept of ‘levels’ seems the only one available to set the limits. Explanation of cancer by tobacco has upward limits that level off around the economics of the tobacco industry (because cigarettes have many additives), and lower limits that level off somewhere around the molecules distinctive of cell chemistry, and similarly for the other examples. With that picture in mind, we can modify the thought we have seized on, and say that to be complete an explanation must reach the *top* of its level as well as the bottom (which is a useful corrective to the reductive approach), and this endorses the idea that a ‘model’ is what does the explaining. Thus a good explanation is a comprehensive and coherent model of all factors relevant to the production of the explanandum, extending to the limits of the level of reality in which it is located.

35. Abstracta

An obvious objection to a wholeheartedly causal and mechanistic approach to explanation is that it will make explanation of highly abstract systems and truths impossible, since they don’t seem to be causal at all. One response to that might be to say that ‘gravity’ or ‘cancer’ are generic and simplified concepts which have been abstracted from a complex physical reality, so that even theories of the behaviour of physical objects will involve some degree of abstraction (Burgess and Rosen, for example, judge that ‘much of what science says about concrete entities is “abstraction-laden”’ (1997:179)). From the other end of the business, we might venture to naturalise even high levels of abstraction such as arithmetic and logic, finding roots for them in the physical world, as when Russell sees the origins of arithmetic in counting sheep (1907:272). One way or the other, we might find unity across the notoriously vague abstract/concrete divide.

Alternatively we might concede the radical separateness of such systems, but loosen the concept of ‘causation’ to embrace them. Thinking of the grounding relation needed for explanation in terms of ‘determination’ rather than the narrower ‘causation’ offers a means (or at least a vocabulary) for broadening the enquiry sufficiently for us to ask whether, for example, Peano Arithmetic is in any sense determined by the truths of some version of set theory, and whether this might be sufficient to offer at least part of an explanation of arithmetic. In that framework, we would say that causation is a particular instance of the determination relation. Key questions will be whether there is a ‘direction’ of dependence within abstract systems, and whether they exhibit the ‘modularities’ that seem needed to restrict the domain of any explanation. We will discuss such systems later to see whether our accounts of explanation and essentialism might have application for them, and whether something ‘determines’ (for example) the relations of squares on right-angled triangles.

36. Convergence and fruitfulness

A final and important observation concerns the concept of ‘convergent’ explanations. There is a familiar idea (known as ‘consilience’) that a scientific theory is greatly strengthened when evidence and measurements converge on a consensus, and a convergence of explanations has similar power. The really wonderful explanations (what Lipton labels ‘lovely’ explanations) are those where a huge number of phenomena are explained by one simple idea. Of the theory of evolution Dennett says that ‘if I were to give an award for the single best idea anyone has ever had, I'd give it to Darwin’ (1995:21), and this is obviously because of the huge explanatory power of the theory (of which Darwin himself wrote that ‘it can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified’ (1859; Lipton 2004:206)). An explanation of one thing may also explain other things. The explanation of a car accident may also explain three other accidents in the same location, or three other accidents in the same type of car, or three other accidents by the same driver. Explanations do not occur in isolation, but frequently converge in a variety of ways. The diverse phenomena of lightning, magnetism, and static electricity gradually turned out to have a single explanation. It seems clear that within each ‘level’ of reality, there are patterns of explanation that converge on certain key concepts, and also that concepts belonging to lower levels can be fruitful in producing explanations in several different areas of the level above. Science investigates these interrelations of converging and fruitful explanations, as much as it investigates the separate explanations of the phenomena. This aspect of explanations will, as we will see, be significant for the essentialist approach.

FOUR

Essence for Explanations

37. Psychology of essences

Essences are controversial in contemporary philosophy, and the word ‘essence’ is often given a meaning very different from the concept we found in Aristotle. We may attempt to eradicate the concept of essence from our best theories of science, but eradicating it from human psychology seems impossible. Susan Gelman’s researches into thinking in human infants has shown conclusively that essentialist modes of understanding are deeply entrenched in young minds, and may well be innate, since they are not learned from parents, and they appear to extend widely across diverse human cultures (2003). Gelman connects the essentialising tendency with two related cognitive needs in children – the need to categorise, and the need to explain, leading them to intuitively assume that many things are members of some natural kind, and that many things have hidden mechanisms and features which explain what is observable.

Her basic finding about essentialism in children is that ‘the three components of essentialism as a folk belief are the idea that certain categories are natural kinds, the idea that some unobservable property causes the way things are, and the idea that words reflect real structures’ (7). In more detail, she summarises her findings on kind essences thus: ‘by five children assume that a variety of categories have rich inductive potential, are stable over outward transformations, include crucial nonobvious properties, have innate potential, privilege causal features, can be explained causally, and are real’ (136). The kinds into which things are categorised do not appear to result from careful research and inductive generalisation, as she reports that ‘with kind essentialism the person assumes that the world is divided up into pre-existing natural categories’ (152), and this seems to arise from a deeply ingrained assumption, more akin to face recognition than to a rational cataloguing procedure. A striking feature of infant thought is the great caution about taking things at face value, seen in the interesting phenomenon that ‘people favour historical paths over outward properties when determining what something is. ...An object looking like a knife is less likely to be called 'a knife' if it is described as having been created by accident’ (151). That example shows that their understanding is not merely concerned with concealed structures, but also with function, purpose, origin and social role. We may be tempted to assume that language imposes this mode of thought on the children, but essentialist categorising occurs earlier than language (179). Her conclusion is that children categorise as a step towards understanding, and their target is a knowledge of inherent hidden properties (286).

Children closely relate essences to kinds of things, though they can hardly avoid the distinct individual characters of people and pets, but what they are after is pretty close to Aristotle’s notion of ‘what it is to be that thing’, or its ‘essential nature’. The topic of study here is related to the vast and interesting question of how we conceptualise the world, but in the distinctively essentialist view of that activity ‘children incorporate a variety of nonobvious features into their concepts, including internal parts, functions, causes, and ontological distinctions’ (13). Very empirical theories of concept-formation which tie concepts closely to our experiences of surface properties seem to be at variance with the facts about how children think. There is also support for the very causal view of essences (which scholastics came to favour), because ‘properties that enter into causally meaningful links are better remembered and are treated as more central to the category than properties that are not causally meaningful’ (116). A more surprising and less Aristotelian finding is that kind essences come in degrees. Earlier views took animal species as fixed, which post-Darwinian thinking rejects, so we may not be too surprised by this, but it is in the conceptualisation of kinds of people that this more flexible essentialism is found, since Gelman writes that ‘kinship is essentialized, but admits of degrees, ...and people can be essentialist even about categories they do not view as fixed over time, such as age groupings’ (88).

The phenomenon of essentialising about groups of people points to the problems which accompany the benefits of essentialist thinking, which can be summed up in the word ‘prejudice’. Racism is the most obvious instance of this, and it has been found that ‘the notion of caste in India is more essentialized among upper-caste than lower-caste individuals’ (179). Other errors creep in elsewhere, and they make mistakes in inductive generalisation by over-essentialising (150), and ‘children overestimate the power of a single example’ for similar reasons (147). Essentialist thought does not guarantee wisdom.

Once children develop language, an interesting new phenomenon emerges, which is that nouns are more essentialist than verbs: ‘children judged personal characteristics as more stable when they were referred to by a noun ('she is a carrot eater') than by a verbal predicate ('she eats carrots whenever she can') (189), and we can see that the label provided by a noun runs much deeper in our understanding of something than the descriptions provided by verbs. Gelman summarises a great deal of research when she notes the phenomenon of the ‘label’, and writes that ‘labels may signal categories that are believed to embody an essence’ (55). A line of thought here (too rich to be properly pursued) is that this sort of ‘noun’ thinking, with its strong essentialist overtones, connects with the recently developed idea that minds are most clearly understood as filing systems. Fodor writes that the mental representations he has defended for so long ‘can serve both as names for things in the world and as names of files in the memory’ (2008:94), and Recanati has explored the concept of mental files most illuminatingly. Files are complex, and beyond their simple labels (which refer to objects) they have a hierarchical internal structure based on increased permanency and importance, with each non-transient file containing encyclopaedic information, and cross-references to other files. There are also ‘indexed’ files, containing other people’s concepts rather than one’s own. This picture turns out to be very fruitful (handling cases such as Phosphorus and Hesperus, or Paderewski, very neatly), and fits the essentialist picture observed by Gelman very well, since there is a core of information which accompanies each label (called the ‘nucleus’ of the file by Recanati, 2012:100). The phrase ‘carrot-eater’ seems to open an object-file, in a way that the verb does not, and an essence is developed for the object.

Given the controversial status of ‘folk psychology’, and the disrepute into which ‘folk’ physics, ‘folk’ reasoning and ‘folk’ ethics have fallen in recent years, it would be very rash to build a philosophical theory of essentialism based on these ‘folk’ essences. The reason why we should attend to them, though, is that they show the ‘grain’ of normal human thought, and it is a rash philosophers who moves profoundly against the grain of humanity. One thing Aristotle clearly stands for is to begin our philosophy in what most people think about a topic, and only modify the normal view if it seems unavoidable. Essentialism is the norm for ordinary thinking, and it is ordinary thinking which the Aristotelian view tries to articulate. There may be a huge dislocation between ordinary thought and scientific thought, but we take such a dislocation to be best avoided if possible. Scientists are, after all, ordinary people.

38. Necessary or essential?

People, then, seek the essences of the things they encounter, presumably aiming to understand the world, in order to cope with it. It may have appeared that in the seventeenth century scientists abandoned this everyday approach to understanding, but a plausible modern view is that they simply raised the standards of enquiry, and that scientists are just as essentialist in their thinking as the rest of us, but better at it. In the early twentieth century theorists of science had firmly abandoned the idea of essences, along with the long-forgotten ‘substantial forms’. Russell, for example, takes the distinction between the essential and the accidental to be a ‘useless’ distinction (1903:§443). With the Aristotelian concept of essence now defunct, this left the word ‘essence’ available for a new usage, and the development of semantics for modal logic had just that result. The syntax for a logic containing ‘necessary’ and ‘possible’, or ‘sometime’ and ‘always’, had been worked out, but what did it all mean? A problem arose: if you write ‘Fa’ you just mean that the predicate ‘F’ applies to the object ‘a’, but if you write ‘□Fa’ you mean that ‘F’ cannot fail to apply to ‘a’, and that seems to require a sharp distinction (since, after all, this is formal logic) between Fs that can’t fail to apply and Fs that can fail. Quine denied such a sharp distinction, and thereby repudiated modal logic (e.g. 1953a; 1953b). In the ensuing discussion, the word ‘essential’ was a convenient label for the Fs that must apply to something (or at least when it exists). A defence of modal logic needed a defence of ‘essential’ predicates, in this new sense of the word.

The problem not only concerns the predicates which apply essentially, but also involves the objects. If we say (with Bishop Butler) that everything ‘is what it is, and not another thing’, there seems to be a necessity in the predicate, no matter what it applies to, but if we say that ‘my cat might have been a bird’ then ‘being a bird’ doesn’t seem essential, and the problem focuses on my cat.

There are two distinct problems here, of reference and of truth. One reading of ‘my cat might have been a bird’ has it that my cat might never have been a cat, because some genetic intervention early in its career turned it into a bird, and we can just allow that as a genuine bizarre possibility. If I say ‘if my actual cat were a bird, it would fly’, I evaluate that by imagining a reality in which there is a bird, and asking whether it is the same entity as my cat, and this invites the question of how you know you are referring to my cat if it appears to be a bird, and thus how the sentence can be meaningful. We might dismiss that problem, by accepting Kripke's account of the matter, and saying that it is my cat because I say so – that is, it is a stipulation of the original sentence that it is my cat that is being considered as a possible bird (1980:49). Some philosophers (such as Plantinga, Adams, Forbes and Mackie) remain worried that the speaker is stipulating that something is a cat when it has lost every known characteristic of cats, and that the original object might therefore need a 'haecceity', a predicate that holds fast to the object's identity through extreme vicissitudes of predicate-change. Leaving that question to the specialist, we can see that the second problem here contains traditional philosophical difficulties: how do you evaluate the truth of a claim concerning what my cat *might* be? If we picture the gradual metamorphosis of a cat into a bird, does it not cease at some point to qualify as a 'my cat' or even as 'a cat', so that the original modal claim (that my cat *might* be a bird) is simply false? That seems like common sense, but what are the criteria for a cat to cease being a cat? Without some such criterion (of what is 'essential' to the cat) we cannot evaluate modal claims of that sort.

The standard modern response is that there must be some predicates which are necessary for the existence of any object. Cats, as long as they exist, must have the features required to be such a creature, and 'my' cat may need to be of a certain breed, size and colour. We find that an 'essence' has now become nothing more than a list of predicates which are necessary for some entity to exist. Doubts about this modern approach began when Marcus observed that 'necessary' and 'essential' are not substitutable for one another, since saying that Socrates was essentially snub-nosed did not mean that he was necessarily snub-nosed, and saying Winston was essentially a cyclist (Quine's example) did not mean that he was necessarily a cyclist (1971:193). The problem becomes obvious when we see that Forbes and Penelope Mackie have to deal with the fact that many necessary features of some existent thing are clearly 'trivial', and yet the epithets 'essential' and 'trivial' contradict one another in ordinary usage. Mackie simply stipulates that some trivial necessities (such as the proposed haecceity 'being identical with Socrates') will be ignored (2006:20), while Forbes itemises three types of trivial necessity for exclusion, which are the entailments of a thing's descriptions, the properties of existence and self-identity, and relations to necessities in a different category (1985:99). Neither of them has any resources for expounding what would make a feature non-trivial, and since their aim is to establish identity conditions for objects across possible worlds, it is a symptomatic weakness of their whole approach that they exclude 'being identical with Socrates' and 'self-identity' as trivial, given that these seem to be of foundational importance to identities. Without a reasonably clear distinction between 'important' and 'trivial' there is not much sense to be made of essences, and that particular debate should really confine itself to 'necessary' features - as indeed does Della Rocca when he begins a paper on essentialism by saying 'some philosophers distinguish between necessary properties and essential properties. This distinction is irrelevant to my purposes; following Yablo, I shall ignore this distinction in what follows' (1996:186 n1).

Fine made the whole issue much clearer when he pointed out that not only are 'essential' and 'necessary' not inter-substitutable (as Marcus had spotted), but that essential relations are directional, whereas necessary relations are not. In a famous example, he pointed out that Socrates is essential to the status of his singleton set, but the singleton set of Socrates is of very little interest to the man himself (1994). Fine offered the traditional locution 'in virtue of' to capture this directed aspect of essences, and referred to the 'natures' of Socrates and his singleton set to explain the relationship (so that the nature of the singleton requires Socrates, but the nature of Socrates does not require set membership). Given the existence of Socrates, it will also be necessary that 2 + 2 = 4, but that has nothing to do with the nature of Socrates. There appeared to be a 'dependence' relation involved in essentialist talk, and more recently the concept of 'metaphysical grounding' has been developed to enhance the picture. Thus space was now created for Aristotelian essentialism to re-enter the stage, especially once it became respectable to again talk of the intrinsic 'natures' of different things, which barely differs from talk of *to ti en einai*.

Given that human psychology seems essentialist, and that there is more to essences than mere necessary properties, and that the natures and groundings of things are required to explain many areas of our understanding, the traditional concept of an essence must be re-examined. A study of Aristotle suggests that he only offered essence as the lynchpin of his metaphysics because it emerged from a study of explanation, and we will now try to see if we can clarify a modern concept of essence for such a context. First we will consider what essences are meant to explain, then what is required of essences to achieve this task, and finally consider how we should characterise the relationship between explanation and essence.

39. Metaphysical explanation

In *Categories* Aristotle said that substances do not come in degrees (*Cat* 3b33), and his ontology admits nothing that is not a distinct and unified substance, but the difficulty of maintaining such a commitment to a cosmos of unified substances was most evident in Leibniz, who desperately tried to give a modern theory of unity, and struggled with his unsatisfactory theory of monads. It is tempting to simply abandon the 'problem' of unity, allowing that unity in physical objects is real enough, but comes in degrees and has no absolute, and that the concept of unity is largely a matter of human convention and human interests (as when Locke observes that we can unify anything we like in thought). However, even the most radical sympathisers with 'nihilism' about unified physical objects tend to find some sticking point, such as Merricks's claim that human organisms are unified (2003), or Van Inwagen's claim that most lives are unified (1990), and it seems that to say that an electron entirely consists of the three properties or tropes of mass, spin and charge, with no reference to their working as a unified team, leaves a gap in the ontology of physics. For Aristotle, unity was taken as primitive in *Categories*, and then explained by the hylomorphic theory in *Metaphysics*. If we try to follow the latter approach, we now seem to face the problem that science has not come up with a feature in physical objects which qualifies as its 'form', and so the Aristotelian approach can only be defended at a certain level of abstraction from the physics, in the aspect Aristotle picked out with the term *arché*, the guiding principle of the thing. There then seems to be a priority problem between the unity and the essence: if we can't identify the physical essence, and infer that we are therefore dealing with a unity, it seems that we must first identify something as being unified, in order to distinguish the principles that do the unifying. There seems no alternative here, and essence will not give us a criterion of unity – but it still may offer an explanation of unity. We can use 'surface' criteria to decide whether something is to be regarded as highly or totally unified, and then ask whether there is some 'deeper' principle operating which generates the unity we have picked out (given that the surface can be seen as no more than a 'mosaic', and so our addition of unities is a puzzle). We will return to this question below, to see whether essences play a role in our attempts to understand the unity of objects.

The puzzles concerning the supposed unity of objects lead to a family of other much discussed puzzles. If a thing is known wholly by its attributes or predicates, what is their subject? How do we individuate particulars and kinds, in perception and thought? How do we track those objects if they are intermittently experienced across time and space? If objects change, what is the subject of the change? How do we assess identity claims made in modal contexts? Each of these questions leads to large clusters of issues in ontology, epistemology, semantics and psychology, so we will impose an account from the perspective of the current enquiry. Let us divide the ways in which we relate to the physical world into five areas: 1) picking things out and counting them, 2) tracking them, 3) filing and conceptualising them, 4) classifying and defining them, and 5) extrapolating from them, judging them, and imagining them. For each of our areas, we must consider whether explanation makes a contribution, and whether the type of Aristotelian essence we have been identifying has a role to play.

40. Explaining individuation

The first area of thought in which essences may have an unavoidable explanatory role is in picking things out in order to consider and discuss them. The term 'individuation' is used rather loosely in philosophy. The concept often has rich connotations which include persistence conditions, amenability to definition, modal profile, and so on, but our main concern is the minimal step of picking something out for consideration, seen as the first step towards understanding.

It is here that we encounter a key question for modern essentialists, which is whether to accept Wiggins's view that it is in the process of basic individuation that essentialism should find its place, rather than in the practices of modal reference or explanation. Wiggins, following Frege, focuses on the identity relationship. On the one hand Quine had understood the concept of identity as a precondition for individuating any object over time. If we somehow pick out some entity, and pick it out again at a later time, the concept of identity is required to unite the entities as one. The process gets started by acts of pure ostension, gradually focusing on some entity by inductive inference, though concepts will soon enter the process (1950:67-8). In his quest for the minimal formal framework required for ontology, he later suggested that this identity relationship could (if so desired) be replaced by a conjunction of formulas of the form 'if Fx then Fy', which effectively replaces identity with indiscernibility (which might be further reduced to sets of objects) (1960:230).

On the other hand, Geach had boldly argued that, since questions about sameness lead to the question 'same *what*?', identity must be an entirely relative matter. Equality, identity and sameness are understood as entirely dependent on the predicate under which they fall on any given occasion (1962:39). Geach's view has found few supporters, and most philosophers endorse Perry's view that Geach has described the relationship of resemblance (in terms of some respect or property), rather than the true identity relation (1970:note 12). To abandon the relation of perfect identity would mean that you could not say two sets are perfectly identical if they have the same members, which is a basic axiom of set theory. The idea that identity has any of the complex problems with which Geach wrestles is dealt with swiftly by Lewis, who writes that 'identity is utterly simple and unproblematic. Everything is identical to itself; nothing is ever identical to anything except itself. There is never any problem about what makes something identical to itself; nothing can ever fail to be' (1986a:192).

Wiggins rejects the simple Lewis view, and takes identity and individuation to fall somewhere between the accounts of Quine and Geach. On the one hand, Quinean acts of 'pure ostension' looked implausible, but on the other hand the Geachian claim that identity actually meant 'the same *sort* of thing', rather than just 'the same thing', seemed wrong. A standard concept of identity must be retained, which is taken to be a transitive, symmetric and reflexive relation conforming to Leibniz's Law (which asserts that if two things are the same then they are *entirely* the same) (Stevenson 1972). If, in addition, the process of 'pure' direct individuation is impossible, then the solution says that it is the *acts* of individuating and identifying which must involve a covering concept. Wiggins's basic position is summarised in the remark that 'understanding the concepts involved in individuation can only be characterised by reference to observable commerce between things singled out and thinkers who think or find their way around the world precisely by singling them out' (1980:2), and the involvement of the thinker means that conceptualisation is an immediate and essential ingredient of the individuation process (a doctrine he calls 'Conceptualism', of which he prefers the 'Realist' version). The notion of the 'sortal' concept, introduced by Locke, seems perfect for the job (*Essay* 3.3.15). The simple conclusion is that 'there could be no singling out *tout court* unless there could be singling out 'as'' (1980:5). The theory is 'realist' because it is not merely a matter of imposing our concepts on the world (though our interests are also involved - 1980:133), but of also drawing on our scheme of sortals in response to experience, which mainly concerns detecting a 'principle of activity' in each thing (2001:137).

This presumed fact about our mode of grasping the things in the world has extensive consequences for how we understand it, since sortal concepts bring a huge array of information with them, which Wiggins is happy to accept. It is here that he draws on the reading of Aristotle which favours *Categories* over *Metaphysics*, so that we have primitive entities satisfying the formal conditions of identity, and then the sortal concepts used for picking entities out can play the role of 'secondary substance', which answer the question 'what is it?' by invoking its kind, rather than its individual form. Hence Wiggins writes that 'answering "what is it?" with the secondary substance identifies an object with a class of continuants which survive certain changes, come into being in certain ways, are qualified in certain ways, behave in certain ways, and cease to be in certain ways' (1995:218), but the theory of individuation dovetails into an account of how we come to understand the wider world, since 'the sense of the sortal term under which we pick out an individual expands into the scientific account of things of that kind, where the account clarifies what is at issue in questions of sameness and difference of specimens of that kind' (1995:242). In general, the appropriate sortal for individuation gives the essence of the thing, the main consequence of which is knowledge of 'what it is'. Effectively, if you identify the thing coming through the undergrowth as a 'tiger', you instantly know you are in trouble, since you know a lot about tigers.

However, the account of Wiggins is quite different from the present one, and we should not think that the sortal essence of a thing offers a comprehensive explanation (for which far more cross-referencing work seems to be required), since he tells us that 'essences of natural things are not fancified vacuities parading themselves ...as the ultimate explanation of everything that happens in the world. They are natures whose possession is a precondition of their owners being divided from the rest of reality' (2001:143). Hence not only does the sortal concept offer a mechanism by which thinkers individuate the contents of their world, but essences are also the 'natures' of things, though of a generic character, such as to mark the contents of the world into kinds. The 'realism' of the theory resides not only in the intransigent reality of the external world, but in the reality of its division into kinds, to which our conceptualisations meaningfully respond. Wiggins's theory has heavy metaphysical commitments, inviting comment at a number of levels.

The most obvious immediate objection to this simple account of Wiggins's rich theory had already been offered by Locke, who wrote that 'if anyone thinks that a man, a horse, an animal, a plant, are distinguished by real essences made by nature, he must think nature to be very liberal, making one for body, another for an animal, and another for a horse, all bestowed upon Bucephalus' (*Essay* 3.6.32), the simple point being that a horse will fall under a large array of sortals, and they can't all be the essence of some horse. Wiggins is aware of this problem, and makes the obvious point that if you are 'tracking' an object, it would hardly lose its sustained identity whenever the sortal under which it was considered was modified (2001:22). Hence he needs to distinguish which sortals will do his individuating and essentialist job, and say why we should give those terms priority over their rivals.

His clearest elucidation of the matter distinguishes 'purely generic sortals', such as *animal*, *machine*, *artefact*, and sortals which are 'pure determinables', such as *space-occupier*, *entity*, *substance*, which leaves the sortals needed for his theory, which have 'the special role of the substance-concept *man*, *horse*, *willow tree*….in marking simultaneously what a thing is, what matters turn on with regard to its persistence and what matters turn on with regard to identity claims relating to it' (2001:69). He had earlier introduced the term 'substance-concept' as the sortals which 'present-tensedly apply to an individual x at every moment throughout x's existence, e.g. *human being*' (rather than those that do not, such as *boy*, which is a 'phased sortal') (1980:24). He allows considerable flexibility, but claims that for any individual there must be 'at least one substance-concept' (1980:65). Thus the essentialist sortals need a traditional concept of substance to latch on to, the hallmark of which is that it reveals a thing's nature, and gives criteria for whether it persists, or partakes of identity relations. Given that the rejected generic and determinable sortals are too wide in application, the preferred substance sortals are close to the traditional 'infima species', the narrowest species to which something belongs (as long as it is not 'phased'). In the taxonomy of nature it is by no means easy to decide which is the 'narrowest' species of a thing, and various criteria such as inter-breeding and genetics need to be invoked. The difficulty of deciding which of the candidate sortals will qualify as the substance-concept is symptomatic of the real difficulty in Wiggins's position, which is that he is trying to run together the rather simple and immediate activity of picking something out for consideration and the rather complex activity of specifying the kind of the thing, along with its essence, which imposes many conditions on the thing. It is hard to see how the simple and immediate activity of individuation could be so rich and laden with information, given that 'what on earth is *that*?' tends to pick something out very successfully in most conversational contexts.

When Aristotle wanted to know what something is, he may have looked to a sortal label in *Categories*, but the key to his enterprise is the requirement of a definition, which is inescapable in Aristotle's account (as seen in *Posterior Analytics*), even if we don't warm to hylomorphism. We can hardly demand that singling out a tiger requires us to first define it, so that Wiggins's view of essentialism, whatever its merits, is quite a long way from qualifying as 'Aristotelian'. If we use the language of 'mental files', we might agree with Wiggins that the *label* 'tiger' needs to be invoked to individuate the tiger, and that the essence of a tiger will be the generic one found when we open the tiger file. In that account, though, the essence in the file and the label used in reference are quite separate (since our well-labelled files sometimes contain false information), and the sensible possibility is allowed for that the file may be almost entirely empty, if we have succeeded in picking out some new object which we cannot classify, and of which we are almost totally ignorant.

There seem, in fact, to be many ways in which basic individuation can occur. Ayers argues that highly generic concepts such as 'object' or 'thing' may be needed, to pick a fish out in its water or a bird in its air, but that Quine's pure ostension is then quite reasonable (1974:139). If we allow concepts and language to participate in basic individuation, then there seem to be many techniques available to us, other than specifying the 'substance-sortal'. We can individuate what is within some boundaries, as when we buy a house and garden. We can individuate by some minor accidental property, as when we pick out pictures with red dots on them. We can individuate things that fall under sortal terms, descriptions, definitions, natural kinds, classifications, properties, rules, boundaries, and temporal or spatial locations. Such a pluralistic account will suit individuation by non-human animals and human infants, who are not, we may presume, well equipped with sortal substance concepts, but navigate the world of objects very successfully, picking things out in their own way. Gelman's research suggests that essentialist thought arrives early, but only as an intuitive curiosity about what has already been individuated.

The proposal that the essence of a thing is invoked in the very act of initially individuating it seems, therefore, to be incorrect, even if we allow that an essence is encapsulated entirely within a substance-sortal concept. Given that one might pick out some object merely by placing it within certain boundaries (such as 'the object in that jar'), it seems clear that the concept of essence which we are investigating is not involved in individuating a thing. Even if the essence of the thing picked out is fairly obvious, as in the case of a geometrical triangle, it does not appear that the essence must be specifiable before the individuation can occur, and one might even say that the individuation *must* precede the identification of the essence. It is often said that sortal terms are what enable entities to be counted, and that offers a promising approach to the issues discussed here, so we will examine it further in the next chapter.

41. Explaining tracking

Picking something out may not require essences or explanations, but keeping track of things is a different matter. Restricting the issue to physical objects, the dimensions of tracking are across time and space, through intrinsic and relational change, and in talk of possibilities. Leibniz said (pessimistically) that ‘it is impossible for us to know individuals or to find any way of precisely determining the individuality of any thing except by keeping hold of the thing itself’ (1710:289). Locke suggested that we simply individuate things by their time and place (*Essay* 2.27.3), but Leibniz pointed out that we individuate times and places by the things occupying them (1710:230), so that won’t do. Once the ‘keeping hold’ is disrupted, our tracking of objects through space and time seems to be largely inferential. I presume that I see the same man each morning on my train, until I learn that he has an indistinguishable twin brother, when I adjust my assessment, in accordance with induction aimed at coherence. If the man’s appearance is greatly changed by maiming or ageing, I try to infer the single space-time path that he has followed between my two encounters with him, since we know from experience that people follow unbroken space-time paths. Then we recognise that there may be ‘intermittent’ objects, such as the Frauenkirche at Dresden (which spent sixty years as a pile of rubble). In neither of these cases does the ‘essence’ of the man or church seem relevant, since the later version is accepted as being the same (or not) on the basis of fairly superficial criteria, such as looking similar, or having a similar function, or simply being referred to by the same name.

Tracking across time and space does not involve essences, but tracking through change may. The Aristotelian approach encountered the problem that we are to say that a thing remains unchanged, or retains some sort of identity through the change, if the ‘form’ is unchanged, but this simply pushes the question of change to a different and more obscure level, and we saw how early scientists turned their back on that approach. Russell takes the view that essences actually obscure our understanding of change (1903:§443), but his approach seems to be to treat change as primitive, and then deal with it in terms of motion (§442). We may say that the persistence conditions for a tiger are intrinsic to the concept of the kind, so that if a tiger somehow (magically) transmutes into a goat, it will undergo substantial change rather than alteration at the point where it ceases to be a tiger. This view, though, seems to rest on the idea that the essence of tigerhood is prior to the natures of particular tigers, and we can at least agree with Jubien that ‘it is simply far-fetched - even incoherent - to think that, given an entity, of whatever kind, its being a single entity somehow consists in its satisfying some condition involving the kind to which it belongs (or concepts related to that kind)’ (2009:47). He points out that the persistence conditions for a particular object heavily depend on its parts, but a kind of object depends mainly on the arrangement (2009:15). If we gave a tiger a brain transplant from another tiger, it would still seem to pass all the tests for being a tiger, but we might think the change to that individual tiger was a bit too fundamental to be mere alteration.

This problem of identity across change seems to fit nicely with the thesis we are propounding. In the example of the tiger with the new brain, fans of individual tigers might claim that it was true change (since a major piece of this tiger was lost), while fans of the tiger kind might claim that it was alteration (since all the requirements for being a tiger are still there). This may seem a matter of taste, until explanations are sought for the two views. The claim that kind is all that matters, and that a brain transplant is a mere alteration, will be defended by referring to the generic kind essence which defines tigerhood. The claim that there is true change here will say that the essence of an individual tiger is in its distinctive individual nature, embodied in a particular brain. Either way, we seem to be faced with claims about persistence or non-persistence of a tiger which are intuitive, but which become explicitly essentialist when explanations are required.

On this view, essentialism does not solve the problem of change, since that would require the prior identification of the essence, which could then be checked for survival (by looking under the bonnet, so to speak) after some event of transmutation. This would be a greater difficulty for the individual approach, since the essence of the tiger kind is an established public fact, but the essence of a particular tiger is almost impossible to pin down. However, we can stand by the claim that the essence of a kind is entirely dependent on the essences of individuals. If, for example, it were agreed that the 'persistence conditions' for human beings involved a maximum age of 122 years, and then someone lived to be 130, this would not rule them out as non-human; rather, the kind essence would have to adjust to the new individual. If we persevere with the view that kind essence can be identified independently, and then used to judge a change, we should note that the essence of the tiger kind itself involves change, because they grow up from being cubs. In general we can conclude that real essences are not used in our normal judgements about change, but that no serious discussion that tries to explain our judgements (or intuitions, or prejudices) about change can fail to make reference to essences.

42. Explaining conceptualisation

The third mode of relating to entities on our list was 'filing and conceptualising'. To suggest that these are separate activities from picking out and tracking might imply innocence of heated modern debates about the degree to which 'given' experience is conceptualised, but they are still activities on which we can focus, no matter how integral to experience they may be. To even list filing and conceptualisation as distinct may also seem surprising, but this draws on the 'mental files' model mentioned earlier (expounded in Recanati 2012). The idea is that if I say to you 'I am going to tell you about Blahongs' (of which you have never heard), you will open a mental file labelled 'Blahong', and await further information, and filing has taken place in the absence of concepts. If I ask you what the word is for a socialist community in Israel, I may be considered to have grasped the concept, but lost the filing label. Since the process of filing is of the utmost simplicity (in the 'Blahong' example) then essentialism is not relevant, but the case of how we conceptualise something is trickier. Recanati makes no mention of essentialism in his account of how mental files work, although we have mentioned a few of the structural features that he proposes. However, the work of Gelman suggests that this may be an omission which students of mental files should consider, since children seem to conceptualise things in terms of their kind, and of their hidden mechanisms, which presumably feature in any file quite early in its development (just as you might press to know the ‘essential nature’ of the putative ‘Blahong’ before your curiosity was satisfied). Conceptualising is part of our drive to understand the world, and is the first step towards explanation, so there is considerable scope for researchers into concept formation to examine whether (or not) concepts develop along essentialist lines.

43. Explaining classification

Classification and definition (our fourth mode) both seem to involve more comprehensive levels of thought concerning the item which has been picked out, tracked, filed and conceptualised. Sortal essentialists tell us that individuating and classifying are closely related activities (though Wiggins allows cross-classification to occur after the more precise individuation (1980:201-4)), but we take the discovery of the most appropriate and revealing ‘substance-concept’ employed in fixing an entity to be a sophisticated secondary activity (even if the results are now ossified in language and culture). A commitment to the reality of ‘categories’ of existence, and a taxonomic system, usually involves a commitment to the thesis that every component of nature (and even of thought) belongs in some determinate category, and that there are precise and real criteria for admission to a category (see Westerhoff 2005:I-§1 for six proposed and rival systems). As a corrective, Wiggins cites a gloriously fanciful classification system from a story by Borges, but observes that such a system seems to be wrong because it explains nothing (1980:144 n18). We might wonder how much explanation is ever to be found in classification, but successful taxonomies help us to cope with reality. If there is a real category for every entity, with criteria for each category, then real sortal essentialism would come into its own, since these criteria would occupy a foundational place in our grasp of the world. However, such optimism about categories finds little support, either in theory or in practice. Anthropologists, for example, find a huge range of specific categorisations across diverse cultures, and when cross-cultural uniformity is identified it is only at a highly general level (Ellen 1996). Anthropological studies of classification have generally abandoned the notion of ‘tests’ or ‘criteria’ for categories, in favour of ‘prototypes’ (Ellen p.4), but that implies that actual categorisation is a much looser and more flexible activity than might have been possible with precise (and increasingly refined) criteria. Among the philosophers, nominalists say that only the particulars exist, and so we can label them in any way we please (though pragmatic factors will favour some obvious systems; Ellen notes, for example, that we struggle with very large things like the sea, and tend to cut them up into smaller parts (p.33)). Nietzsche condemns the obsession with categories, writing of the a priori categorisations of Kant that ‘philosophers, in particular, have the greatest difficulty in freeing themselves from the belief that the basic concepts and categories of reason belong without further ado to the realm of metaphysical certainties’ (*Late Notebooks* 6[13]). Westerhoff rejects any thoughts of categorization by essence with the remark that ‘what ontological category a thing belongs to is not dependent on its inner nature, but dependent on what other things there are in the world, and this is a contingent matter’ (2005:218), and he expresses the overall finding of his enquiry thus: ‘my conclusion is that categories are relativistic, used for systematization, and that it is not an intrinsic feature of an object to belong to a category’ (2005:9).

To adjudicate between the extremes of category realism and category relativism, we can consider how classification fits within the practices of explanation. Consider the categorisation of a seagull as a 'bird', a 'fledgling', and a 'pest' (without entering into the more precise science of the matter). These everyday categorisations are entirely appropriate in different contexts, and are employed when people have varied concerns; birds fly over us, fledglings are vulnerable, and pests must be dealt with. In each context some important feature invites the category, such as flying, surviving, and interfering with human activity.

These features will hardly qualify as 'essences', but the concept of an 'important feature' points in that direction. The relativist will rightly reply that 'important' is a concept amenable to Borgesian whimsy, since a person can decide that a random speck of dust is important. We can narrow the field by considering what is 'widely agreed' to be important, but that will necessarily be grounded in human concerns. Even the boldest metaphysician is unlikely to defend 'absolute importance', but any objective notion of importance must probably appeal to the structure of reality, with certain things playing a more pivotal role than others in the structure. We have suggested that if we accept dependence relations and stratified levels to reality, a plausible notion of relative importance within the whole of reality seems to emerge. Insofar as such a case can be made, there is a view of classification which will track the importance, and in that context we will find that the practices of classifying and of essentialising are closely allied.

44. Explaining definition

We have seen that for Aristotle the relation between definition and essence is so close as to approach identity. He offered a procedure of 'division' which became increasingly fine-grained as it approached the definiendum, but was confined to generic features at the last stage by the inescapable need for the account [*logos*] to employ universals. We also noted that Aristotle's definitions can be quite extensive, something like a scientific monograph, and their aim is to spell out 'what it is', or the essential nature of the thing. We saw earlier that Lowe, following Fine, was inclined to say that an essence *is* a definition, but Aristotle firmly rejects this view at *PA* 90b17, and the idea that an essence actually is a set of sentences seems a very long way from the Aristotelian approach. If the *psuché* is the form or essence of a human being, it seems misguided to identify the human soul with a set of sentences. The correct account seems to be that the belief that something has an essence is a prerequisite for an attempt at defining it.

Aristotle thought that there existed a unique definition for each entity, but Leibniz knew that varied definitions of a thing were inevitable, though he hoped for eventual convergence on an ideal. Nowadays definitions are varied in purpose, starting from the distinction between extensional and intensional definitions, which tracks back to Frege's separation of reference from sense. That is, you can define in order to pick out what the definiendum refers to, or to say what it means (by descriptions, or properties). If the intensional definition is sufficiently precise it will narrow the resulting extension down to a single item, thereby fixing the reference (the discredited 'descriptive theory' of reference). The aspiration to define individual entities has subsequently become unfashionable, and definition is now of most interest for stipulating the natures of new entities in logic and mathematics. This need not be the end of the Greek dream of defining the central concepts of philosophy, but few people imagine that each of those concepts has a unique definition. Such definitions seem inescapably contextual and pragmatic, and dependent on language, culture and accompanying metaphysical scheme; Gupta, for example, observes that 'some definitions aim at precision, others at fairness, or at accuracy, or at clarity, or at fecundity' (2008:2). If that were the case we would have to concede that not all definitions aim to give the essence, since Gupta's list is too varied, but there is no escaping the aim of specifying what is 'important' (for some context), and we have offered the important as the first step on the road to the essential. We may plausibly conclude that although the definition is certainly not the essence itself, there would be no practice of definition if we did not possess the concept of essence, and the aspiration to understand it. If we take the more comprehensive view of definition employed by Aristotle, there is even a temptation to identity a good definition with a successful explanation, but that is probably a step too far. We have seen that explanations are fairly diverse and draw on wide-ranging models of a whole level of reality in ways that vary with context. They will also involve accidental features of the objects involved, and not just intrinsic natures. Better to say that a comprehensive definition specifies the core of a successful explanation, which is precisely what we are taking essences to be. The definition of a tsunami will give us the essential features of its generation and effects, but will not give the transient details of each individual tsunami, or events that only have remote connections to such things.

45. Explaining extrapolations

We have examined the role of explanation and essence in individuation, tracking, filing, conceptualising, classifying and defining. Our fifth and final mode of comprehension (extrapolating, judging, imagining) is the widest one We mention 'extrapolating' because essentialism relates interestingly to inductive thought, we mention 'judging' because we are interested in what is true, especially in scientific contexts, and we mention 'imagining' because we need to look at the role of essence in counterfactual situations. Earlier we concluded that inductive generalisations do not in themselves explain, especially if induction is seen as the mere accumulation of instances, but that if induction aims at coherence (rather than increased probability) then it will approach explanation. The question for induction is not whether there is 'more of the same', but in what 'respects' things resemble, and what is 'typical'. Thus Bonjour writes that 'inductive explanations must be conceived of as something stronger than mere Humean constant conjunction; …anything less than this will not explain why the inductive evidence occurred in the first place' (1998:214). The Humean problem of induction rests on regarding reality as a mosaic of unconnected events, containing regular patterns, which are denoted as 'laws'. Induction is then the poorly motivated activity of leafing through the events, hoping to spot the patterns. If we aim to explain, rather than to describe, then we are motivated to dig into the pattern, to see where it comes from. As Lipton points out, 'one of the problems of the extrapolation and instantial models of confirmation is that they do not cover vertical inferences, where we infer from what we observe to something at a different level that is often unobservable' (2004:66).

At a lower level we might hope to find laws, necessities or essences to explain the observed patterns. The Humean view, not lightly dismissed, is that you might indeed move to a ‘different level’, and successfully find laws, but that these laws will just be further patterns within the structure. We have already rejected the idea that mere statement of a law can offer illuminating explanations, apart from revealing connections across reality. If that is the best we can manage, then of course laws are very welcome. If we could find necessity in the laws, that would be a step closer to satisfying explanations, but orthodox Humeans reject necessity in laws, since the laws derive their contingency from the presumed contingency of the pattern which they describe. Empiricists are typically distrustful of anything other than conventional necessities. The most satisfying explanation would give us both the real necessities (if they exist), and the reason for those necessities.

It would be naïve to leap to the simple conclusion that there are discoverable essences which give rise to necessities, and the practices of science do not encourage the direct quest for essences. One shortcut to the direct inference of necessities from essences is captured in the remark that ‘there is indeed natural uniformity in the negative charge of electrons, but the reason for this is that it is an essential property of being an electron that something be negatively charged; it would not be an electron otherwise’ (Mumford/Lill Anjum 2011:142). This seems true enough, but tells us more about our words than it does about electrons, and we can take it that the mere semantics of natural kind terms will not be sufficient to offer essences which can truly explain the world (as Nathan Salmon has persuasively argued, in 1980/2005). We can be fairly sure that electrons don’t have a uniform charge simply because we have called them ‘electrons’, and then stipulated the charge that is necessarily required for our concept.

Nevertheless, the process of inference to the best explanation drives us to dig deeper beneath regularities, and seek for stable sources from which phenomena derive. It is well observed by Gelman that ‘inductive success is rewarded with more induction’ (2005:316), which we can connect to Russell’s claim that the concept of a ‘dog’ is an inductive generalisation, bringing a comprehensive file of canine information to bear on any discussion of dogs. Further inductions and explanations only flow if you discover what is of central importance, and you get it right, and we can label this as the ‘essence’, without specifying what such an essence will look like. Indeed, as we observed earlier, the best criterion for ‘importance’ seems to be explanatory fecundity, which is closely allied to our modes of classification.

46. Explaining natural kinds

The concept of a ‘natural kind’ stands somewhere between the concepts of law and of essence. On the one hand it is said that laws rest on natural kinds, since universal laws of nature must refer to unchanging types of objects and properties. On the other hand, one approach to Aristotle sees essences as being exclusively ‘kind essence’ – the set of features which fix a kind of thing - so that essences also rest on the concept of a kind. An objection to the necessary connection between laws and natural kinds is voiced by Chakravarty, who writes that ‘causal laws often do not make reference to kinds of objects at all, but rather summarize relations between quantitative, causally efficacious properties of objects’ (2012:3), to which Ellis offers a reply that properties and relations can themselves be considered as natural kinds (2005:90). We must start by asking what is meant by a ‘natural kind’, and rather than jumping to the idea that they are the blueprints upon which nature is constructed (the concepts of ‘moulds’ denigrated by Locke (*Ess* 3.3.17)), we should start with the observed aspect of nature which gives rise to the idea, and that seems to be stability. The earliest and best definition of a natural kind is in the *Upanishad* ‘Chandogya’, and tells us that ‘by knowing one lump of clay, all things made of clay are known; by knowing a nugget of gold, all things made of gold are known’. Aristotle expresses a similar thought about water (Top 103a20), and Peirce adds an interesting foil in making the same point: ‘the guiding principle is that what is true of one piece of copper is true of another; such a guiding principle with regard to copper would be much safer than with regard to many other substances - brass, for example’ (1877:8). Modern chemists probably would agree about gold and copper, express slight caution about water, and probably reject clay, while brass, as man-made, self-evidently does not have sufficient regularity for the role. Aristotle took animal species for granted as natural kinds, but evolutionary theory has undermined our belief in their long-term stability; hence they were rejected as natural kinds in recent times, but a rethink is currently occurring, as the concept of a ‘natural kind’ has come up for reassessment. One view is that natural kinds have perfect stability, and we might conclude that some types of matter such as the natural elements will therefore qualify, but we then realise that some of them undergo radioactive decay, and some theories suggest that even protons may decay after 10^{36} years (*Wikipedia*, ‘Proton’). If something endures for 10^{36} years and then decays, that is exceptionally stable, but it is not permanent, and yet if the proton is not a natural kind then very little else will qualify in that way. Permanence does not seem to be a reasonable criterion, but we can say that when elements or particles decay they cease to be that kind of thing, so they may have a perfect stability without permanence.

If we base natural kinds on the ‘Upanishads Test’ – that any random sample will suffice for an intensive study of the kind – this will define natural kinds relative to human epistemic capacities, which is not very objective, and it may also drastically narrow the number of kinds, since each tiger is slightly different, and trees vary enormously. Tin is an element with 21 isotopes (10 of them stable), so we must choose whether tin is one natural kind or 21 natural kinds. Tin is used to make the alloy we call ‘solder’, and the distinction between isotopes is not important in that respect, so that for solder-makers all tin will pass the Upanishads Test, but for chemists it will fail. Ordinary people may have low standards for the Test (so that even clay may pass it in a sculpture studio), but cutting edge science seems to demand total uniformity of natural kinds. But the idea of total uniformity may be an illusion, since it requires all tin-16 atoms to be totally indistinguishable, and yet we know that their electron shells continually fluctuate in energy levels, and perfect type-identity between atoms is impossible. Hence the concept of a natural kind has to relax to be of any use (in our inductions, for example) – which means that animal species might be acceptable natural kinds after all.

Testing one sample enables us to ‘know’ all the other samples, but ‘know’ is a loose term here, best understood as growing in strength as inductive inferences increase in extent and coherence. A natural kind offers an important step in the inductive process (and Koslicki offers inductive fruitfulness as a distinguishing feature of a natural kind (2008:204)). We learn that tigers are all fairly similar, and then, given the slowness of evolution, we can gain a fair mastery of the tiger species from the study of one standard-looking tiger (by cataloguing its genome, for example), without worrying what its ancestors were like. This is even more true of the proton. This means that we can allow essences to have a modest transitory aspect, without giving up essentialism, and we can allow tigers to have the sort of essence that many consider necessary for a natural kind. If we connect 'knowing' tigers and protons with explaining them (since a good explanation scores quite high as a justification), then we can say that a prime qualification for being a natural kind is staying still long enough to support an explanation which can be generalised across related instances of a thing. This won't tell you what 'related' means, but there is a potential induction leading from one instance, to something which qualifies as the 'essence' of that instance, to a general entry qualification for the kind. Again, though, we mustn't expect essences to solve our problems, because we will find ourselves hesitating over whether to admit some feature into this 'qualification', when the only criterion we can apply concerns the boundaries of the kind which will result. That is, our concept of the kind will dictate our choice of essence, rather than the other way around. We come back to explaining the kind that has been designated, and find no escape from a certain degree of conventionalism in the choice of kind designation we have made.

The fact seems to be that human beings divide reality into chunks in order to grasp it. As Devitt puts it, 'our explanatory purposes in introducing a name for a species demand that we draw the lines around a group that is small enough to share a whole lot of important properties and large enough to yield broad generalisations' (2008:243). We prefer the divisions to match the apparent 'joints', such as (for species) breeding groups, or ecological niches, or branches in the evolutionary tree of life, but the real situation is revealed when we are faced with a continuum where we can't find any 'joints'. Ellen offered the nice example of the sea. The British Isles are surrounded by continuous water, yet Britons are familiar with The Channel, the Solent, the Irish Sea, the North Sea, the Wash, and so on. Darwinian evolution has shocked us by revealing that the flow of life which was assumed to have nice 'species joints' is actually continuous over long periods of time. If we time-travelled back to a remote period, we could doubtless name the species we observed, and yet our palaeontological books might reveal to us that what we had named was understood today as a 'transition' between species. Early hominids acquire names as if they were neatly divided groups, but the underlying assumption is a continuum. If we hold strongly to the clear speciation of the life forms existing today, then Dupré points out that while animals exist in fairly distinct groups, vegetable life is far from clearly distinguished, and apparent species flow into one another across continents (1995:23). If speciation across time were not a continuum, we would have to assume that two parents could have an offspring that was a different species, and none of us is likely to accept that our own children could be of a different species. This fact does not stop us from naming species, though, if we rely on the Upanishads Test, and see that extensive knowledge and explanations can be had of the transitional species we choose to pick out. If any slice of reality is amenable to a coherent explanation that results in extensive and stable generalisations, we can reasonably say that the slice can be thought of as having an essence. Coherence in our explanations will drive us towards coherence in our slicing.

We do not need to follow Dupré into a fairly anarchic 'pluralism' about natural kinds. His main thesis about kinds is given in the remark that 'the question of which natural kind a thing belongs to ....can be answered only in relation to some specification of the goal underlying the intent to classify the object' (1995:5), and a typical example he gives is that cooks make an important distinction between garlic and onions, but this is of no interest to scientific taxonomists (1995:34). The implication is that the concept of a natural kind has no basis at all in nature, and is merely the result of human interests, but that seems to be a misrepresentation of his own examples. Cooks do not disagree with taxonomists, and know perfectly well that taste is no basis for cataloguing the structures of nature. Taste tells us nothing about the reproduction, ecological niche, genealogy, or inner structure of the onion genus (allium). The best divisions of nature are the most explanatory divisions, and explanations require core features (essences) to do the explaining. Since good explanations are true, they must reflect external reality, and not our culinary preferences. Dupré rejects essences for natural kinds, mainly on the inevitable grounds that he doesn't believe in the kinds, but the rival view espoused by Devitt seems much more plausible – that the classifying into kinds is a quite separate activity from the study of essences, summarised in his remark that 'essentialism is concerned with the nature of a group, whatever the category it falls under' (2008:228). The sensible procedure seems to be the imposition of kind-divisions at the points where the underlying natures exhibit very distinct change, but this won't stop us from dividing the sea around Britain into sections (or dividing garlic from onions), despite all of the underlying seawater passing the Upanishad Test. Thus we see that essences have the further role in our thinking (in addition to tracking, classifying, defining etc.) of providing the focus for explanatory understanding of distinct groupings of entities in nature, even if these groupings change over time, and even if we squabble over how to classify them.

47. Explaining modal reference

Next, we will consider whether an entity must have an essence in order to make sense of reference to it in modal contexts. That is, if I say 'I might have been in Paris right now', most views of metaphysics will allow that to be both meaningful and true (because no intrinsic changes to me are entailed by the possibility), and if I say 'I might have been sitting in the centre of the Sun right now' that is taken to be meaningful and false (because I would entirely vanish). The tricky cases are where the possibility considered refers to some variant of me, such as that I might have been taller, or a frog, or a dustbin. If we accept Kripke's view that such a claim is always meaningful because reference to myself is stipulated (1980:44), the assessment of truth-value is still problematic, because the claim that I might have been a frog seems to be false precisely because if the subject of the sentence were a frog that would entail that it wasn't me. I have said that I am referring to myself, and yet I appear to be failing in the attempt. A claim that I might not be myself seems to be a breach of the most basic rule of identity. But why does the possibility of being a frog preclude the possibility of being me? Two bold options are to say that the only 'me' is the actual one, so that all modal talk refers to 'counterparts' (who might differ greatly from me), or to embrace ‘haecceitism’ and allow that something might still remain the same no matter how different its appearance became. Neither solution seems appealing, since if a magician threatened to turn me into a frog that wouldn’t bother me if the result was my mere counterpart (and thereby not me), and the haecceitist approach allows that my pen may actually be Julius Caesar (despite all appearances), which undermines our capacity for any kind of sensible talk about normal objects. Philosophers (such as Lewis, Forbes or Mackie) may opt for surprising solutions to this problem, but for ordinary thinkers only essentialism is plausible. Counterfactual claims about the possibilities available to me become impossible at the point where I am no longer ‘essentially’ the subject of the sentence. So when is that? Our concept of essence will need to make a contribution to this modal dilemma.

We have considered physical objects, but what about abstract things? The problems of modal variants or counterparts seem inapplicable, since there seems to be no possible world in which this triangle might have been a circle. The standard view of abstract entities is that they exist in systems of interconnected necessities (perhaps because they are analytic, or because they reveal timeless truths), so that the idea of relative ‘importance’ or ‘dependence’ (with a ‘direction’) seems unable to get a purchase, and so essentialism has nothing to contribute. However, the concept of ‘determination’ seemed quite comprehensible in such a context, as when we say that complex truths are determined by their simple components, or theorems are determined by definitions, or whole theories are determined by axioms. In outlining an approach to explanation, we proposed that explanation rests on a view of reality as structured into ‘levels’, with the unanalysed modular components at the bottom of a level ‘grounding’ the structures of that level, with each level grounding the one above it, and that we should keep essential explanation within a level distinct from foundational explanation. The proposed picture is that foundational explanations invoke the ‘bottom’ of the hierarchy (though that may be a forlorn hope), while essentialist explanations invoke the bottom of each level. We will examine below the possibilities for such a view for abstract systems, but for now we can say that the role of an essence must arise from whatever ‘levels’ can be discerned within an abstract system, and that such essences will need to be more than mere ‘atoms’ in the system, since they must have the power to explain. For example, the concept of a ‘point’ in geometry may well be foundational, but no explanations seem to flow from the mere contemplation of an atomic point. A quest for an essence among abstracta would hope to find convergence on small unified conceptions, built from the raw conceptual atoms, where these conceptions were taken to determine an array of more complex abstract facts. We might say that essences are the focal hubs in networks of explanation.

48. Role of essence

We can now summarise our findings about explanatory essences. It seems unlikely that essence can give a criterion of unity, but perceived unity may require an essentialist explanation. Given that individuation by means of a sortal concept, perceived as an essence, was unsatisfactory, and that individuation of physical objects seemed better understood in empirical terms, with further information emerging from inductive inference and classification, the task of individuation seemed irrelevant to the concept of essence we are studying. In the next chapter we will consider the relationship between unity, individuation and counting, to see whether explanation in metaphysics encourages essentialist thinking. Tracking an individuated object through space and time also did not seem to require any essentialist thinking, since more superficial criteria would do the job. Tracking through intrinsic change seemed closely related to the question of unity, and again it didn't seem that essences can offer a criterion that would distinguish substantial change from alteration (since circularity always threatens), but explanation of perceived degrees of change seemed inescapably essentialist, since the very nature of the entity must always be the central issue. The first step in thinking about an individuated object appeared to be no more than the labelling of a mental file, with no commitment to contents, but the steps towards conceptualising the contents of the file were another matter. Children build their concepts around essence, and to fully grasp a concept, it is best to seek its essence.

Classification seems to be a higher order activity than the mere apprehension of an object. While pragmatic and even whimsical classifications are a fact of life, there seemed to be a more realist notion of classifying by what is 'important' within a natural structure and this is exactly the context where we are claiming that essences have a role. The Aristotelian concept of a definition is the expression of a *logos* for an essence, which could be equated with the 'core' of an explanation (ignoring accidental features of particulars). While some might equate definitions with essences, it seems much better to say that the conceptualisation of an essence is the necessary precondition for definition. Successful definition appears to be the main test for whether something has been picked out which can justify the label of 'essence'. Hence we do not require essences to fit the definition process, because the dependence goes the other way, and definition is not one of the constraints on our concept of essence.

Induction, we suggested, should not be the mere compilation of a list of unconnected observations, but a quest for what is important, typical, coherent and explanatory. The best summary of thorough inductive enquiry is that it aims to identify essences – where essences are understood as those underpinnings of a hierarchical model of reality which seem to explain it best. In this account, essences are the target of induction, and may even illuminate by revealing necessities. The concept of a natural kind connects essences to the idea of general laws of nature, but it seemed unwise to assume perfect stability and perfect uniformity in each natural kind. Resting on the idea of a good explanation seemed to provide the concept of natural kind we require, as illustrated by the fact that we can hope for a comprehensive explanation of tigers, without a commitment to either the permanence of the tiger species, or perfect type-identity between individual tigers. Much higher degrees of stability and uniformity (in the proton, for example) will take us much closer to universal laws, but short of the implication that the laws are necessary. There is an unavoidable contextual aspect to the selection of our natural kinds, but without the concept of an essential nature for each one, it is hard to see how the concept of a natural kind would be possible.

On the question of modal truths, neither stipulation, nor counterparts, nor haecceitism, nor mere necessary properties seem to suffice as criteria for what 'I' might possibly become, and only some sort of explanatory account (such as the grounding of my personhood, whatever we take that to be) will do the job. In this instance the role of essences is to perform the whole of this task, as no other concept seems suited for the role. Without some version of essentialism, we have no concept at all of when changes to an entity become its termination. This conclusion may seem at odds with the conclusion that essentialism does not provide a criterion for normal change (though we decided it might provide an explanation). It is best to say that small changes do not require an essentialist criterion, since the persistence of the underlying substance can be taken for granted; it is only when we try to explain why small changes don't matter much that our talk becomes essentialist. When, however, we consider the more extreme (or even absurd) changes entertained in the literature on modal identity, then recourse to talk of essentialist criteria is the only strategy available. In both cases the story is essentialist, but in the normal easier cases enquiry into essences is overkill.

Hence we must now ask how an essence should be understood, given that it is to have a role in our first-order practices of conceptualisation, classification, ambitious induction, grasp of natural kinds, and modal judgements in extreme cases, and in our second-order practices of explaining the first-order practices of unification, tracking through change, and definition. In each case we can point to the line of approach implied by explanatory essentialism.

49. Essence and concepts

Concepts seem to have two aspects - connections both to the world and to other thoughts - and theorists often divide over which of the two aspects has priority. Fodor has championed their role in representing the world as prior (1998), whereas Peacocke sees them as individuated by their role in reasoning (1992). Since the purpose of concepts is to bridge the gap, there seems little point to any disagreements here. We can take Fodor as reliable when he writes that 'I don't know how concepts are acquired. Nor do you. Nor does anybody else' (2008:146), and neuroscience has yet to deliver a clear picture. However, any account which individuates a concept (or most other things) by its role or function will tend to present it as a 'black box', a location in a flow diagram, with no explanation of how it fulfils its role. If we enquire what needs to be the 'essence' of any concept, we can say little more than that it must involve the concept's 'bridging' ability, of connecting world to thought, and for that 'representation' seems to be a promising term, because it offers a mechanism which can do the job. It seems wrong to connect concepts entirely to language, since animals and infants can evidently think (and Gelman says infants have pre-linguistic categories (2005:179)). It illustrates the essentialist view that to enquire about the nature of concepts leads to what is essential to them (to fulfil their role), and that their representational capacity is a candidate for the answer. An inviting thought here is that the essence of a concept is whatever explains its role, and that in general essences are what explain roles and functions.

If we seek the essence of some first-order concept, such as that of a CAT, then mental files provide a good theoretical picture. The general picture is given by Fodor, who writes that 'we think in file names, and file names are Janus-faced: one face turned towards thinking and the other face turned towards what is thought about' (2008:100). Recanati tells us that he wants 'mental files (properly speaking) to serve as individual concepts, i.e. thought constituents'

(2012:64), and that ‘a mental file plays the role which Fregean theory assigns to modes of presentation’ (2012:221). The Fregean view of concepts makes their function central (in that they output an extension of appropriate objects), and we can approach explanation of this function through mental files. Recanati’s files function in reference, and in the provision of information. Their main components are the *label*, the *nucleus* and the *periphery*, and the main types of file are *demonstrative*, *recognitional*, *encyclopaedic* and *indexed*. The foundation of the system is simple *proto-files*, with a rising hierarchy of orders. Without exploring the niceties of this picture, we can see that the whole story is explanatory in intention, and that the file for ‘cat’ will be hypothesised as explanatory of our relation to actual cats, and our modes of thinking about cats. We must then agree with Lowe that ‘things must have an essence, in the sense of 'what it is to be the individual of that kind', or it would make no sense to say we can talk or think comprehendingly about things at all. If we don't know what it is, how can we think about it?’ (2008:35). If we speculate or introspect about the contents of our personal ‘cat’ file, we do not (in my experience) immediately confront ‘essence of cat’, but the activities of thought and speech seem to require the notion of a basic ‘nature of cats’ to which Lowe refers, and that must be the lynchpin of a substantial ‘cat’ file. Further theorising has emphasised the role of ‘prototype’ cats and ‘typical’ cats in this story, and that seems preferable to the ‘theory theory’ of concepts, which makes their role paramount (Margolis and Laurence 2009 offer a survey).

50. Essence and classification

We have rejected sortal concepts as simultaneously achieving ‘individuation’ and giving us a thing’s essence. That approach places the kind before the individual, and fits some normal experience, but without giving the whole story. It may be correct that the whole ‘cat file’ is opened as soon as we spot a cat, but archaeologists digging up unusual objects, or explorers in the jungle, take their time over the allocation of sortals. Classifying only seems possible after a number of other preliminaries. While classification can follow any principle we choose (group by price, size, colour, proximity, beauty….), there seemed to be a ‘natural’ mode of classification which described the consensus on classification in most of the familiar cultures, and rested on explanatory importance, judged by role in natural dependence structures. Given that this makes ‘explanation’ the target of a culture’s classifications, does this lead us to classification by essence, or for essences to have a prominent role? Not all explanations are essentialist, and certainly not all classifications are essentialist, because we are free to classify by what is trivial or superficial. Westerhoff writes that ‘systems of ontological categories are systematizations of our intuitions about generality, intersubstitutability, and identity’ (2005:55), and each of these three aims implies a structure to the concepts (of nesting, embracing and overlapping), which presumably aims to map the structure of reality. Within such a tangle of classifications, though, it is not clear that essences are what launch or underpin or fix the structure. Locke’s ‘nominal essences’ seem to play a major role in everyday classification, and these are not essences at all (in our present usage), because they are superficial.

Nevertheless, the aspiration of both normal thought and of scientific rigour is that nominal and real essences should coincide – that is, that everyone wants to get the classifications right. The most revealing cases are re-classifications that occur in the light of new knowledge, such as the general view that whales are not actually fish (though Dupré, in his attack on ‘natural’ classifications, nicely challenges that revision (1993:30)). We can say, in general, that deeper knowledge of the nature of any entity is always likely to raise the question of reclassification. This may imply that essentialism is driving the discussion, but the reclassification may be responding to ‘relations’ (such as genealogy, or ecological niche) rather than intrinsic ‘natures’. There is no consensus about classification, and essentialism is one approach, though arguably the best, given that more is explained by the nature of a tiger than by its place in a genealogical tree. Devitt’s view that the essence of a group is distinct from its category implies that classification has little connection to the essentialism he defends, but one could defend the essentialist approach to classification, whatever practices actually dominate the field.

If we attempt classification by essence, what would an essence then have to be? Presumably we are rejecting superficial resemblances in favour of ‘deeper’ resemblances, but issues of context and pragmatics still arise. In the case of tin we were faced with either one essence or 21 essences, with the isotope level being the ‘deepest’ we could hope for, and yet it seemed that this classification depended on what needed to be explained. But this is not encouraging for the plan of discovering essences and then classifying accordingly, and we are also faced with massive ignorance about the essences of many things that need immediate classification. Issues of classification, we may conclude, are closely related to essentialism, but it throws little light on what we should take an essence to be, since many worthy classifications are not guided by essence. We will, though, examine the development of the periodic table of elements in the next chapter, to see whether essentialist considerations seemed inescapable in that famous act of classification.

51. Essence for induction

Scientific induction seems boring if it just maps patterns of regularity, and exciting if it digs down to explanations. If the rewards of the ‘boring’ strategy are laws of nature which offer predictions and generalised connections (not unwelcome, of course), but the ‘exciting’ strategy offers essences which explain the regularities, what would such essences then consist of? If we return to the example of smoking and lung cancer, the boring strategy has saved millions of lives by reliably predicting likely death from too much smoking, but the exciting strategy is closing in on the chemistry that causes (or even necessitates) what has been predicted. The obvious word which philosophers have embraced here is ‘mechanism’, and we might say that essences are simply explanatory mechanisms. For biology this seems right, given the complexity of biological structures, and their active nature, since biological discovery largely consists of simply showing ‘how it works’. However, we saw reasons to think that many explanations might better be seen as ‘models’ of the situation (as in the car crash example), where the whole causal sequence and structure is laid open for inspection, with no particular part of it counting as a ‘mechanism’, unless that word is used somewhat metaphorically. In the car crash example, we only seem to think in an essentialist way when we pick out some aspect of the event as having prime importance. An enquiry by a tyre manufacturer, for example, considers contact between road and tyre to be essential, where the police may consider intoxication of the driver to be of the essence. Scientific essentialism may seek to avoid the contextual character of its more everyday cousin, but similar problems seem to arise. For example, if we say that an essentialist explanation descends to the bottom of the level where the phenomenon arose (such as the geology of tsunamis), this fixes a lower limit within the natural structure for an essence, but it will not give an upper limit, since extending to the ceiling of the level would seem too extensive to count as an 'essence'. There is subjectivity in the 'importance' of what is picked out, and the power and scope of the explanation that is accepted, and the interests of the enquirer will control those issues. Essentialist explanation of a tsunami must focus on the violent rift in the sea floor, and takes for granted the chemistry and physics of rock strata, but should it also take for granted the transmissive capacities of sea water, or the distance from the coastline, or the height of coastal walls? There is no criterion available here, and we can say no more than that an essence will be a low-level determiner of the object of enquiry, taken to be as minimal or as comprehensive as the context requires. Essences are no less real for all that, but we must acknowledge that there are loose aspects to the concept of 'essence'.

52. Essence and natural kind

If natural classification focuses on 'natural kinds', might this get us closer to a clear concept of an essence than the mere practice of classification can? For example, we might say that 'tin' is a classification, suitable for manufacturers of solder, but that only an isotope of tin is a true natural kind. This will be because a tin isotope can pass the Upanishads Test to the limits of our ability, offering no discernible difference between samples. We might then ask how we should treat a case of two samples in which there was a real underlying difference, but conditions placed the difference forever beyond the grasp of humanity. That is, are natural kinds understood realistically, or operationally? If we opt for realism, we may not actually know what the natural kinds are, because various underlying differences may be hidden, never to be discovered. The operational view, however, allows us to be as strict or lax as we like about the demarcations of the kinds, and similar laxity will apply to the essence of the kind. The realist view seems in tune with the essentialist approach, and if we wanted exciting induction to explain regularities, then we should look for exciting essentialism about kinds, and this will not only define their nature and what is stable about them, but also explain their stability. The stability of natural kinds was exactly the reason why Albert the Great believed in substantial forms. This is an area where essentialism must become 'scientific' in character, because the stability of metals is unlikely to be explained by metaphysicians. We may persuade scientists to identify every substance which can fully pass the Upanishads Test, and label them as 'natural kinds', and presumably treat what each set of samples has in common as its essence. This would apparently make the essence into a list of properties considered necessary for the existence of the kind, and mere lists of properties seem insufficient for the explanatory task required of essences. If we ask for an explanation of the stability of the kind, that seems to go 'deeper' than the observable properties, suggesting foundational mechanisms and primitive powers. If philosophers cannot prejudge what scientists will offer as the basis of each stable kind, they also cannot prejudge what the essence might look like, and again the only certainty about essences for natural kinds is that they must occupy a fundamental role in a certain level of explanation.

53. Essence and possibilities

We saw that the only approach offering any sort of criterion for deciding the truth or falsity of ‘I might have been a frog’ is essentialism. If we take essences to be sortal in character, that means that I couldn’t have been a frog because I would have lost my essence, which is being human (my ‘substance-concept’), but it does mean that I could have been be a small Chinese girl, I could have lived 30,000 years ago (though perhaps not a million years ago), and I could have been the person correctly designated as ‘the most unusual human who ever lived’. This gives further grounds for doubting the sortal approach (if I thought my sex, my origin, or my normality were as essential to me as my humanity). If I take essences to be sets of necessary properties then the frog will retain the trivial ones which the frog and I share (of existence, or self-identity), but the only criterion available for deciding on the more intrinsic and personal properties will be imaginative or intuitive assessments of which deprivations will bring my being to an end, which seems to require assessment of the identity or non-identity to precede evaluation of the properties, instead of following from it (since we must first determine my ‘end’). If there is to be a criterion for deciding which features can and cannot be lost, there has to be a notion of the role which these features are playing. But then the only relevant role we can find is performing the task of making sure that I am still me, which is the problem we were trying to solve, and presupposes a ‘me’ to be evaluated.

We might say that the essential features are those that explain who-I-am, which we can take to refer to who I am as an individual, rather than to the various ways I am classified. The literature on personal identity offers an array of theories for who an individual person is (such as a pure ego, a psychological continuity, a bundle of mental events, a human animal, and so on). For each theory there will be an essentialist account of how this personhood is possible (such as a mental substance, a system for remembering, a space where mental events meet, or a human body). The essence is what explains the personhood, and that explanans is what must be retained if I am to remain the subject of the speculation about frogs. This story is loose, and isn’t the ‘criterion’ we want, but it still feels right, and matches the intuitions which ordinary people have about when someone (in extreme old age, perhaps) has lost identity with their former selves. The prospect of finding an essence of me, and using it to decide the point at which I cease to exist as possible versions of me approach froghood, doesn’t look good without a prior decision about whether the essence is ‘mine’, but deciding that some greenish flippered humanoid has ceased to be me, and then offering an explanation for that intuitive judgement, is bound to make use of some account of my essence. The essence is whatever explains our willingness to assert an identity between two objects across counterfactual situations (or across possible worlds).

54. Essence and unity

The criteria which explain an assertion of the identity of a thing across its range of possibilities are likely to be similar to the criteria we offer to explain assertions of its unity in a single situation. Philosophers have not yet produced a criterion for the unity of an entity. No deep general puzzle can ever be a closed case in philosophy, but there must be the suspicion that there is no absolute unity to be had, but merely degrees and modes of unification. Electrons, molecules, crystals, cells, animals, persons, shoals of fish, planets, crowds, a library, sets, and the universe all exhibit various sorts of unity, but it looks more like a family resemblance than a fact of nature. The explanatory principle of essence can still apply, however. If we say that a shoal of fish is a unity (quite reasonably, if seen on a trawler's radar), we can ask for criteria for the unity, and expect to hear of physical closeness among the fish, keeping together while moving, and being of fairly uniform species. This gives a loose notion of the essence of a shoal, and is responsive to rational criticism, and to explanatory success or failure, in discussions of fishing patterns. Many thinkers take a person to be the epitome of unity, and if we try to explain that judgement or intuition, then philosophers can only appeal to the theories of the nature and underpinnings of personhood mentioned above. The unity is derived from the ego, memories, ideas or body of the person. This convergence of essentialist explanations in two distinct (though related) areas of thought is important for the current account, because it is only when what is proposed as essential to something seems to lead to a array of explanations that we can say we are talking of the essences that Aristotle argued for. The extent to which such explanations of unity are rational or successful is of considerable interest, and will be examined in more detail in the next chapter.

55. Essence and change

In making the Aristotelian distinction between mere alteration and true substantial change (depending on whether the entity remains 'that thing' after the change), it was hard to see how the essence on its own could make the required distinction. For scientists, or other experts, to straightforwardly identify the essence of a thing, and then use the essence to settle how fundamental a change has been, looks a forlorn hope, given the fall from grace of essentialist investigations in science laboratories. To assemble an account or definition of the nature of a thing seems to require some prior knowledge of what the thing is, and after change has been undergone, the essence of the emerging entity seems only discoverable after the assessment has been made of whether this is still the same thing. Hence it seems that essence is not a criterion of change, but figures in the explanation of a change, but only after the event. We might judge an apple to be the same apple after a degree of decay or partial consumption, and then offer (say) the survival of the core and some healthy flesh as the grounds for the judgement. Whether it is actually still the same apple seems an open question (affected contextually by whether it is in a nursery, a shop or a restaurant, where different standards apply), and there is no agreed essence to act as a criterion. Some might consider the survival of pips sufficient to make it the same apple, and then the pips would be the essence. 'Essence' thus remains a perfectly meaningful concept, even when we take a very relaxed view of identity through change, and what counts as the essence may vary.

56. Essence and definition

We proposed that a definition is not an essence, but is better seen as a test for essence. It is pointless attempting a definition of something which is thought to have no essence, and definition presupposes a commitment to an essence, because it is an attempt to state what the essence is. Lexicographers have drilled us into assuming that a definition is rigorous and maximally concise, but a rambling definition full of metaphors could equally well qualify in the present context, if it facilitated the successful grasp of a thing's nature. However, the Aristotelian Method of Division offers the best model for how definitions should proceed (if their target is the grainy facts of the individual, rather than the smoothed out essence of the natural kind). The method is a gradual refinement which closes in on the essence of the definiendum, but for that we need not only a conviction that the target has an essence, but some general appropriate concept of what an essence is. A good definition produces understanding of the nature of something, which implies that it explains that nature. It seems obvious that we have a third notion of explanation here which converges on the same essence on which explanations of unity and of modal identity converged. We might adopt a pluralistic view of essence, with a different concept for each of the focal issues that we have investigated, but the notion that when you fully understand the nature of a thing the explanations then 'flow' into a range of aspects seems to fit normal experience. Family members are far more likely to understand the odd behaviour of a person than strangers are, and experts who understand horses, bread, diseases, and geology can explain most aspects of their speciality (within what is known). Definitions can be whimsical, but good definitions aim at expressing the core of expert understanding (just as a novelists reveal the wellsprings of character motivation). The essence needed for definition is the same essence that is needed to explain unity and continuity. In terms of dependence relations, the definition will pick out the foundations of the target 'structure', the grounding of the qualities and behaviour by which we identify, unite and classify things.

57. Explanation aims at essence

Our findings so far reveal two quite distinct concepts of an essence. On the one hand any object of thought, from the simple and minute to the complex and massive, and from concrete to abstract, must have an essence which makes it available for that thought. That concept of essence has been presented in terms of 'structural levels' to each target object or area of study, thus requiring there to be 'lower' levels as grounding. The essence is the core of that lower level, known by the way in which our modes of apprehension converge on it, and confirmed by the flowing of further explanations from it, and the possibility of definition. This is the realist view of explanatory essences. On the other hand we start from the activity of explanation, and face the fact that most explanations are driven by contextual interests, some personal and others shared across a culture. We attend to the fact that children (and hence people) naturally essentialise what they encounter, but that the essentialising faculty sometimes misleads. We conclude that essences are creations of the human effort to understand, and may reflect the limitations of the human mind (with its need to simplify) more than they reflect the supposed structure of reality.

We may recognise that we all have minds that are limited in this way, but can rise above the limitations to a more objective approach (typically preferring a scientific account of what we label as a 'folk' activity). No scientists wants to give up explanation, but this second view of essences may make them redundant. No one can claim that all explanations are essentialist, since there are quick one-off superficial explanations which work well in their context. So does it serve any purpose to distinguish between the powerful essentialist explanations and the more limited simple explanations? We have suggested that essentialist explanations reach to the bottom of a level of study, but that how high they extend is prey to pragmatic influences. The only criterion might be that the explanation must remain sufficiently low in the level to generate the flow of explanations across a considerable range at the top, but that seems an unreasonable requirement, since we can envisage a perfect essentialist explanation which only explained one thing. If no objective criterion for topping off the essence arises, maybe we just have a vertical continuum of degrees of explanation, and the best account of explanation is the creation of a 'model' (as in the car crash case), ranging across the whole height of the level, and made available to all enquirers, with their varied interests. This offers a Model Theory of explanation, and rejects the Essentialist Theory of explanation.

The Model Theory aims at a complete and accurate description of the nexus of determinations (built from dependencies and causes) within a level. The implication is that actual explanations are selected from the model according to need, the only objectivity being whatever objective truth the whole model can achieve. This seems, however, to abnegate the requirement of priority that is contained in the concept of a determination. If A determines B, then in some sense A is 'prior' to B, and however one understands 'prior', it must lead to A having more explanatory power than B since (irrespective of any Cs and Ds involved) we saw that direction of determination implies direction of explanation, and A will (at least partially) explain B, whereas B cannot (in that same context) explain A. If A picks out some modular feature which spreads across the bottom of a level, then an economical account of the operations of A will ground almost every complexity in the higher structures. In this way, the grounds of a level are 'prior to' and thus explanatorily more important than what is grounded. Since the levels described as 'lower' are those that do most of the grounding work, those levels must have the prime role, which is the minimum requirement for the Essentialist Model.

On the realist view essences exist (in a loosely determined way) low down in a level of reality's structure, identified by convergence, and confirmed by fruitfulness and definitional success. There seems to be an acme for such essences, in which a narrow core is identified which fully explains the whole entity or area of study, to the top of its level. The foundational mechanisms of a highly uniform natural kind, such as a proton, or a gold atom, might qualify in that way, but there is no clear line between explanandum and explanans. Are the main relational capacities of a thing, for example, part of the problem or part of the solution? We cannot use the distinction between evident surface properties and hidden essential ones, since there is no principled reason why the essence should not be quite evident to us (and Oderberg defends the view that essences are largely knowable to us (2007:18)). It is hard to find an objective ground for the distinction, other than in the structure of explanations involved in the case. The intersection of converging and flowing explanations picks out, in the clearest cases, a fairly distinct ‘core’. If it were claimed that the valency of an atom is part of its essence, this might be rebutted by an explanation of the valency in terms of more fundamental features of the atom, such as electron structure, and so the convergence is taken to a lower level, and the division between explanans and explanandum is correspondingly shifted. Without that account of the boundaries of essence in terms of explanation, essentialism is open to the traditional charge of vagueness, as when Richard Cartwright remarks that ‘I see no reason for thinking essentialism unintelligible, but a chief perplexity is the obscurity of the grounds on which ratings of attributes as essential or accidental are to be made’ (1968:158).

In what sense, though, is this a ‘realist’ view of essences, given that while the lower limit of the essence is determined by the lower limit of the level of the activity (and even that may be imprecise), the upper limit of the essence is purely determined by the institutions of explanation, rather than by the facts? The reply must be that it is not ‘purely’ determined in this way, because explanation is not a ‘pure’ activity, involving as it does a meeting of mind and world. Context and interests may dictate which explanations are being sought, but they do not dictate what the successful explanations then turn out to be. Debates between realists and anti-realists have proved inconclusive, with their foundational positions left as acts of faith, but realists can at least claim that their faith is rational, since it is fully coherent, and is never contradicted by reliable reports of experience. A slogan for realism is that reality exists independently of any experiences of it, but there are further ramifications of the position, and one of them would be that the convergence of good explanations reveals the real structure of the world. Given that view, we can believe in the reality of the ‘top’ of the essence as much as the ‘bottom’, but always with the caveat of limited precision. The world is a complex shifting pattern (as Dupré urges), and for humans to express precise truths about it may be as much of an illusion in physics as in metaphysics. We are never going to perfectly describe the weather, and the cosmos is like the weather, but an essence seems to have at least as much precise existence as a storm, and you can’t even talk about the weather if you don’t identify its messy ingredients.

We conclude that the essence of something is a real fact, but one which can only be picked out in the context of explanation. The main hallmark of an essence is that the explanations become more powerful when the essence is identified. Essences are not necessary for all explanation, but can a non-essentialist explanation be ‘powerful’? It is fundamental to an essence that it be the essence *of* something, the ‘something’ frequently being a physical object, but often an indeterminate number of other types of thing, including events, states of affairs, relations, properties, laws, general truths, and even abstract objects and systems. Harré points out that Newton’s First Law (that all physical change requires a force) is so idealised and general that it could never be experimentally tested (1993:22), and yet it is a cornerstone of traditional mechanics, which has given us powerful explanations. An obvious response from the essentialist would be to say that the law describes ‘the essence of everything’, but is that trying too hard, and forcing an unnatural framework onto the situation? The essentialist is the person who keeps asking *why?* (just as the essentialist approach to induction digs deeper). If Newton’s First Law is true, then why is it true? It is either primitive (Maudlin’s view, 2007:17)), or imposed on nature from without, or it arises from within. Mumford's attack on the whole concept of a 'law of nature' rests on the view that no decent account can be given of either the internal or external views (2004:144), but essentialism suggests that the internal view of so-called 'laws' is worth examining. If we describe this law as the Law of Inertia, that points to the nature of any body that falls under it. Newton's First Law is not usually presented in essentialist terms, but to explain the Law, reference to the essence of the bodies that exhibit inertia seems inescapable. The most generalised powerful explanations, which do not pick out any specific natural kind essence, still fall within the essentialist approach, by approaching the lowest level of all in our current grasp of nature.

Explanations of human events, such as the accession of King Richard III, are usually given in strikingly different terms from the explanations of physical science, and yet they may well qualify as 'powerful'. A 'powerful' explanation of that event would need to be simple and wide-ranging, such as a core of fifteenth century beliefs about hereditary monarchy, legitimacy and honour. Identifying such a 'core' would presumably arise through convergence from other explanations of the period, and we might expect the 'power' of the explanation involving such key beliefs to be revealed in fruitful explanations of many other events of that era. However, if the components of the weather are somewhat indeterminate, how much more is that true of the 'ingredients' of historical explanation? The problem here seems to be the extreme rarity of any sort of consensus about the true explanation of some historical event. Each event is unique, and all events depend partly on the particular motivations of a number of different people, so that the phenomena of convergence and fruitfulness are extremely rare in social and historical studies. Bold simplistic historical theses that might count as essentialist tend to soon fall by the wayside. We could try to impose the essentialist framework, but it is probably better to say of this example that the lack of consensus limits the power of the explanations.

It would be rash to proclaim that 'all powerful explanations are essentialist' (particularly if the definitions of 'powerful' and 'essential' were interdependent), but it is hard to see what 'powerful' could mean if it did not identify some central mechanism of the matter which produced wide-ranging effects. That claim presupposes a picture of reality that involves grounding and levels, but if those concepts are given up it not only undermines essentialism, but threatens the very notion of explaining anything.

All powerful explanations can be seen in essentialist terms (so that the essence of an essentialist explanation is its power, and the essence of a powerful explanation is the identification of essence), but are all essences explanatory, and are essences only explanatory? There is an affirmative answer to the first question: given that features of reality are only ever picked out as essential in the context of an explanation, and that the essence is picked precisely because it explains, it is virtually tautological that all essences do is explain. To describe that as absurdly narrow would be to underestimate the all-embracing nature of explanation. Traditional individual essences are held to deliver unity, continuity, a predication subject, dispositional nature, qualitative nature, kind and classifications. It is hard to envisage our being curious about any aspect of an individual which did not fall under those headings. Wiggins says that sortal essences merely individuate, and are not explanatory (2001:143, quoted above), but the assignment of a sortal brings with it the inductive generalisations that have accrued to that class of things, and so such essences are at least very informative – though it is obvious that merely classifying something will not offer a powerful explanation of it. Other modes of explanation, such as the covering law approach, are not essentialist.

Given that all essences are intrinsically explanatory, does it follow that essences make no contribution other than to explanations? That is, might explanatory essences have interesting non-explanatory side-effects? Given that we are taking the concept of essence to have little meaning in any mind-independent conception of nature (because the concept arises from human modes of enquiry), the only candidate for side-effects will be in human experience, thought and language, and the answer there seems to be strongly affirmative. Explanation intrudes into most of our daily thought, but there is plenty more that doesn't seem to qualify as explanatory, and yet seems to be rich in a range of essences, as focuses for thought. We suggested that the study of concepts should adopt a more essentialist approach. Nearly all of our concepts have originated in explanatory contexts, but then settled into the conventions of language and thought, with their explanatory heritage abandoned. We don't explain all day, but every moment is filled with the outputs of long-forgotten explanatory enquiries, all stamped with the essentialism which that required.

58. Value of essentialism

That concludes the enquiry into the relationship between explanation and essence. Perhaps the biggest objection to this sort of Aristotelian essentialism is not some specific case, but a general sense of intellectual apathy about the matter. Scientific metaphysicians will note that the admired scientists eschew essentialist talk, and their accounts of the world seem nevertheless to flourish. Opponents of metaphysics will see essentialism as the embodiment of everything they dislike. Aristotelians tend to imbibe the scholastic interpretation, which emphasises *Categories*, genus and species, and essence residing in the kind, leaving only limited interest in essences as explanation. Even followers of Kit Fine, who offers an admirable defence of the role of essence, will tend to see definition as all that matters, making superfluous our talk of actual explanatory features of the world. Fine's defence presents the opposition as between essence as 'conceived on the model of definition', and the concept as 'elucidated in modal terms' (1994:2), but this unduly emphasises the purely verbal activity of definition, and underplays the much closer engagement with the world found in explanation. So why should the explanatory view defended here matter, to scientists, or to philosophers, or to the rest of us? The answer has to be connected to their metaphysics.

Pasnau, following Leibniz, defends the separation of metaphysics from science, arguing that even if science gives up substantial forms, we can 'still think a genuine substance requires a form of some more abstract kind, not for a physical explanation, but for a full metaphysical understanding of how things are' (2011:580). We take Pasnau's view as to be avoided, since it invites the sort of attack launched by Ladyman and Ross, who write that 'the metaphysician has no test for the truth of her beliefs except that other metaphysicians can't think of obviously superior alternative beliefs. (They can always think of possibly superior ones, in profusion)'

(2007:58). Their preferred view is that ‘metaphysics is the enterprise of critically elucidating consilience networks across the sciences’ (2007:28), which makes metaphysicians into scientists, distinguished by their interdisciplinary and purely theoretical approach, and qualified for the job by their scientific knowledge. Some middle ground seems desirable, where metaphysicians are neither isolated from science nor absorbed by it. No theory of anything has any value if it contradicts the facts, and metaphysics must thus be answerable to the settled facts that emerge from science. Modern metaphysicians must work in the framework of modern facts, which was equally a requirement for Plato and Aristotle, except that more hidden facts have become accessible. It is also self-evident that metaphysicians work at a high level of abstraction and generality, using an appropriate vocabulary, and that ‘high level’ thought operates at a great distance from empirical evidence. Hence we will take metaphysics to be the attempt to construct a highly generalised and accurate account of the nature of reality, answerable to both everyday and scientific facts, but principally aiming for perfect coherence at its own level. This places the enterprise at a higher level of generality than the consilience of Ladyman and Ross (and the level of Maudlin 2007, who has metaphysics doing little more than describing the general structure of physics), on the assumption that they overestimate the difference which scientific discoveries should make to our metaphysics. As long as we bracket off the weirder fringes, little seems to have emerged from the sciences that challenges Aristotle’s view of nature (though the hidden nature of matter, and the evolution of species, are exceptions).

This view of metaphysics unifies the understanding of ordinary people with the understanding of scientists. If we respond that science continually contradicts everyday understanding, we should observe how happily ordinary people will absorb the findings of science (if they know about them), and accept that the understanding of modestly educated people is now full of talk of vitamins, black holes, subconscious thought, radioactivity, and plate tectonics. Metaphysics concerns the scheme of understanding employed by all parts of our community, and we hope to have shown that explanation is possibly the defining activity of the human intellect, and that we cannot see how this activity operates if we deny the central role played by essentialist thinking.

FIVE

Cases of Explanatory Essentialism

59. Testing the theory

In the light of the proposals of the previous chapters, we will now look at a few cases to see if they can be illuminated in this way, for ordinary people, for the philosopher, and for the scientist. That is, are there cases which are best understood as shaped into an essentialist structure largely by the explanatory practices that we bring to them, where an essential structure is one in which some key simple facts at the lowest level determine the nature of the structure? In order to fit that account, each case will need to meet the preconditions for successful explanation given in chapter three (of directionality, grounding, and bounded levels), and will respond to the challenge for an explanation by revealing convergence on certain low-level features which fruitfully illuminate the example. The best support for the present thesis will be cases where a direct request for the essence of the case cannot be satisfactorily met, but where an essentialist account has to be formulated when we demand to know *why* this case is as we find it. The vagaries of real world explanations, with their personal interests and cultural presuppositions, must be acknowledged here, but the assumption is that the bland question 'why is this thing the way it is?' has certain core answers (of history, motive, function, structure, components) which constitute a model of the case, and a basis for any subsequent enquiry that has narrower and more personal interests. The fact that an essentialist view is being offered which begins from human practices, rather than from objective features of the world, does not imply the kind of anti-realism espoused by Goodman (1978:Ch 1)) and Putnam (1981), since we assume that the levels, divisions and structures involved are thoroughly real. It is merely urged that if we are to understand essentialism, we must see the interventions of minds as an indispensable part of the picture.

Case 1: Unity and Counting

60. Symbols for counting

It might be said that arithmetic is a local invention of our cultures, and not a necessity for powerful thought, but our own most ambitious cultural achievements are now so enmeshed in mathematics that its superfluity looks exceedingly unlikely. It therefore seems worth picking out the phenomenon of counting, as one approach to a potential link between unified objects, to see whether that activity throws light on the unifying powers of the mind, and the role of unification in thought. Does the existing institution of counting have the unification of the counted entities as one of its preconditions? If so, does the explanation of this unifying process need to be conceptualised in an essentialist manner?

Counting seems to be a three-place relationship, requiring a mind to count, entities to be counted, and symbols for the counting. We will assume that to qualify as ‘counting’ the activity must be successful. If I learned by heart the natural numbers in an obscure language and used them to count sheep, but didn’t know how many sheep I had finally counted, I can be deemed to have failed. If I can manage ‘intransitive’ counting (recitation of familiar numerals) but am unable to label the objects numerically, that too would be failure. Heck asks whether counting might be ‘fundamentally a mindless exercise’ (2000:202), but makes the distinction that ‘the numerals are not mentioned in counting ….but are used’ (2000:194). Since I don’t previously know how many sheep I am counting, I can’t predict the final number, so to be a proficient counter I must know the cardinal value of each of the symbols I employ. There seems to be nothing special about the symbols I employ for my counting, since I could use the names of the villages in Hampshire as my symbols, but I would need to know how each one cashes out in terms of the actual underlying number of objects. As Heck puts it, ‘counting is not mere tagging: it is the successive assignment of cardinal numbers to increasingly large collections of objects’ (2000:202) – known to psychologists as the ‘cardinal word principle’. If I assign a unique cardinality to each of the village names and then assign a village to each sheep, this will still not tell me the total number of sheep if I am not systematic in my assignment. If there are three sheep, the third village I assign had better pick out the required number, and so on. Hence, for counting objects, the symbols must start with a symbol which refers to the concept of ‘one’, with subsequent symbols in a fixed order, known formally as a ‘well-ordering’ (a strict total ordering, in which every subset has a least member). We could, for example, lick the villages of Hampshire into appropriate shape by imposing an alphabetical order on them, and then require persons doing the counting to learn them off by heart. When we say that one of the symbols must refer to ‘three’ this may seem to pick out a platonic form which underlies all the symbol systems used for counting, but that would still not get us what we want if we didn’t know what the form meant in terms of sheep. As Russell puts it, ‘we want our numbers to be such as can be used for counting common objects, and this requires that our numbers should have a definite meaning, not merely that they should have certain formal properties’ (1919:10). So our symbols must be arranged appropriately, and must also have ordered cardinal meanings. For finite numbers the ordinals and cardinals are effectively the same, so the successive magnitudes of cardinality will match the successor operation for the ordinals. In transfinite arithmetic the two types of number behave differently, but we will focus on the normal practical counting of finite totals of objects.

61. Concepts for counting

Given that we have a set of well-ordered symbols, to which successive cardinal numbers have been assigned, we must now turn to the world and consider what might be necessary and sufficient conditions for identifying appropriate entities to be counted. We cannot just stand a person in a kitchen and say ‘go ahead, count!’ with any hope of a determinate result. Some specification of what to count is obviously required. An offer to ‘count the things!’ won’t do, because ‘thing’ is not sufficiently specific, and could include the tiniest of things (the quarks in the kitchen), the largest of things (if the Milky Way is partially ‘in’ the kitchen), or parts of larger things, or even the abstract objects embodied in the kitchen. It might also include all the mereological sums, and may generate infinities that exceed the capacity of a mere well-ordered sequence of symbols.

We might demand that 'individuation' of items is what must precede counting, but cases such as individuating 'all the coal in Scotland', which doesn't facilitate counting the coal, shows that this won't do. We need some concept which will separate out the things to be counted. Frege argued persuasively that even if we manage to individuate some distinct physical objects we are still not in a position to begin the counting, as illustrated by picking out two boots, but still not knowing whether you face two boots or one pair of boots (1884:§25). Being a pair is not some property of the boots, but a concept which we bring to bear on the situation, so that counting requires a further contribution from the mind of the counter. It is tempting at this point to think that the sortal concepts discussed earlier might do the job, as when we say 'count the sheep'. You say what sort of thing is to be counted, pick out things of that sort, and counting can then proceed. The traditional term for items that are counted is 'unit', implying some form of unity in each item, and the assignation of ones (which can be summed) to each entity. When counting, we can commonly ask 'what are the units for the count?'. It may seem that 'sheep' would exemplify what is required; however, that tells you to start counting with sheep as units, but not when your count is complete. In formal language, we need a 'domain' for the count, or a designation of 'scope' for the sortal concept. The obvious domain would be spatial (in a field, for example), but domains could be much quirkier. You could count for one minute, or count sheep images in Google, or count complete sheep on the butcher's mutton counter. This is significant, because we see that there is a 'normal' mode of daily counting, but that the institution of counting has enormous flexibility, invoking despair in an analytic philosopher who demands necessary and sufficient conditions.

The difficulty is illustrated in a criticism of Frege's approach to the matter. Frege felt that counting entirely rested on the application of a concept which possessed a determinate extension. His definition of a number was the cardinality of a set of equipollent sets (so that three is the cardinality of the set of all trios of items, such as the Graces, the Triumvirs and the prime numbers between six and fourteen – sets which exhibit one-to-one mapping between their respective members). For counting Frege wanted concepts which would generate collections which could be members of that number's constitutive collection, such as 'wheels of a tricycle', which picks out another trio of items which can meet the mapping requirements of the family of sets which embodies 'three'. This approach obviously elucidates interesting facts about cardinality, but does not give us all that we need for counting. We can 'see' the number three in a trio, but we can't do the same for 157, so we need to know about the procedure that will be involved. In summary, when we count we must apply our symbol series to the selected items, but we must also ensure that we meet three conditions: no item must be counted twice, no inappropriate item must be counted, and no items must be missed out (Rumfitt 2001:65).

Frege offered criteria for the sort of concept which would do this job. His best known proposal says that the concept must 'isolate [*abgrenzen*] what falls under it in a definite manner', and that the concept 'does not permit arbitrary division of it into parts' (1884:§54). The word 'isolate' is

Austin's translation, but this seems to be a rather strong requirement, as it would prohibit the counting of overlapping entities. If we draw a rectangle composed of various smaller rectangles, they seem to be countable in various ways (which could be specified according to boundaries or according to areas), but they are not isolated (Dummett 1973:549). However, the literal translation of *abgrenzen* is 'delimit', rather than Austin's 'isolate'. In Koslicki's discussion of Frege's proposal, she brings out his principle idea here, which is that countable things need boundaries, rather than full separation (1997). The second of Frege's proposals is to cover the difficulty of counting something under the concept 'red', given that a red thing is constituted of further redness. It is unclear how many reds one is looking at when faced with a red patch, because it endlessly divides into further red things (unlike a sheep, whose subdivisions are not sheep). The two proposals, that the concept should 'delimit' and should resist sub-division, both aim to give a determinate boundary to the target entity, so that we narrow down to a prescribed boundary (the skin and wool of a sheep, for example), and the narrowing then comes to a halt (since further subdivision of the animal no longer qualifies as 'sheep'). Rumfitt quotes a further remark from Frege, in an 1885 lecture, that our concept for counting requires not only 'sharpness of delimitation', but also 'a certain logical completeness' (2001:54). This seems to aim at the further conditions for successful counting just mentioned, such as ensuring that no qualifying item has been omitted from the count. If we say, with Frege, that the concept that generates the units of a count must home in on some boundary for a qualifying entity, halt at that boundary, and completely embrace whatever items have such boundaries, a satisfactory picture of counting according to a concept seems to have emerged.

It is when we step back from this promising proposal and view it more broadly that the interesting criticisms begin to appear. In brief, there are doubts about whether the number involved is actually rooted in the sortal concept, and there are problems when we investigate more complex instances of the concepts that meet Frege's nice criteria. Frege's account of the role of concepts in determining the number in a given situation seems correct and illuminating, but it does not give the whole story, especially when the actual procedure for counting various types of objects is considered.

In Yourgrau's summary of what he calls Frege's 'relativity argument' (that counting is entirely relative to the sortal concept employed), he sees Frege as employing an unstructured mereology of unprioritised parts, which are then determinately partitioned by appropriate sortal concepts. It is part of Frege's logicism that the cardinality of the resulting set arises from pure logic, since a number is nothing more than a property of a particular second-order set (1985:357). In defence of his view that a concept must precede counting, Frege said that a pack of cards can only be counted if the aspect of the pack which is to be counted is specified (1884:§22). In Yourgrau's view, this will not achieve the required result. He considers the set {Carter, Reagan}, and observes that if you ask 'how many?' of it, you encounter the same difficulty of someone counting in a kitchen – that you don't know whether to count the presidents, or their feet, or their relatives. Yourgrau's general summary is that 'we can address a set with any question at all that admits of a numerical reply' (1985:358). Hence the establishment of a determinate set does not solve the problem of how to count.

Of course, the Fregean will reply that the number attaches to the concept (‘Presidents…’, or ‘feet of…’, or ‘relatives of…’), rather than to the set, so if the feet of those presidents are to be counted, the resultant set will contain four objects rather than two. Yourgrau’s challenge is evaded, because the number always attaches to the elements of the set which is generated by the concept, and that is an unambiguous totality. However, consider the case where Carter unfortunately loses a foot. The extension of the concept ‘feet of those presidents’ now becomes three instead of four. The Fregean approach suggests that this must now be a different concept (perhaps by time-indexing), since it has a different extension, but it seems better to say that the feet themselves are the source of the number (three now, instead of four), rather than the covering concept that leads us to the feet. The sortal concept is an important feature of normal counting, but the numbers seem tied more closely to the members of the set than to the concept that generates the set.

If we then consider what types of sortal concept can do the job which Frege has specified, we meet further problems of a similar kind. If I am faced with a bowl of apples, pears and oranges, and told to count the apples and the pears, we must ask (in our Fregean context) which concept will generate a successful count. ‘Apple’ and ‘pear’ are just the type of natural sortal divider that we would hope for, offering limits and a criterion of completeness. ‘Fruit’ will not do, because we must not count the oranges. We can claim that there is a concept ‘apples-and-pears-in-this-bowl’ which produces the right set, but in actuality we do not employ such concepts. The actual procedure will pick out the units separately, under the concepts ‘apple’ or ‘pear’, with a further psychological act required to flip between the two as we count. A Fregean might reply that the count produce two separate totals, each under a plausible concept, followed by an act of addition, but this is not what actually happens, since no intermediate total of pears is ever registered during the count. The third item in the count can be a pear, and the fourth an apple, so the conceptualist account requires flipping between concepts during a single count, which seems to require control by an implausible meta-concept (especially in more complex cases). It seems that the combination of apples and pears is not the product of a sortal concept, but is a psychological achievement of the counter. We see that the psychology of counting is rather more anarchic than the neat Fregean picture appeared to suggest, since this combination of concepts could be extended to counting the apples and all the prime numbers I can think of, plus the names of any Prime Ministers who float into my mind during the first minute of the count.

62. Perception and counting

None of this challenges the requirement that concepts are necessary to generate a clear and determinate count. However, if we ask whether animals can do anything like counting, we may even begin to wonder about this too. No one (except the most gullible) thinks non-human animals can answer ‘five’ to a request for counting, but it seems obvious that many animals can register that five objects are more than four, and can respond to simple facts of cardinality. It also seems clear that puzzling objects, for which no obvious concept leaps to mind, such as weird artefacts dug up by archaeologists, or emerging shadows in the mist, can also be counted, long before they need to be clearly conceptualised. If they just have to fall under the concept ‘thing’ or ‘physical object’ (as Ayers suggest, 1974:139), that weakens the supposed requirement for a more determinate sortal that Fregeans had hoped for. If we imagine a test constructed by psychologists, in which there are six frames shown on a screen, and a rapid series of interesting objects are shown in the frames, the concepts that cover the nature of the objects would continually fluctuate, but the fact that there were six of them would be fixed in the mind of the observer, and the fixing would be by their position (as a three-by-two pattern, for example), rather than by a concept such as ‘frame’ or ‘apple’. The last ditch insistence that the concept of ‘position’ is needed does not seem plausible.

The point here is that an excessive reliance on concepts to give us the theory of counting will be at the expense of the vital role of more primitive perception, in which objects rather than concepts will have priority. If we hear three rings of the bell, or three explosions, it is not the concepts of ‘ring’ or ‘explosion’ which underpin the immediate apprehension of ‘three’, but the trio of sensual impacts. To point us towards this aspect of counting, Ayers gives the example of being told to count ‘the coins in the box’ (1974:139). The Fregean view implies that there is a concept *coins-in-the-box*, and an extension to the concept, of certain objects which fall within its meaning. The traditional ‘unit’ which counting requires would be coin-in-the-box, but Ayers observes that this activity could equally focus on the concept ‘coin’, with ‘in the box’ merely indicating the scope of the operation. The question of ‘scope’ arises because for Frege variables ranged over all objects (including, it gradually emerged, some rather incoherent ones), whereas modern logic requires the specification of a ‘domain’, which could be of quite limited extent (Dummett 1973:475). A Fregean concept is an open sentence, of the form ‘x is a coin’, where x ranges over all objects. If the coins in question are more restricted (for counting purposes), then the concept must do the work of restriction, so we need something like ‘x is a coin-in-this-box’. In modern logic, however, the domain is established first (‘the contents of the box’, perhaps), and only then does the concept begin its work, by just counting ‘coins’. We might be asked to count in the box-domain the objects that fall under the sortal ‘coin’, but we might equally be asked to count anything in the box, in which case *in-the-box* becomes the nearest we can get to a meaningful sortal concept. This drift between domain and sortal concept suggests that the simple approach in which the concept does all the prior work no longer seems sufficient. The domain itself might be established by a sortal concept, but it might equally be established by a list (which could include items of fruit, prime numbers and presidents in one indiscriminate collection), and one could be asked to count all the listed members of the domain, with no reference to a concept.

A further point which Ayers makes is that if we plausibly pick out an entity for counting, such as a sheep, by means of a concept, we must must still face the continuity of the object concerned. Events such as storms may be individuated by concepts, but their temporal boundaries are perceived rather than conceptualised. The case of the caterpillar and butterfly must be accommodated by the conceptualist. If every few weeks I take a census of ‘insects in this field’, I must count the later butterfly as one with the earlier caterpillar. The Fregean might say that the potential butterfly is part of the ‘caterpillar’ concept, and so on, but this must rest on a perception of the transition, and not on the concept of the intrinsic nature of the object perceived. Our perception of the world intermingles with our conceptualisation of it when we count, and Ayers argued that the primitive recognition of continuity must precede any use of sortal concepts (1974:117).

Frege was very critical of rival theories, and Tait makes a good case for the view that he seriously underestimated some interesting thinkers (1996). Frege was particularly severe on the view that numbers emerge from the mere perception of physical objects, by a process of psychological abstraction, and his attack caricatures the abstractionist approach. If a 'unit' is needed for counting, this seems to require the units to be somehow indistinguishable, and yet distinct from one another. If we count black and white cats we have to 'abstract' away the colour, position etc. of the cats, to turn them into countable units, which seems to mean that what is counted is no longer cats (1894:324). When we count the population of Germany, we would apparently have to turn each German into an 'average' German to achieve the identity between units that is required (1884:§42), which Frege rightly takes to be absurd. But no one ever thought that units should be understood in this way. Aristotle observes that if we count ten sheep and ten dogs 'the number is the same…, but it is not the same ten (because the objects it is predicated of are different)' (*Phys* 224a2), and he tells us that the unit is '*stipulated* to be indivisible' (and not that it actually *is* indivisible) (*Met* 1052b33). We should say that a unit sheep and a unit dog are the same in their *role* as units, but not that they *are* the same units. For Aristotle units of number are more enmeshed in the world, whereas for Frege they are pure phenomena from the 'third realm' of logic. For example, where Frege said that boots are either two boots or one pair, depending entirely on the sortal concept employed, Aristotle asserts that 'a pair of men do not make some one thing in addition to themselves' (*Met* 1082a26). Tait's analysis of the situation is that Frege has confused equality with identity (1996:59; Frege 1884:§39 would exemplify the confusion). If units are identical, then all units will merge into one unit, making a mockery of the concept, but they can be equal in respect of their magnitude, while retaining their distinctness. A unit dog will bark, and a unit sheep will bleat. In this way we can achieve the abstraction required for counting according to units, while retaining the relationship with the individual objects which are counted. For Frege the counting occurs in a world of concepts and sets, but for Aristotle counting remains in the physical world. A possible explanation of Frege's difficulty is his profound distaste for psychology in the philosophy of mathematics, when actually the simultaneous treatment of something as both a 'unit' and as a 'dog' can only be explained by a psychological operation. This approach seems to fit Dedekind's view that numbers 'serve as a means of apprehending more easily and more sharply the difference of things', and he also observes that counting shows 'the ability of the mind to relate things to things, to let a thing correspond to a thing, or to represent a thing by a thing, without which no thinking is possible' (1888:Pref).

A possible bridge between the approach to counting that relies entirely on concepts, and the approach that brings us closer to the direct perception of objects, is offered by Jenkins (2008), who makes an interesting case for an empirical account of arithmetic, which does not unwisely rely on the direct perception of number properties in groups of objects (Mill's hope, rightly rejected by Frege), but gives instead a more empirical account than is normal of the concepts themselves. The suggestion is that to think conceptually about the physical world is to remain highly engaged with the world, rather than retiring into the realm of pure thought. This is so, she says, because ‘the physical effects of the world on the brain explain our possessing the concepts we do’ (2008:224). This contrasts with the view implied by Frege (and endorsed by Geach (1957:40)) that the mind generates the concepts from within, as part of our rational endeavour to master the world. The consequence of Jenkins’s view is that ‘concepts which are indispensably useful for categorizing, understanding, explaining, and predicting our sensory input are likely to be ones which map the structure of that input well’ (144). Exploring this would take us far afield, but the account of concepts which Jenkins offers fits well with an picture of counting in which objects and concepts are jointly engaged. The division between the two is somewhat Humean in character, and suggests that some counting concerns ‘matters of fact’, while other counting concerns ‘relations of ideas’. Modern theorising entangles the two more richly than the neat Humean picture suggests, but it implies a loose division between ‘normal’ and ‘abnormal’ counting. ‘Normal’ counting responds to our direct perception of facts in the world, and involves Jenkins’s empirically rooted concepts, and the sort of abstracted pattern recognition which is explored in the ‘structuralism’ of Resnik, who argues that ‘mathematical knowledge has its roots in pattern recognition and representation’ (1997:9). The more ‘abnormal’ and fanciful examples of counting show an increasing involvement of the intellect, and more generalised concepts (such as modular components of the patterns, or intersections of the concepts). The intervention of the intellect to count both apples and pears in a single count is a first step in this direction, and the culmination is counting prime numbers, and the branch of mathematics called ‘number theory’.

63. Unity and counting

If counting is rooted in the perception of objects and patterns, guided by a conceptual scheme which itself arises from such things, human counting seems to be best explained by its probable origin, in coping with the physical world. The procedure is then generalised, and the question ‘how many?’ (the first step in counting, according to Frege) can broaden into seemingly unlimited areas. In all cases, though, of humble normality or bizarre abnormality, we have not parted from the idea that finite counting relies on the ‘unit’. Leibniz spoke of numbers as actually being mere collections of units (e.g.1686:121), but to understand ‘nine units’ you need prior understanding of ‘nine’, so the counting procedure that linked the cardinalities of numbers to actual groups of objects still seems required. The present concern is the extent to which ‘unity’ is a precondition of treating something as a ‘unit’, and (if so) whether we might explain that unity by postulating an essence which supports it.

Aristotle tells us that ‘arithmeticians posit that a unit is what is quantitatively indivisible’ (PA 72a22), in which we should register that the thing treated as a unit is not ‘indivisible’, but only indivisible in quantity. The notion of ‘quantity’ has a rich history in scholasticism, where it was offered as the feature that most clearly demarcated a distinct and perceptible substance. According to one theory, says Pasnau, quantity ‘is what makes the body's parts be spread out in a continuous and unified way’ (2011:280). Aristotle was struck by indivisibility as leading us to the unified entity (*Met* 1016b3), and also the unit, but he draws attention to the way in which such indivisibility hovers between mind and world. Early in Book *Iota* he outlines four theories of unity, and then comments that ‘the reason why all these things are unities is indivisibility; in some, it is indivisibility with regard to movement, in others with regard to thought and the account’ (*Met* 1052a33). The best cases of objective indivisibility are where the cause of the unified movement is ‘contained in itself’ (rather than being glued or nailed together – *Met* 1052a22), best fulfilled by animals. If unity can be just in the account [*logos*], however, then we have the sort of fairly unrestricted unification by the mind that Locke drew attention to when he observed that we can treat the Universe as a unity. What we treat as unified entities, and also as units, seems to exist on a continuum, with the world imposing unity (and unified concepts) on us at one end, when we experience the Moon, or an animal, or a loud bang, or a boot, and the mind freely perceiving unities wherever it likes (in a trout-turkey, for example) at the other. We may feel that there can be no common ingredient in such diversity, but this is where Frege’s contribution is so useful, because what is counted must meet the criteria for counting which he elucidated. If a Lewisian philosopher counts three trout-turkeys, they must be ‘delimited’, they must not be subject to ‘arbitrary sub-division’, and the group must be ‘logically complete’ (in concept or domain). Trout-turkeys are in disrepute because the only principle which can explain the delimitation and completion of their collection is the mind of the person counting. We can all count a few trout-turkeys, but the essence of the trout-turkey is entirely in the mind of the counter (where the explanation of its existence is to be found). At the other end of the spectrum of unity, where the world seems to do the unifying work and offers objects naturally suitable for counting (with any normal mind responding accordingly) then the explanation is to be found in the object. The unity of a sheep or the Moon do not have their source in human thought, since they meet Frege’s countability criteria with no help from us. Sheep come ‘complete’ and equipped with boundaries, with intrinsic unity in their movement as the obvious evidence. At that end of the spectrum it is plausible to say that the ‘nature’ of the object is what makes it countable, while at the other end it is the ‘nature’ of the person doing the counting which generates the countability criteria. An individual bee has an obvious unified life, but while treating a whole hive of bees as also having a unified life may be understandable, it seems to go ‘against the grain’ of the observed phenomenon. That there is uncertainty in the middle of the spectrum is shown when Koslicki asks ‘why do speakers of English count carrots but not asparagus? - there is no 'deep' reason’ (1997:424).

At the external and natural end of the spectrum of unification many objects seem to invite counting because they intrinsically meet the Fregean criteria for countability, with delimited boundaries, fairly distinct ‘logically complete’ kinds, and parts that differ in kind from the whole. It is here that the current thesis finds its best support, if we try to explain why an entity meets these criteria, and are faced with something like the essential nature of the thing. Animals offer the best evidence for this, but the tricky case of the unity of a mountain throws clearer light on the situation. Chambers dictionary defines a mountain as ‘a very high, steep hill, often of bare rock’. This captures the concept well, and seems to offer prima facie countability, since it is fairly ‘complete’, is easily individuated, and the rocky material is not itself high or steep. The problem, though, familiar to philosophers, is with the outer boundary of the object. If you walk up a mountain, at what point are you first ‘on’ that mountain? For us this generates the difficulty of counting a group of connected mountains, and to ask ‘how many mountains are there in the Alps?’ is clearly absurd. Everyone knows what a mountain is, and yet a precise count of some mountains is rarely possible. Mountains fail on one vital Fregean criterion – the need for an outer boundary. One might attempt a precisification of the outer boundary, by obtaining a consensus from walkers of when they were definitely on the mountain, or one might paint a white line in a plausible spot (if one were selling a mountain), but none of these will satisfactorily produce a total for the number of mountains on a long, undulating jagged ridge. The underlying problem is found in the dictionary definition, which gives a good account of the mountain’s peak, and makes no mention of the outer border. If we thought that Chambers had given us the *logos* for a mountain, we would have to accept that the outer boundary is not relevant, because our concept only concerns peaks. Hence the very nature of mountains precludes the delimitation of the object which Frege showed to be essential to the count. We can’t count mountains because unity is not part of their character, and they resist unification by the mind in any way other than stipulation, on which consensus seems unlikely.

We may have to conclude that there is no neat essentialist underpinning to the institution of counting, even when we pursue the explanation of some count to its most basic elements. Too many cases of counting rest on convention, stipulation, context and the interests of the person doing the counting. One may even reject the whole essentialist picture by simply specifying an object to be nothing more than a bordered region of space-time (as Quine does – 1970:36). However, there is a case to be made for counting at one end of our proposed ‘spectrum of unification’ to be understood in essentialist terms. If you stare at your kitchen, or the local landscape, and seek a ‘natural’ count of the ingredients, you can at least begin counting objects which hang together and have sharp borders, and these (it seems reasonable to suggest) are features dictated by the determinate ‘nature’ of each item. Without such items it is hard to imagine how the institution of counting would ever have begun. No one would suggest, of course, that we can only count some ‘natural’ unity if we have first grasped its essence, but the discussion seems to show a necessary connection between whatever is labelled as ‘essence’, and the features that are actually required for counting. Items that naturally fall under appropriate sortal concepts, or even impress their distinctness directly on our perceptions, will do so because they have an intrinsic nature which determines sufficient unity and borders for the role. Our study of counting does not clinch the case for essentialism, but it shows how essentialist thought fits into a satisfactory and coherent explanation of the way our counting procedures connect us to the world.

Case 2: Axiomatised Systems

64. Explaining systems

Having considered the unity and countability of physical objects, we will now look at more theoretical areas. Most ontology of the abstract focus on 'objects', but we will focus on the idea of 'systems', since they offer more of the preconditions for explanation in their clearly demarcated structures. To permit explanations, we have proposed that there must be an inherent structure, a direction within the structure, relations of determination and dependence that can be tracked through the structure, and something conforming to 'high' and 'low' levels, arising from modular components. The concept of a 'powerful' explanation requires a convergence within the determinations, and more fruitful explanations arising from certain parts of the system. It is hard to envisage interesting explanations without such a context. The sorts of systems that we have in mind are the well known areas of study involving inferences, numbers, and collections – that is to say, logic, arithmetic, and set theory. Our main question is whether the studies of such formal systems aim at explaining them, and (if so) whether the types of explanation that emerge fit the pattern we have been calling 'essentialist', even if there is no agreement about explanatory success.

65. The kernel of a system

We will begin with some remarks from Frege about the concept of a theoretical system. His 1914 'Logic in Mathematics' lectures assert that current mathematics is fragmentary, and needs to be shaped into a 'system'. This begins with the concepts of an 'inference' and a 'theorem'. This creates a 'chain' of inferences, proceeding into ever greater complexity. But he then notes that you can move backwards in the chain, so that 'the circle of theorems closes in more and more', eventually arriving at truths which are not inferred, and these are the 'axioms, postulates or definitions'. He then observes that

> *Science …..must endeavour to make the circle of unprovable primitive truths as small as possible, for the whole of mathematics is contained in these primitive truths as in a kernel. Our only concern is to generate the whole of mathematics from this kernel. The essence of mathematics has to be defined by this kernel of truths.* (1914:204-5)

Russell endorsed this view, and Hilbert made it into the prime quest of mathematics, but the modern response is that such dreams have been dashed, largely by Kurt Gödel. The simple idea that one could identify the axioms by Frege's method, and that the consistency and truths of the whole system would thereby follow, certainly met strong challenges from Gödel's Incompleteness Theorems (the First undermining a complete proof of the truths, and the Second undermining the internal establishment of consistency (Smith 2007:343)). However, the flat rejection of Frege's approach is a simplistic response to these developments, since Gödel himself did not share such a view. Of his famous First Theorem, he wrote in a 1932 letter that if one adds a definition of truth, then 'with its help one can show that undecidable sentences becomes decidable in systems which ascend further in the sequence of types' (Koellner 2006:6). By 1961 Gödel optimistically wrote upholding 'the belief that for clear questions posed by reason, reason can also find clear answers' (Koellner 2006:12). Contemporary theoreticians pursue axiomatic theories of truth and set theory aimed at finding appropriately expressive systems which can settle difficult questions such as the Continuum Hypothesis, the Liar Paradox, and incompleteness, while retaining something like the sensible kernel invoked by Frege (Maddy 2011; Halbach 2011). Frege's essentialist view of mathematics remains viable, and worth examining.

Not all thinkers take our preconditions for explanation to be found in such formal systems. A common modern view is that the sorts of system which can be implemented on machines are primarily syntactic in nature, and work out the inevitable consequences of prior assumptions, which do not even need to be true. There is nothing more to be known than the sequence of operations involved. Curry took a different view when he wrote that 'in the study of formal systems we do not confine ourselves to the derivation of elementary propositions step by step; rather we take the system, defined by its primitive frame, as datum, and then study it by any means at our command' (1954:204), and Kreisel said that 'it is necessary to use non-mathematical concepts …for a significant approach to foundations' (1958:213). The sort of 'direction' within a system that is required for explanation might be provided by ordering relations, or part-whole relations, but Aristotle tries to express a different notion of 'priority' when he writes that 'one is prior to two because if there are two it follows at once that there is one, whereas if there is one there is not necessarily two' (*Cat* 14a29), and elsewhere he is explicit that this priority is not the parthood relation (*Met* 1034b24). Similarly he tells us that in the syllogism 'the first figure has no need of the others, while it is by means of the first that the other two figures are developed… and therefore the first figure is the primary condition of knowledge' (*PosA* 79a31). This concept of priority and direction in reasoning was endorsed in the rationalist tradition, and Leibniz talks of 'the connection and natural order of truths, which is always the same' (*New Ess* 1710:412).

There is a modern consensus that strong (Euclidean) foundationalism, beginning with self-evident certainties from which the system is constructed, will not do. Thus Zermelo, the main founder of the axiomatic approach to set theory, writes that 'principles must be judged from the point of view of science, and not science from the point of view of principles fixed once and for all' (1908:189). His introduction of the controversial Axiom of Choice simply on the grounds that it facilitated very useful proofs is the classic instance of such a view. Russell adopted a similar approach to the axioms and practices of arithmetic, when he wrote that 'it is an apparent absurdity in proceeding ...through many rather recondite propositions of symbolic logic, to the 'proof' of such truisms as 2+2=4: for it is plain that the conclusion is more certain than the premises, and the supposed proof seems futile' (1907:272). This picture fits the view we have been propounding, because explanation must begin with a puzzle, and the principle puzzle in formal systems seems not to be the grounding of the truth of the axioms, but the coherence and fruitfulness of the ongoing system (just as the spectacular success of the physical sciences is the main datum facing philosophers of that enterprise, and linguistics addresses language as a going concern).

The idea that the world of truths has a direction and inherent structure was developed by Frege. His central idea is that 'proof has as its goal not only to raise the truth of a proposition above all doubts, but additionally to provide insight into the interdependence of truths' (1884:§2). This explanatory aspect of Frege's thought has been examined sympathetically by Burge and Jeshion (though Heck demurs, describing Frege's claim that there is a dependence relation between truths as 'obscure and suspect' (2002:190)). Faith in the total interconnection of truths has declined in modern times, but Burge's view is that 'Gödel undermined Frege's assumption that all but the basic truths are provable in a system, but insofar as one conceives of proof informally as an epistemic ordering among truths, one can see his vision as worth developing' (2000:361). The standard view of this matter is expressed by Hart: 'Frege thinks there is a single right deductive order of the truths. This is not an epistemic order, but a logical order, and it is our job to arrange our beliefs in this order if we can make it out' (2010:44). Jeshion, in support of Burge, responds to that view by saying that 'Frege thought that the relations of epistemic justification in a science mirrors the natural ordering of truths: in particular, what is self-evident is *selbstverstandlich* [self-standing]' (2001:944).

The idea that foundational truths are self-evident is appealing, if our understanding of truths is to give an accurate picture of their structure, and Frege tells us that 'it is part of the concept of an axiom that it can be recognised as true independently of other truths' (Burge 1998:326). There is a traditional distinction (found in Aquinas) between what is intrinsically self-evident and what is self-evident (or obvious) to us. Burge's summary of Frege takes the first view, that a self-evident truth is one believed by an ideal mind, on the basis of understanding rather than inference, and unavoidably believed when fully understood (1998:350). Jeshion, however, argues that this underestimates the way in which the minds of actual believers (rather than ideal ones) are involved in Frege's account (2001:937). We take basic ideas which seem obvious to us, and judge that this obviousness is merited by rational assessment, offering objectivity. This allows a more plausible fallibilist view of self-evidence, by maximising the endeavour of normal minds to fulfil the requirement of truth. Another feature of Frege's basic beliefs picked out by both Burge and Jeshion is that they must exhibit a high degree of generality. This seems an inevitable requirement, given how many truths are to be supported by these very few basic truths, but Burge detects a difficulty because highly generalised a priori insights will struggle to get back to the particular truths which must be our ultimate aim (and which Kant held to be basic to mathematics) (Burge 2000).

Thus we are offered an optimistic rationalism which claims epistemic success, where pessimistic rationalism accepts the ordered truths but offers less hope of understanding them. Confidence in critical self-evidence and in the logic is the basis for the optimism, offering the prospect of convergence on the lowest level, maps of dependence relations, the clarification of boundaries and overlaps, and the identification of a fruitful kernel to the system. Essentialism flourishes best in this optimistic scenario, where the ordered structure offers an explanation of the system. Thus Burge says of this approach that 'understanding logical structure derives from seeing what structures are most fruitful in accounting for the patterns of inference' (1998:354). When Frege uses the German word *Grund* he means not only 'ground' but also 'reason', and so the basic ingredients of a proof not only explain its logic, but are also the groundings which justify our beliefs. Russell endorsed Frege's view when he wrote that 'in mathematics, except in the earliest parts, the propositions from which a given proposition is deduced generally give the reason why we believe the given proposition' (1907:273).

An illustration of the dependence relation in Frege is his discussion of the 'direction' of a straight line. Initially it seems unclear whether this concept depends on that of 'parallel', or whether the priority goes the other way. Frege appeals to intuition, which seems to give 'straight line' and then 'parallel' as fairly self-evident concepts, leaving 'direction' of a line to be defined as the extension of the concept of being parallel to that line (1884:§64-68). Dummett observes that 'Frege appeals to a general principle that nothing should be defined in terms of that to which it is conceptually prior' (1991:33), and so the procedure of definition accompanies the proofs as a further technique for tracking dependence relations within a system.

We are not only looking for structures with a direction in their dependence relations, but also for 'levels', with roughly demarcated upper and lower boundaries. In formal systems it might be better to speak of 'outer' rather than 'upper' boundaries, and these are often demarcated by the concept of the 'closure' of a well specified theory or model. This would lead us to talk of the 'core' of the system rather than of a broad 'lower level', and this is exactly the word used by Frege in that context (1879:§13). This gives us a metaphorical bullseye, rather than a bottom layer, but talk of foundations still seems appropriate. We suggested that modular construction would generate the foundation section of a system, and this is very evident in formal systems. For example, Walicki begins an introduction to mathematical logic by writing that 'in order to construct precise and valid patterns of arguments one has to determine their 'building blocks'; one has to identify the basic terms, their kinds and means of combination' (2012:2). Burge quotes Frege as using the same language, when he wrote that 'the properties belonging to these ultimate building blocks of a discipline contain, as it were in a nutshell, its whole contents' (1998:320). All of the systems under discussion exhibit the same phenomenon of a small number of very simple or primitive ingredients, and so the picture of a 'level' suggested earlier becomes particularly clear. Jeshion is explicitly essentialist when she makes this point: 'the primitive truths contain the core of arithmetic because their constituents are simples which define the essential boundaries of the subject. …The primitive truths are the most general ones, containing the basic, essence determining elements' (2001:947). Frege said he was searching amongst the multitude of laws in a system for 'those that, by their power, contain all of them' (1879:§13).

66. Axioms and explanation

Typical building blocks of systems are objects, rules, truths, and definitions. Explanations begin with puzzles, and it was an interesting feature of the modern visual proof of the Pythagoras Theorem (cited above) that it seemed to dissolve the puzzle, rather than give a formal explanation. In such cases we see directly the essence of the phenomenon – a common view of geometry in the seventeenth century. Steiner offers the unusual view that it is the objects (such as triangles and circles) which are foundational, and in his defence of explanation in mathematics he writes that 'an explanatory proof makes reference to the 'characterizing property' of an entity or structure mentioned in the theorem, where the proof depends on the property; if we substitute a different object, the theory collapses' (1978:34). Such essentialism is appealing for Aristotelians, but most modern thinkers assume we can dig deeper than 'objects' or whole 'structures', and look to rules and truths for foundations. Both Steiner and Mancosu (2008) offer examples of alternative proofs in mathematics, where one proof is more explanatory than the other because it reveals more of what underpins the result. Definitions can conjure up new primitive objects, but the definitions rest on truths and rules, and these tell us what we can and cannot do with our 'objects'.

The axiomatic approach to foundations focuses on a set of initial truths. It is tempting to think (with Euclid) that our explanations will simply converge on self-evident axioms, which can then be denominated as 'essential', but for Frege this is not correct, because axioms are too dependent on the particular system in which they have their role. Frege has a modern awareness (stimulated by challenges to the Euclidean axioms) that a system may have more than one axiomatisation, and that modifications of axioms can generate new systems. Russell agrees when he writes that 'premises which are ultimate in one investigation may cease to be so in another' (1907:273). Frege's view is closer to Euclid than to modern views, though, because he insists that axioms have to be true, whereas in modern thought an axiom just has to play a formal role, and could just as well be false. Frege writes that 'traditionally, what is called an axiom is a thought whose truth is certain without, however, being provable by a chain of logical inferences. The laws of logic, too, are of this nature' (quoted by Burge, 1998:323). Being unprovable is obviously a main hallmark of the axioms, but an unproved truth is only an axiom if it is used in a system. If there were agreed sets of true axioms for set theory, classical logic, arithmetic and geometry, these would seem to fit nicely the explanatory essentialism we have been developing, since they would be the inevitable and unique focus for the understanding of each of those systems. Even false axioms might, of course, offer us the kernel of the formal system which they support.

67. The essence of classical logic

The question 'does classical logic have an essence?' requires the prior question 'is there an intrinsic explanation of classical logic?'. Classical logic may have an extrinsic explanation (in human psychology, or the abstract structure of nature, or the character of pure reason), but that has no bearing on the aspect that interest us. We want to understand the nature of logic, not its cause. There are, inevitably, thinkers who see logic as too flexible and human to have an essence: Carnap said you can use any logic you like, Goodman says you give up rules if you don't like their results, Quine says any logic can be changed if the science needs it, and Nietzsche saw logic as the will to power. However, classical logic is a going concern, and it centres on implication relations, resting on the foundational idea that a contradiction is unacceptable. We can test any assumption, by seeing if it implies a contradiction. The steps in the proof involve a set of rules, connectives and basic principles.

In the early stages of classical logic it was normal to characterise the system axiomatically. This followed the example Hilbert had set with geometry, and Hilbert optimistically promoted such approaches, until Gödel famously showed the limitations of axiom systems. Rumfitt summarises the modern view of this matter thus: ‘the geometrical style of formalization of logic is now little more than a quaint anachronism, largely because it fails to show logical truths for what they are: simply by-products of rules of inference that are applicable to suppositions’ (2010:41). There is now a consensus that the best exposition of the basics of classical logic is Gentzen’s system of ‘natural deduction’. Rather than defining logical connectives by intuitive statements of their nature or meaning, they are treated as rules for their role, showing when a connective can be introduced and when eliminated. Since (according to Gentzen himself) the elimination rules are implied by the introduction rules, we boil the logic down to a set of introduction rules (Read 1995:229). Thus given two affirmative propositions, we can link them with ‘and’; if we have one proposition, we can affirm it ‘or’ some second proposition. Bostock demonstrates how not only all the simple connectives but also traditional rules such as Modus Ponens or Conditional Proof can be expressed (with a little ingenuity) in terms of such introduction rules (1997:Ch.6). Part of Gentzen’s achievement was to present arguments in atomic steps, applying one rule at a time, which made proof fully transparent to anyone puzzled by it (Prawitz 1971:202). The standard alternative to the natural deduction account of connectives (in terms of their role) is to first assert their meaning (as truth conditions, in terms of truth tables). Thus Mill argued that introducing ‘and’ added extra meaning to two propositions, rather than melding them into one (1843:1.4.3). Prior parodied the natural deduction account (in terms of mere role) by offering the connective ‘tonk’, which combines or-introduction with and-elimination, and which leads to deductive anarchy (1960). This shows that the essence of logic involves a little more than the mere rules, and Belnap responds to Prior by adding that natural deduction rests on a prior grasp of deduction as a ‘going concern’, which requires consistency in the connectives. The addition of ‘tonk’ is inconsistent because it is not conservative, in that it allows new deductions not involving ‘tonk’ itself (Belnap 1962). If ‘tonk’ itself is inconsistent, it fails the first requirement for admission into classical logic.

This seems to narrow down the number of plausible introduction rules, and Russell surmised that there exist eight or nine authentic connectives in logic (1903:11), though this rests on a shared intuition about the going concern of reasoning, which may rest on extrinsic explanations of the matter. Gentzen felt that his rules constituted definitions of the connectives, but saying how something can be used may not suffice for a definition (which should give the nature of the definiendum). We will assume, though, that any further explication of the nature of the connectives will have to be extrinsic in character. Thus the rules for connectives might constitute the essence of the logic, but the essences of the connectives would take us out of the level we are considering.

If the natural deduction rules are the kernel or core of classical logic, then all proofs are fully explained in this manner. There is one caveat, however. The idea of an explanatory essence involves a rich system, with some core aspect which does the explaining (in the manner of Aristotle’s ‘formal cause’), but in the case of natural deduction it is not clear what is being explained, given that there is nothing more to it than the application of these rules. It may be that Gentzen has achieved what our visual proof of the Pythagoras Theorem achieved – of making what is happening so obvious that he has dissolved the puzzle rather than explaining it.

If the rules *are* the logic, rather than explaining it, then essentialist thinking may no longer be relevant. We may want to dig to the next level, by enquiring where the self-evidence comes from, and by comparing classical logic with the multitude of alternative systems, but we will leave the matter there.

68. The essence of set theory

Set theory operates in the world of standard mathematical logic, but with the addition of one two-place relation ∈, read as '…is a member of…'. This requires a 'set' on its right-hand side, a unified entity which can have members. To explain set theory we must also know what constitutes a set, and what relations between the sets themselves are permissible. The early history of the subject reveals that attempts to constitute sets according to a defining property (all the red things, for example) hits trouble because certain properties result in impossible sets, so in most modern theories the sets are constituted as their inventor (Cantor) intended, by simply specifying their members (Maddy 1988; Lavine 1994:Chs 4 and 5). There may be two concepts of set here, but the extensional approach has seemed safer. This gives us a first truth of set theory, that if two sets have the same members they are the same set. This is the most basic axiom (Extension), and set theory has become the best known exemplar of a system built entirely on axioms. If we think of the right-hand side of the membership relation as offering a container for things on the left-hand side, this needs the container to occasionally be empty, and so another axiom allows there to be an Empty Set. We then add that all the members buried within a set can form a set of their own (Union), and that two sets can be combined (Pairing), and the most obvious features of sets have then been specified, largely in terms of freedom to manipulate sets as we wish, as long as membership is respected. Other axioms followed, the most interesting being the Axiom of Choice. It was found that for certain key proofs in the going concern of set theory, it was desirable to generate new sets by selecting one element from each of a collection of sets, even when no obvious principle for the selection was evident. Zermelo went ahead and did this to prove a useful result, and then proposed that Choice should be axiomatic. The ultimate authority for axiomatising Choice was no more than an intuitive sense that the principle of choosing elements seemed reasonable, and that the results that followed also seemed reasonable. Zermelo gave as his guiding principle that 'starting from set theory as it is historically given ...we must, on the one hand, restrict these principles sufficiently to exclude as contradiction and, on the other, take them sufficiently wide to retain all that is valuable in this theory' (1908:200). Choice offered many successes, but one notable failure: the Banach-Tarski Theorem used Choice to prove that a single sphere could be decomposed into two spheres both identical to the single one. Because Choice is so attractive, its supporters respond by rethinking how to represent geometry, rather than how to cramp set theory (Maddy 2011:35). The end result of this approach is that the Axiom of Choice remains uncontroversial, and orthodox in ZFC set theory.

Various other difficulties emerged, but by only working with sets whose members had been specified, and adding that sets can't be members of themselves (Foundation), the repeated application of the Power Set axiom (making a new set using all the subsets of a given set) gives us the set hierarchy (V) in what is called its 'iterative conception', which largely avoids controversy and paradox. The standard set theory is ZFC, as loosely described here, but there are rival versions which evidently have their uses. For those who resist the explosive force of the Power Set axiom, for example, there is the more cautious Kripke-Platek system, which leaves it out.

Given the orthodox account of ZFC, we can wonder whether the explanatory picture we have developed is applicable to it. The hierarchy of constructed sets which arises from the emptiness at its base by means of the axioms certainly exhibits priority, dependence and direction. The difficulty with whether this picture conforms to Frege's essentialist account, which seeks a kernel with the power to generate the whole system and enable us to understand it, is illustrated by the account of the Axiom of Choice. We are looking for an essence with at least some degree of self-evidence, and security rooted in some lower conceptual level, but the acceptance of Choice seems remarkably pragmatic and conventional. One view of the matter is illustrated by Gödel's bold claim that 'we do have something like a perception of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true' (1947:483-4). That very platonist view is controversial, and a commoner view is seen in the comment on Gödel by Boolos, that while the minimal axioms of Extensionality and Pairing may seem obvious, since without them there wouldn't be any set theory, none of the others has such a firm status (1998:130).

Maddy argues that the axioms have some 'objective' truth, because they model mathematics so well (2013), but the fact remains that new axioms are still being explored, especially with a view to settling the difficulties of the Continuum Hypothesis ('that there exists no cardinal number between the power of any arbitrary set and the power of the set of its subsets' - Gödel 1944:464). The delicate balance between achieving exciting results in mathematics, while founding them on axioms that seem reasonably plausible, suggests that we are not yet looking at a distinct essence of set theory. The difficulty for our picture in that balancing act is not that we have not settled on the essence, but rather that it is not quite clear what the puzzle is. Set theory is certainly a going concern which needs to be understood, but the axioms do not seem entirely obvious in themselves, and they appear to actually dictate set theory practice, rather than explaining it.

69. The essence of arithmetic

Mayberry writes that 'if we grant, as surely we must, the central importance of proof and definition, then we must also grant that mathematics not only needs, but in fact has, foundations' (1994:405). Putnam expresses the rival view that 'I do not believe mathematics either has or needs 'foundations'' (1967:295). As we have seen, essentialism need not be foundationalist, since it requires only that a modular construction implies a lower level (below which an infinite regress, or a blurring into vagueness, would be an independent problem). Candidates for the foundational discipline of mathematics include various set theories, type theory, category theory, model theory and topology, but if we focus on arithmetic the possibilities for foundation or essence may become clearer. The quests for foundation and essence are related, and foundations and axioms for arithmetic have been proposed which are candidates to do both jobs.

Mill suggested that the two Euclidean axioms (sums of equals are equal, and differences of equals are equal) together with a progression based on units, suffices for arithmetic (1843:2.6.3; Shapiro 2000:95), but Kant had already rejected the Euclidean axioms, on the grounds that they were purely analytic, and he denied that arithmetic had any axioms (1781:B204). The Euclidean axioms make addition and subtraction central, but treat 'equal' as primitive, without revealing what sorts of things partake of equality. At the centre of the modern debate we find the Peano Axioms, as either giving the simple nature of arithmetic, or else stating most clearly the puzzle that must be explained. Frege reduced arithmetic to what we would call axioms (and he eventually derived the Peano Axioms from his system), but for him the essence of arithmetic was found in the underlying logic, which he expressed axiomatically (Dummett 1991:12). Modern logicists are inclined to take Hume's Principle as the single truth from which arithmetic is built. An alternative widely held view is that arithmetic either is set theory, or reduces to set theory, or is most revealingly modelled by set theory. Since set theory is a thoroughly axiomatic activity, the axioms of arithmetic would then be found in the axioms for sets. We will examine whether these three modern views offer a 'kernel' for arithmetic, and thus clarify the target and limits of the essentialist quest.

70. Peano Axioms as the essence of arithmetic

It was Dedekind who formulated the best known axioms for arithmetic (now known as the Peano Axioms). Informally, the axioms tell us that zero is a number but not a successor, all numbers have a unique successor, and whatever holds of zero and of some number and its successor will therefore hold of all numbers (Wright 1983:xiii). This generates the complete and well-ordered sequence required, and implies the operations of standard arithmetic. For theorists of arithmetic the Peano Axioms have been major success story. Russell describes them as recommended by 'their inherent obviousness' (1907:276), and Potter says 'it is a remarkable fact that all the arithmetical properties of the natural numbers can be derived from such a small number of assumptions' (2004:92). It is a commonplace criterion for any theory of numbers that if the Peano Axioms cannot be inferred from it then it will have few adherents, because that is the benchmark for telling the story correctly.

The Axioms are built around the successor relation, and so they emphasise the primacy of the ordinal numbers rather than of the cardinal numbers. It is by no means clear whether we should give priority to cardinality or ordinality when considering the natural numbers (though In the world of transfinite numbers they come apart and are quite different concepts). Russell says there is no logical priority between the two (1903:§230), but that the cardinals are simpler because they rely only on one-one relations (Frege's view), whereas ordinals also involve serial relations (§232). Also cardinals can be seen in progression before any theory of progressions is applied to them (§243). Wright even makes the claim that someone 'could be familiar with the natural numbers as objects ….without conceiving of them as ordered in a progression at all' (1983:118). Dedekind, on the other hand, says he regards 'the whole of arithmetic as a necessary, or at least natural, consequence of the simplest arithmetic act, that of counting, and counting itself is nothing else than the successive creation of the infinite series of positive integers' (1872:§1), which gives priority to the ordinals. In counting objects we saw that we needed both ordering and cardinality, but Dedekind takes ordering to be more fundamental. The most extreme defence of the ordinal approach comes from Quine, who says that for the explication of the natural numbers 'any progression will do nicely', and he rejects Russell's further criterion that the progression must also 'measure multiplicity', on the grounds that any progression will inevitably do that (1960:262-3). Benacerraf rejects Quine's view, and asserts that 'transitive counting …is part and parcel of the explication of number' (1965:275 n2).

We can hardly settle such a dispute here, but we can focus on the question of what constitutes the essence of number, meaning that we want an explanation of numbers, which results in a full understanding of their nature. If you take ordinals as basic, you must give an account of the cardinals, and vice versa, and it turns out that both of these can easily be done. If cardinal numbers are collections of each magnitude (partitioned by the one-to-one relation), then the successor relation can be defined as the union of that set with its own singleton set (Frege 1884:§76; George and Velleman 2002:58), and an ordinal progression of numbers results. On the other hand, Hart tells us that Von Neumann simply treated cardinals as a special sort of ordinal (2010:73), and for Dedekind and Cantor 'the cardinal number of a set S is the least ordinal onto whose predecessors the members of S can be mapped one-one' (Heck 2000:200). Thus for a set with five elements, the ordinal '5' can map 0, 1, 2, 3, and 4 onto them.

There is no Aristotelian deductive priority between the two types of number, but there are plausible intuitive grounds on both sides of the argument. Russell proposed that cardinals are prior because they are simpler, so that there is a concept of number available before the introduction of the idea that they can form a progression. On the other hand Cantor derived the simplicity of the cardinals by abstraction from the richer concept of the ordinals. Fine defends this approach, and quotes Cantor's claim that cardinal number is 'the general concept which …arises from the aggregate M when we make abstraction of the nature of its various elements m and of the order in which they are given' (1998:599). The response to Russell would be that cardinals are an incomplete account of numbers, since ordinals contain more information, and thus give a better explanation of numerical phenomena.

If good explanations focus on the essence, then Koslicki offers support for the priority of the ordinals when she writes that 'being the successor of the successor of the number 0 is more explanatory of the essential nature of the number 2 than …being the predecessor of the number 3, since the first mirrors more closely than the second does the method by which the number 2 is constructed from a basic entity, the number 0, together with a relation which is taken as primitive, the successor relation' (2012:199). Sympathisers with structuralism in mathematics will find this sufficient for our purposes, since the structural position of 2 seems to be fully illuminated by giving easily grasped primitives, and a rule for construction.

In addition to the importance of progressions and explanatory power, the main reason why mathematicians favour the Peano Axioms is that they are 'categorical'. Dedekind proved that ‘all simply infinite systems are similar to the number-series N, and consequently also to one another’ (1888:theorem 132). That is, that any systems which fit the requirements of the Peano Axioms (the second-order version) will be identical ‘up to isomorphism’, meaning that they will exactly map onto one another, and exclude non-standard models (Read 1995:49). They will be the same in the way that all the standard chess sets in the world are the same. Cartwright’s verdict on this situation is that many people think that ‘the concept of natural number is adequately represented by those axioms’, but that each model of the axioms may only be ‘taken as’ the natural numbers (rather than identified with them) (1962:48). In the spirit of Structuralism (of which Dedekind is the patriarch) this is seen as sufficient, since the structure is all that is required; that is as close as you can get to the essence of the natural numbers. Mayberry observes that ‘the central dogma of the axiomatic method is this: isomorphic structures are mathematically indistinguishable in their essential properties’ (1994:406), so that categoricity will be the main aspiration of any axiomatic system.

A critic of this approach is Almog, who says ‘there are reasons to worry about Dedekind’s essentialist project’ (2010:363). The gist of his criticism is that while all of the models of the axioms are isomorphic to one another, a model is not the real thing, and there are what he calls ‘unintended twins’ of the system being modelled (just as Putnam proposed an unintended twin for water on Twin Earth) (2010:366). Almog’s solution is to endorse another option in this area, which is Skolem’s proposal that the series of natural numbers is itself treated as primitive (and so doesn’t require ‘models’). Skolem writes that ‘the initial foundations should be immediately clear, natural and not open to question; this is satisfied by the notion of integer and by inductive inference’ (1922:299). To treat the numbers as primitive is to give up on explanations of progression and counting, and rival views at least grapple with the task of trying to give us a deeper understanding than the fallback position which Skolem offers, so we will merely note his approach. The Peano Axioms meet the needs of mathematicians, and seem to be self-evident truths that illuminate the structures of arithmetic. Some take them to be the solution for our quest, but others take them to be a good expression of the puzzle.

71. Logic as the essence of arithmetic

If the Peano Axioms are felt to adequately explain arithmetic, then the essence of arithmetic is a combination of three primitive concepts (number, zero, successor), and arithmetic is the implications of the progression defined by a few axioms built from the primitives. The meaning of the progression must rest on our intuitions about the meanings of the primitives, which may well be rooted in the fact that we can count things in the world. The rival approach of logicism rests on a belief expressed by Russell, that this account is not enough: ‘Peano's premises are not the ultimate logical premises of arithmetic. Simpler premises and simpler primitive ideas are to be had by carrying our analysis on into symbolic logic’ (1907:276).

Wiggins says ‘it was a justly celebrated insight of Frege that numbers attach to the concepts under which objects fall, and not to the objects themselves’ (1980:44). We examined this idea in the context of counting, but the key idea is that the essence of number is not to be found in physical objects (Mill’s hope), and it is not to be found in mere sets of objects (Maddy’s view), and it is not to be found in mere formal progressions (Dedekind's axioms). Rather, number is a second-order concept, concerning classes of classes of things (with 3 being associated with the class of all trios of objects). Frege's own view, in the early part of his career, was that number is a property of this second-order class. Thus he wrote that 'a statement of number contains a predication about a concept' (1884:§46; also 1894:329). This is illustrated by saying that 'the number of moons of Jupiter is four', where the predication 'is four' attaches not to actual moons, and not to a 'set' of the moons, but to the concept which produces that set. That particular concept is very stable, but we saw that other concepts (such as 'persons on this bus') might need time-indexing, or more extensive description.

In this account, the number is essentially a property of a concept, but the difficulties begin when we ask *why* some concept has that numerical property, and see that it can only be in virtue of the extension of the concept, which consists of the entities which are conceptually picked out. In his later work Frege tried to bring precision to the idea of extensions of concepts (in his Basic Law V), but famously encountered paradox, and gave up. The culmination of Frege's account was that numbers emerge as abstract 'objects', required for the endless supply of objects needed for accounts of infinity, and this committed him to a platonist view of numbers. The objects are extensions under the control of a second-order concept. These numbers are cardinals, from which ordinals can be derived, and the principles of arithmetic can be specified. Because of the difficulties with rogue extensions, few people think that Frege's original theory is correct, but Frege's project certainly fits the parameters of his essentialist remarks in 1914. He tried to identify a single master conceptualisation of number, and explore ways in which arithmetic, from pure infinities to physical counting, could be explained by the resulting system. Yourgrau, for example, says that the Fregean account (as developed in Maddy 1981, based on proper classes, rather than on sets) addresses the pure set theoretical accounts, but also 'explains' why those accounts work. For example, Von Neumann says there are three items if the items match his set-theoretically defined three, but (Yourgrau suggests) only the Fregean account offers a 'principle of collection' that ties together all the threes (1985:356).

Hodes offers a common criticism of Frege, that his account of arithmetic rests on austere logic, but culminates in a rich ontology of objects. Later, though, Hodes says that the inconsistency in Frege's later work is 'a minor flaw', because 'its fundamental flaw was its inability to account for the way in which the senses of the number terms are determined' (1984:139). Effectively, Frege has given a formal framework which can refer to the numbers, but he hasn't told us what numbers actually mean – and that was the essence he was trying to pin down. Wright addresses this difficulty, and finds a strategy in avoiding the extensions of concepts, and concentrating on the essence of cardinal numbers as the lynchpin of logicism.

Wright tells us that the three reasons why the original logicist project foundered were Russell's paradox, the non-logical character of the axioms of Russell and Whitehead, and Gödel's two incompleteness theorems (1983:xxi). Russell's paradox blocked the hope that every concept would generate a set, the axiom problem was that the logic was 'impure' (because it invoked objects), and Gödel seemed to show that there were aspects of number which exceeded the grasp of logic. Wright therefore proposed to focus on the uncontroversial 'Hume's Principle' –

that two collections have identical cardinality if there is one-to-one correspondence between their members. The derivation of the Peano Axioms from this starting point is now known as Frege's Theorem. The principle is central to our interest in this case, because Heck tells us that 'the interest of Frege's Theorem is that it offers us an explanation of the fact that the numbers satisfy the Dedekind-Peano axioms' (2000:204). Hume's Principle, Heck tells us, 'is supposed to be more fundamental, in some sense, than the Dedekind-Peano axioms' (2000:189).

For logicists the Peano axioms are the puzzle to be explained, rather than being the explanatory essence. Wright presents them in this way in his introduction when he says that the Peano Axioms seem to be 'truths of some sort', and so there 'has to be a philosophical question how we ought to conceive the nature of the facts that make those statements true' (1983:xiv). Wright presents his project in explicitly essentialist language when he says that for Fregeans like himself 'number theory is a science, aimed at those truths furnished by the essential properties of zero and its successors' (1983:xiii). This treats essences as pertaining to 'objects', in Steiner's manner, rather than as the natural deduction rules which we surmised were the essence of logic. The logicism will be found in the definitions that pinpoint the members of the progression, since the definitions are taken as analytic, and hence purely logical in character.

Fine summarises Wright's subsequent procedure thus: 'the Fregean arithmetic can be broken down into two steps: first, Hume's Law may be derived from Law V; and then, arithmetic may be derived from Hume's Law without any help from Law V' (2002:41). The controversial Basic Law V (which led to paradox) thus drops out, and arithmetic rests on the intuitively appealing definition of equinumerosity in terms of one-to-one matching between two collections. Heck argues that we can directly discern equinumerosity without either matching or counting (2000), but Wright's principle is still appealing. Wright's summary of what he is proposing is expressed in the language of our present discussion when he writes that 'the Peano Axioms are logical consequences of a statement constituting the core of an explanation of the notion of cardinal number. The infinity of cardinal numbers emerges as a consequence of the way cardinal number is explained' (1983:168). Thus we find both Frege and Wright trying to locate the essence of number at a more fundamental level than the structure produced by the Peano Axioms, and presenting what they offer as both a 'kernel' and an 'explanation' of arithmetic. There are plenty of criticisms offered of both old and new logicism, but the objective of the quest is clear enough.

72. Set theory as the essence of arithmetic

We have sampled the ideas that the essence of number is found in axioms, or in concepts, or in a definition, and we can conclude with the possibility that numbers are sets. Early set theory fell into some disrepute because of paradoxes, but theorist battled through to a fairly standard account (ZFC, and the iterative conception). The big step towards the modern view was (as Lavine puts it) that in 1923 Von Neumann 'had shown how to introduce ordinal numbers as sets, making it possible to use them without leaving the domain of sets' (1994:122). Zermelo had already championed the central role of set theory in mathematic, and Maddy, a modern scion of this tradition, summarises the situation thus: 'Zermelo was a reductionist, and believed that theorems purportedly about numbers (cardinal or ordinal) are really about sets, and since Von Neumann's definitions of ordinals and cardinals as sets, this has become common doctrine' (1988:489). Standard set theory passed a key test by proving the Peano Axioms, and nowadays not only does set theory encompass arithmetic, but 'it is well known that virtually every field of mathematics can be reduced to, or modelled in, set theory' (Shapiro 1997:5).

We cannot simply accept set theory as the essence of mathematics. We have already seen that if we seek the essence of set theory itself, there is no clear answer, because there is some arbitrariness in the axiomatisation, and it is not clear that anything we call 'set theory' actually exists apart from the axioms that generate the system. In addition, the more powerful second-order set theory needed for mathematics is (unlike the Peano Axioms) only 'semi-categorical'. That is, there is no situation in which there exists a full set of models which are all precisely isomorphic to one another, because the system endlessly expands (Mayberry 1994:413). Yourgrau's conclusion is that 'sets could hardly serve as a foundation for number theory if we had to await detailed results in the upper reaches of the edifice before we could make our first move' (1985:356).

Despite these difficulties, Maddy supports the view that 'numbers simply are certain sets', and her defence (in her early work) rests on the view that 'this has the advantage of ontological economy, and allows numbers to be brought within the epistemology of sets' (1981:347). Mayberry presents the strongest possible account of this approach when he writes that 'one does not have to translate 'ordinary' mathematics into the Zermelo-Fraenkel system: ordinary mathematics comes embodied in that system' (1994:415). On this view, then, the essence of a number will be a set, though set theoreticians are less inclined to use essentialist language than the other theoreticians we have mentioned, perhaps because sets have the sort of self-evidence which dissolves the explanatory puzzle, rather than solving it.

If we are dropping Frege's second-order concepts, we might object that the simple set of wheels on my tricycle and the set of Graces may be trios, but they are different sets. Hence the recourse of the set theoretician is to define three in 'pure' sets, leaving out the elements. This, however, leads to a notorious difficulty. Zermelo found that the number 3 could be captured by the pure set {{{Ø}}}, meaning that 3 is the singleton of the singleton of the singleton of the null set, so that it stands on its own, and does not contain 1 or 2. Then Von Neumann showed that 3 could be captured by {Ø,{Ø},{Ø,{Ø}}}, in which 3 is a set which contains 0, 1 and 2 (Shapiro 2000:265). Both versions offer an analysis of the concept of 'successor', which we previously met as a primitive. However, if we wish to know the essence of 3, Von Neumann tells us that 1 *is* a member of 3, and Zermelo tells us 1 is *not* a member of 3 (a question discussed by Aristotle in *Met* M.7).

The difficulty arose when Benacerraf then observed that both of these sets capture 3, but obviously 3 can't be both of them, and so is probably neither, and hence that identifying mathematics with set theory is a confusion. His conclusion was that 'the fact that Zermelo and Von Neumann disagree on which particular sets the numbers are is fatal to the view that each number is some particular set' (1965:279). The idea that numbers are actually to be identified with sets seems to be undermined, which leaves mathematics to be either ‘reducible’ to set theory, or ‘modelled’ in set theory. If it is merely modelled in set theory then this seems to banish any essentialist hopes of pinning down the true nature of numbers, and Brown writes that ‘maybe all of mathematics can be represented in set theory, but we should not think that mathematics *is* set theory; functions can be represented as order pairs, but perhaps that is not what functions really are’ (1999:102), reminding us of Almog’s similar doubts about the Peano Axioms. Read makes a similar point when he says that ‘the Von Neumann numbers have a structural isomorphism to the natural numbers - each number is the set of all its predecessors, so 2 is the set of 0 and 1. This helps proofs, but is unacceptable. 2 is not a set with two members, or a member of 3’ (1995:106).

It doesn’t look as if numbers *are* sets, and it doesn’t seem that modelling numbers in set theory reveals the true nature of numbers. Perhaps if numbers were reducible to sets we might feel that the essence of the former was revealed (rather as reducing lightning to electrical discharge is so revealing). This may be the best the essentialist can hope for in the set theoretic approach, despite a difficulty spotted by Benacerraf (1965:290), that one version of set theory can be reduced to the ordinal numbers (rather than the other way around), and Hossack speculates that ‘we might reduce sets to ordinal numbers, thereby reversing the standard set-theoretical reduction of ordinals to sets’ (2000:436). If the sets reduce to the ordinals, rather than ordinals reducing to sets, this would return the priority to the Peano Axioms. The problem points to a further question: if the two modes are mutually reductive, does one direction of reduction give greater understanding of the situation? Do we best understand ordinary arithmetic when we think of it in terms of ordinal numbers, or in terms of sets? The evidence leans a little towards the former, but we can leave the question open.

In defence of set theory, Maddy gives a nice example of the sort of explanatory fruit we are hoping for. The commutativity of multiplication (that n.m = m.n) can be proved from the Peano Postulates, but the proof (she says) offers no explanation of the phenomenon; set theory, on the other hand, shows exactly why the commutativity occurs. Roughly, a grid of items of sides n and m contains the same total whether you view it upright, or turned through 90°, which is a set theoretic way of viewing it (in terms of Cartesian products) (1981:347). This seems to invoke a perceived pattern, but Maddy would doubtless argue that patterns are visualised set theory. Such an example certainly weakens the claim that the Peano Axioms will do the full explanatory job we expect of an essence, and strengthens the case for set theory. Among the supporters of set theory, Maddy comes the closest to the sort of essentialism about systems we have been exploring when she writes that ‘our set-theoretic methods track the underlying contours of mathematical depth. ...What sets are, most fundamentally, is markers for these contours ...they are maximally effective trackers of certain trains of mathematical fruitfulness’ (2011:82). While explorers of this aspect of foundational mathematics make little reference to essentialism, our discussion shows that each of the speculative proposals that has been investigated fits the Aristotelian pattern of explanatory essentialism we have been discussing. Perhaps a more explicit account of this general quest would help to focus its objectives more clearly.

Case 3: The Periodic Table

73. The Nature of gold

We have looked at the role of explanation and essence in the cases of counting and the unity of ordinary objects, and in the case of the way we understand abstract systems of thought. To gain a different perspective, we will now look at a case from the physical sciences, since these are taken to be the paradigm contexts for our best explanations, and science is generally thought to be the arena in which the essentialist attitude must either thrive or be abandoned. Prior to the writings of Putnam and Kripke, any modern suggestion that science should be understood in essentialist terms would have been met with incredulity (except from Copi, who had suggested it in 1954). We have seen the reason for this view in the predominance of the Humean view of the laws of nature as regularities, and the covering-law view of explanation that accompanied it. The new shift in attitude resulted from the simple and bold proposal in the theory of reference that a singular reference (and perhaps a reference to a natural kind) is not achieved by means of a cluster of descriptions (probably implying a cluster of intersecting regularities), but by actually invoking the thing itself. Some initial process of picking out occurs, of the single thing, or of a paradigm instance, with the reference thereafter maintained by a linguistic community, perhaps guided by expert knowledge. While clusters of descriptions may shift and change, the thing itself does not, and so we find ourselves referring to an agreed concept of the underlying fixed nature of the referent, and this begins to look like a traditional essence. Since this underlying nature is usually discovered by experts rather than by ordinary speakers, the doctrine of ‘scientific essentialism’ emerges (of which Ellis 2002 gives a nice survey). However, the background to this move is found not in the philosophy of science, but in the semantics of modal logic, and this makes it difficult to disentangle a number of strands from the new picture. Kripke introduced the concept of ‘rigid designation’, so that when we discuss what President Nixon might have done, or how gold might behave in other circumstances, or what heat might be in some counterfactual scenario, we can retain the idea that the referent remains unchanged from whatever was initially picked out (1980). If we say ‘Nixon might have been taller’ we are not referring to a taller Nixon, but are referring to the shorter Nixon and discussing his possibilities. Hence if we are discussing the possibilities for gold or heat, there is (if the theory works as well for natural kinds as it does for Nixon) no question of these being anything other than the items with which we are familiar. Since experts have pronounced, with considerable confidence, that gold has atomic number 79, this means all talk of the possibilities for gold will also refer to a substance with atomic number 79. Thus, on this view, ‘gold has atomic number 79’ is deemed a necessary truth, since it will be true in all the possible worlds in which there might be gold. It is acknowledged that (though unlikely) the experts may have got the atomic number wrong, but the point is that the ‘nature’ of gold, whatever that may, could never change. If we contemplate some possible substance which lacks this agreed nature, then we are not contemplating gold.

74. Scientific essentialism

Some optimists hoped that a metaphysical revolution could be directly effected by this revolution in semantics, but the sustained criticism of Nathan Salmon (1980/2005) has turned most philosophers away from such a simple and unlikely view. If we ask whether gold or heat or President Nixon could be somewhat different from the instances of them within our common acquaintance, and yet still qualify for those labels we currently use, it does not seem that mere linguistic usage can settle the matter. As we saw earlier, rigid designation is, in fact, a 'stipulation' (Kripke 1980:44), which will impose a conceptual necessity on our subsequent uses of a word, even though we are free to countermand one stipulation with a quite different (and less rigid) one. As Mumford observes 'an electron would not be an electron if its behaviour were different from the behaviour it has in the actual world, but this necessity is purely conceptual' (1998:237). The semantics may only offer a conceptual necessity to natural kind terms, but Kripke and Putnam had still effected a revolution, by shifting the attention of philosophers of science away from 'laws of nature', and towards the 'natures' of natural kinds. The huge shift in our metaphysics of nature that resulted is found in the simple thought that rather than things obeying laws, maybe laws obey things. That is, that laws of nature are not prior and independent authorities, with an ontology derived from either theology or total obscurity, but are actually the consequences of the natures of the kinds of entities that constitute the world. The philosophically exciting possibility of identifying natural necessity with metaphysical necessity is still on offer, provided we are willing to say that some substance would cease to be that substance if it in any way lost the 'essential nature' that we have assigned to it. In such a case, then any world made up of the substances that constitute our actual world will therefore have the same natures with which we are familiar, and since it is those natures which generate the 'laws' of our world, then any such world will necessarily have the same laws that we experience. When, for example, Lowe writes that 'it is not metaphysically necessary that water is composed of H_2O molecules, because the natural laws governing the chemical behaviour of hydrogen and oxygen atoms could have been significantly different, so they might not have composed that substance' (2013:6), he reveals that (though he supports Fine's definitional view of essences) he is not a scientific essentialist. In the latter view, the laws for the combination of hydrogen and oxygen necessarily arise from the natures of those two elements, and could not have been different. There is, of course, room to demur from this view, if there is thought to be more to the laws of nature than the mere expression of the essential natures of natural kinds, but the supporters of scientific essentialism would say that the onus of proof lies with the champions of such poorly supported laws, and not with the more economical essentialist view.

75. Mendeleev

Against this scientific essentialist background, we will now look at the most important development in the history of chemistry – the acceptance of the periodic table of elements. For this we will rely particularly on Scerri's excellent guide to the subject (2007). Our interest is in whether this development was explanatory in purpose, whether something like the essences of the various elements emerged from the project, and whether those identifications of essence were the consequence of the explanatory motivation.

Mendeleev has achieved fame as the man who discovered the periodic table, but the history is complex, and his achievement certainly seems to be one of those developments in the history of science that had a certain inevitability, so that if Mendeleev had not produced his account then someone else would soon have done it. Mendeleev still deserves great honour, of course, but the inevitability simply reflects the situation that the facts about the elements were steadily emerging, becoming obvious to anyone who studied the new research. According to Scerri, Mendeleev got there first because he made certain assumptions about the object of enquiry which other scientists did not make. A key question facing researchers of the time was why two elements, sodium and chlorine, which in isolation are poisonous to human beings, can combine to make ordinary beneficial table salt, or why the metal mercury and the gas oxygen can combine to make mercury oxide. This demanded an explanation, which would show what was lost from the elements and what was retained. Mendeleev's solution was to distinguish between the material element and the 'substance' of the element. Thus Mendeleev wrote in 1868 that 'neither mercury as a metal nor oxygen as a gas is contained in mercury oxide; it only contains the substance of the elements, just as steam only contains the substance of ice, but not ice itself' (Scerri 2007:115). Scerri's interpretation of this remark is that 'for Mendeleev, the element was an entity, which was essentially unobservable but formed the inner essence of simple bodies; whereas a particular 'element' was to be regarded as unchanging, its corresponding simple body aspect could take many forms' (115). Given this rather metaphysical view of the physical problem, he then examined the evidence for any symptom of the hidden 'substance' that was sought, and (says Scerri) Mendeleev's 'genius' lay in spotting that unchanging atomic weight was that symptom. Other researchers had taken an interest in atomic weight, but it was this essentialist picture of things which motivated Mendeleev to pursue the matter, and soon arrive at his famous breakthrough in laying out the atomic table. In later writings Mendeleev referred to 'matter, force and spirit' in physical substances, and Scerri endorse the view that the reference to 'spirit' amounts to 'the modern notion of essentialism' (2007:118).

Another interesting aspect of Scerri's account is that while Mendeleev is often accorded his fame because of a few spectacular predictions made and then confirmed by this theory, the reality of his predictions is that 'at best, only half of them proved to be correct', with his failures being quietly forgotten (123). When Mendeleev was awarded the Royal Society's Davy Medal in 1882, the citation made no reference to predictions and only praised his accommodations to the troublesome data – which confirms the view of explanation developed above (146).

76. Explaining the periodic table

We seem to have good grounds for thinking that Mendeleev's project was an explanatory one, and that essentialist thinking not only motivated his task, but also gave him an edge over his rivals, because it focused his efforts. Of course, what appears to be essentialist thinking on Mendeleev's part may just be good luck resulting from a fruitful delusion, so a more significant question for our enquiry is whether the periodic table has revealed to us anything which can justifiably be called the ‘essence’ of each element. To answer that, the history of the affair must be pursued further, to see the later adjustments that had to be made. When the ‘noble gases’ such as argon were discovered in the 1890s, the periodic table was so secure that the new elements could be slotted onto the end to form a new ‘group’ (or vertical column). The pressing question remained of *why* the elements naturally fell into this tabular form, characterised by features such as that group one (lithium, sodium, potassium…) showed increasingly dramatic reactions with water as they descended the ‘periods’ (or rows), whereas at the other end, in the new group 18 of noble gases, there was no reactivity at all to be discerned. It took the unravelling of the structure of the atom to produce explanations of that puzzle. For example, it turns out that the possession of eight electrons in the outer shell of an atom is the most stable configuration, and this is found in the noble gases, explaining their reluctance to interact with other elements.

The status of the table met its next crisis in the 1910s when it was realised that a number of subtly different atoms seemed to qualify for the same place in the periodic table, and researchers were confronted with ‘isotopes’ of the same element. We have already seen the difficulty this creates for the application of the concept of ‘natural kind’ to the elements. For the periodic table it raised the question of whether scientists should switch to an entirely new isotopic table, abandoning the great discovery of Mendeleev. The dilemma was resolved by a parallel and very revealing development at around the same time. By the use of X-rays, Moseley showed that the steps of the table proceeded in exact whole numbers, and that these corresponded to the positive charge (the protons) of each type of atom. Not only was a precise indication of the gaps in the table finally agreed, but a new and powerful guiding principle for the whole system had emerged. Since this meant that since all elements had an ingredient which was an exact multiple of one, they could be understood as ‘composites of hydrogen’, and Scerri’s comment is that ‘this revitalized some philosophical notions of the unity of all matter, criticised by Mendeleev and others’ (2007:175). A consequence of this discovery is that ‘the elements are now believed to have literally evolved from hydrogen by various mechanisms’ (2007:250). A conference in 1923 decided to base the periodic table on the ‘charge’ of the atom (its number of positively charged protons), rather than atomic weight, and ignore the awkward differences between isotopes (which reflected variety in the number of neutrons in the nucleus), and thus the modern system for the periodic table was largely settled. This is because isotopes rarely make much difference to the chemical behaviour of an element, but a striking exception is the lightest element (hydrogen), where the neutron makes a much greater proportional difference to the weight and behaviour. The philosopher must never forget that a certain amount of fudging was required to arrive at the neat modern picture.

Two interesting questions remain after that simplified account of the development of the periodic table: has a full explanation of the nature and behaviour of the elements resulted, and (if so) what is that explanation? There is no consensus on the first question. On the one hand it has been said that the phenomenon of radioactivity meant that all explanations of chemistry must be traced back to the underlying physics, and Niels Bohr and others have attempted to explain the main features of the periodic table in terms of quantum mechanics. On the other hand, Scerri and others observe that there are several features of the table for which no quantum mechanical explanation is even remotely available. This includes a reason why the period length seems to be eighteen, a reason 'why a given collection of atoms will adopt one molecular structure (and set of chemical properties) or the other' (Weisberg/Needham/Hendry 2011:35), and a reason why the electron shells are filled in a particular order (Scerri 2007:229). For some philosophers of chemistry this is enough to affirm that chemistry can never be finally reduced to physics, since the detailed revelations of quantum physics have not answered these questions. On the whole it seems that explanation can terminate at the lowest level of chemistry, without spilling into physics, which conforms to the picture we have been developing.

77. Essence of an element

This leaves the most important question: given that the essence of an element is what explains that element, can we identify such an essence? We have begun to develop some criteria which are applicable to the question. Do explanations converge on some narrow feature of the atoms of an element, and is there some feature which generates comprehensive explanations of its behaviour? Is there some aspect which unites an element within a single concept, determining its boundaries and overlap criteria? Is there some revelation which allows us to say that we now fully understand the nature of this element? Is there a model of the atom which brings to light a clear grasp of its mechanisms?

Portides refers to a number of models of the atomic nucleus which have been proposed, and concludes that 'the unified model can be considered a better representation of the atomic nucleus in comparison to the liquid-drop and shell models, because it explains most of the known results about the nucleus' (2008:391). However, this leave unsettled the question of whether it is the nucleus or the electron shells which should most concern us. A nice example of what we would like to explain is the colour of gold. This is exactly the sort of thing which Locke cites as to be explained by a real essence, though he despaired of such an achievement (*Essay* 2.31.6*)*, and Kenelm Digby in 1644 ventured the corpuscular speculation that 'the origin of all colours in bodies is plainly deduced out of the various degrees of rarity and density, variously mixed and compounded' (Pasnau 2011:506). Scerri is now able to tell us that 'the characteristic color of gold ....can best be explained by relativistic effects due to fast-moving inner-shell electrons' (2007:24). We may have further questions, but this suggests that the essence of gold (in the explanatory context we have been discussing) may be found in the electron shells. If so, it would seem that gold is classified according to its units of charge, but that its essence is elsewhere, suggesting that we should not place too much emphasis on classification when we are in quest of the essence. However, while the sequence of elements is decided by the protons, 'the modern notion is that atoms fall into the same group of the periodic table if they possess the same numbers of outer-shell electrons' (Scerri 2007:192), so that the structure of the table rests on both the protons and the electron shells. So do we settle for the combination of protons and electron shells as the issue, or is there some explanatory priority between them? If we settle that by a dependence relation, then the matter is decided in favour of the protons; Hendry, for example, writes that 'nuclear charge determines and explains electronic structure and spectroscopic behaviour, but not vice versa' (2008:523) - but is the nucleus too remote from the action we wish to explain (perhaps implying a move to a different ‘level’)?

The outstanding features of the basics of the chemical world which seem to require explanation are the distinctive spectrum which marks out each element, and the richness of the relations that elements enter into. There are, for example, 92 naturally occurring elements, but over 100,000,000 compounds have been discovered or synthesised (Weisberg /Needham/Hendry 2011:25). If we think that the explanation resides in molecular structure, we find that CO_2 and SO_2 differ despite having the same structure, and if we think that only the atoms matter we find that ethanol and dimethyl ether contain exactly the same atoms (structured differently), but one boils at 78.4° and the other at -24.9° (Hendry 2008:523). Hendry takes the firm view that ‘in general, nuclear charge is the overwhelming determinant of an element's chemical behaviour’ (2008:522), but the matter may be too subtle for the sort of simple answers philosophers prefer. The key to many of the explanations is to be found in the chemical bond, which was formerly thought to involve a transfer of electrons, but is now understood in terms of shared pairs of electrons (Scerri 2007:207); clearly this is more concerned with the shell than with the nucleus.

Actually these details are of little importance for the present thesis. An exploration of them makes the necessary point, that once we begin to focus on the core explanations of our area of study, we find ourselves led to a narrow group of features, which have the potential to meet anyone’s normal concept of the ‘essence’ of the matter. The periodic table itself is not the explanation we seek. The news that gold has been placed in the eleventh group of the sixth period of a table elements reveals nothing about the sources of gold’s behaviour, and the information would have intrigued Locke and Leibniz, but it would hardly have satisfied them. It is only when the causal mechanisms are described and modelled that we feel we are very close to our explanatory goal. However, the development of the table is what has revealed these explanatory structures to us most clearly, and we should not underestimate the revelatory nature of such relations, just as we noted the illumination that can come from pinpointing a precise regularity.

78. Weak and strong conclusions

These short case studies seem to show that essentialist explanation has an important role in the metaphysics of ordinary objects, in our grasp of formal systems, and in a major theory of the physical sciences. The findings of our enquiry seem to at least conclude with a plausible weaker claim, to which an accumulation of evidence has given good support. There is also a more speculative stronger claim, which is suggested by the current approach, and merits further investigation.

The weaker claim can be seen in its response to the view put forward in 1949 by Weyl, a mathematician and physicist. We have met the question concerning arithmetic of whether we can grasp the nature of numbers, or must settle for a set of isomorphic maps. Weyl wrote that ‘a science can determine its domain of investigation up to an isomorphic mapping. It remains quite indifferent as to the 'essence' of its objects. The idea of isomorphism demarcates the self-evident boundary of cognition’ (quoted in Shapiro 2000:160). Weyl’s thought encapsulates an older view of the aspirations of science which many are now rejecting. The admirable empirical restraints which science places upon itself produced a picture of theories as mathematical pattern construction, guided by the logical boundaries revealed in model theory. Such models are constructed linguistically, from sets of sentences, but there is a richer concept of a model which involves representations, causal powers, and closer ties to the physical world.

In the discussion of induction, we saw that curiosity can drive us beyond the strictures of empiricism, and nothing can prevent us from asking *why?*, even after an isomorphic mapping of theories has been achieved. The simplest curiosity wonders what mechanism exists which produced the map. The offer of explanatory 'laws of nature', with no ontological status accorded to those laws, can look like a mere placeholder for ignorance, as much of an I-know-not-what as essences seemed to Locke. For Weyl the word 'essence' needed distancing quotation marks around it, but it is hard to resist the desire to understand the fundamental natures of the more basic natural kinds, such as particles and elements, and even biological cells and tigers, and the belief that a grasp of those natures is taking our understanding of nature to a deeper level. Debates can still rage at that level, concerning the status and role of laws, or concerning the rival claims of dispositions and categorical properties, but grasping the natures of things should be on the agenda of science. These natures constitute the focal point of each of the successful explanatory structures which is generated in modern science. The quest by theorists to identify the essence of mathematics is still worth pursuing. That is the weaker claim, with the addition that 'essence' is much the best word for the target of such enquiries, built on the increasingly widespread belief that the aim of the sciences is explanatory rather than descriptive.

The stronger claim gives explanation a much more central role, not only in the modes of justification on which we build our beliefs, but also in the shaping of the metaphysical schemas by which we all live. Harman has argued that inference to the best explanation is the key concept in epistemology, being the mechanism which produces all of our beliefs (1974). The present discussion is sympathetic to such a proposal, but has not argued for it. What has been argued for is that a concept (essence) which occupies a very prominent position in the history of our metaphysics, rather than of our epistemology, has its actual source not in an attempt to see reality from the point of view of eternity, but to see reality through the very human focus of our attempts to explain it. If it is right that such a powerful concept employed in many accounts of the metaphysics of nature has originated in this way, might that account fit other aspects of our metaphysics? The stronger claim, then, is that the account of essence given here is not an aberration, but is typical of the way we think. That is (to give the thought its industrial strength expression) that we will only understand our own metaphysical schemes clearly when we see that they have arisen from the very basic human drive to explain the world. If that were correct, it would not only open a way to understand our own thinking more clearly, but also a way of subjecting it to constructive critiques, by demanding that our metaphysics do the job that is required of it.

The stronger claim has much in common with Kant's critique of pure reason, but it differs in its stronger commitment to a realist basis. When Kant writes that 'space is a necessary representation, a priori' (1781:A24), or that 'it is only because we subject the sequence of appearances and thus all alteration to the law of causality that experience itself is possible' (1781:B234), he will find little agreement among current thinkers. The conception of reality that accompanies Kant's views is so weak that his thought can be interpreted as anti-realism or idealism. However, it would set philosophy back several hundred years if we ignored Kant's plea that we attend carefully to the role of our own minds in the general picture of reality which guides us. If we adopt a more robust view of the reality of the external world, including its space, time and causation, but acknowledge (with Aristotle) the necessary influence of our own cognitive interests in what happens next, then it seems an obvious possibility that our metaphysics is dominated and shaped by explanatory concerns.

Bibliography

Almog, Joseph (2010) 'Nature Without Essence'. *Journal of Philosophy* 107/7

Aristotle (c.330 BCE) *The Basic Works.* ed. McKeown. Modern Library Classics 2001

_______ (c.330 BCE) *Categories*. trans. J.R. Ackrill. OUP 1963

_______..(c.330 BCE) *Posterior Analytics*. text, and trans. by Hugh Tredennick. Loeb Library 1960

_______..(c.330 BCE) *Posterior Analytics*. trans. Jonathan Barnes. OUP 1975

_______ (c.330 BCE) *Physics.* trans. Robin Waterfield. OUP 1996

_______ (c.330 BCE) *Topics*. Text, and trans. by E.S. Forster. Loeb Library 1960.

_______ (c.330 BCE) *Metaphysics 1-14.* trans. Hugh Lawson-Tancred. Penguin 1998

_______ (c.330 BCE) *Metaphysics 1-9, 10-14.* text, and trans. by Hugh Tredennick. Loeb Library 1935

_______ (c.330 BCE) *Metaphysics 7-10.* trans. and annotated by David Bostock. OUP 1994

Armstrong, D.M. (1983) *What is a Law of Nature?* CUP

Audi, Paul (2012) 'A clarification and defence of the notion of grounding', in Correia and Schnieder (eds)

Ayers, M.R. (1974) 'Individuals without Sortals'. *Canadian Journal of Philosophy* Vol.4 No.1

Bacon, Francis (1605) 'The Advancement of Learning', in *Advancement of Learning and New Atlantis.* OUP 1966

Beere, Jonathan (2009) *Doing and Being*. OUP

Belnap, Nuel D. (1962) 'Tonk, Plonk and Plink', in *Philosophical Logic* ed. P.F. Strawson. OUP 1967

Benacerraf, Paul (1965) 'What Numbers Could Not Be', in Benacerraf and Putnam (eds)

Benacerraf, P. and Putnam, H. (eds) (1983) *Philosophy of Mathematics: selected readings (2nd edn)*. CUP

Bird, Alexander (2007) *Nature's Metaphysics.* OUP

Blumenfeld, David (1982) 'Superessentialism, Counterparts, and Freedom', in Hooker (ed)

Bonjour, Laurence (1998) *In Defence of Pure Reason.* CUP

Boolos, George (1998) 'Must We Believe in Set Theory?', in *Logic, Logic and Logic* (ed. Jeffrey). Harvard

Boyle, Robert (1666) 'The Origin of Forms and Qualities', in *Selected Philosophical Papers* ed. Stewart, M.A.

Brody, Baruch (1980) *Identity and Essence*. Princeton

Brown, James Robert (1999) *Philosophy of Mathematics*. Routledge

Burge, Tyler (1998) 'Frege on Knowing the Foundations', in Burge (2005)

_______ (2000) 'Frege on Apriority', in Burge (2005)

_______ (2005) *Truth Thought Reason*. OUP

Burgess, John P. and Rosen, Gideon (1997) *A Subject with no Object*. OUP

Cartwright, Nancy (1983) *How the Laws of Physics Lie.* OUP

Cartwright, Richard (1962) 'Propositions', in *Philosophical Essays*. MIT 1987

_______ (1968) 'Some Remarks on Essentialism', in *Philosophical Essays*. MIT 1987

Chakravarty, Anjan (2012) 'Inessential Aristotle: Powers without Essences',
ttp://www3.nd.edu/~achakra1/research.html

Copi, Irving (1954) 'Essence and Accident', in *Journal of Philosophy* vol.51 no.23

Correia,F and Schnieder,B (eds.) (2012) *Metaphysical Grounding*. OUP

Cover, J.A. and O'Leary-Hawthorne, J. (1999) *Substance and Individuation in Leibniz*. CUP

Curry, Haskell B. (1954) 'Remarks on the Definition and Nature of Mathematics', in Benacerraf and Putnam (eds)

Dedekind, Richard (1872) 'Continuity and Rational Numbers', in *Essays on Theory of Numbers*. Dover 1963

_______ (1888) 'The Nature and Meaning of Numbers', in *Essays on Theory of Numbers*. Dover 1963

Della Rocca, Michael (2002) 'Essentialism versus Essentialism', in the *Journal of Philosophy* vol.93 no.4

Dennett, Daniel (1995) *Darwin's Dangerous Idea*. Penguin 1996

Descartes, René (1646) *Philosophical Essays and Correspondence,* ed. Ariew,R. Hackett 2000

Devitt, Michael (2008) 'Resurrecting Biological Essentialism', in *Putting Metaphysics First.* OUP 2010

Drewery, Alice [ed.] (2005) *Metaphysics in Science*. *Ratio* Vol.XVIII/4

Dummett, Michael (1973) *Frege: philosophy of language (2nd edn).* Duckworth 1981

_______ (1991) *Frege: philosophy of mathematics*. Duckworth

Dupré, John (1993) *The Disorder of Things*. Harvard

Ellen, Roy (1996) 'The Cognitive Geometry of Nature', in *The Categorical Impulse*. Berghahn Books 2006

Ellis, Brian (2001) *Scientific Essentialism*. CUP

_______ (2002) *The Philosophy of Nature: new essentialism.* Acumen

_______ (2005) 'Katzav on limitations of dispositions', in *Analysis* Vol 65 No 1

Fine, Kit (1994) 'Essence and Modality', in *Philosophical Perspectives* 1994

_______ (1998) 'Cantorian Abstraction', in *Journal of Philosophy* 95 No 12

_______ (2002) 'The Varieties of Necessity', in Fine 2005

_______ (2005) *Modality and Tense.* OUP

_______ (2012) 'Guide to Ground', in Correia and Schnieder (eds)

Fodor, Jerry A. (1998) *Concepts: where cognitive science went wrong*. OUP

_______ (2008) *LOT2*. OUP

Forbes, Graeme (1985) *The Metaphysics of Modality*. OUP

Frede, Michael (1987) *Essays in Ancient Philosophy*. Minnesota

Frege, Gottlob (1879) *Begriffsschrift ('Concept Script').* in *From Frege to Gödel* (ed. van Heijenoort). Harvard 1967

_______ (1884) *The Foundations of Arithmetic*. ed/tr. Austin,J.L. Blackwell 1980

_______ (1894) Review of Husserl's *Philosophy of Arithmetic*. tr. Kluge, E.W. *Mind*, July 1972

_______ (1914) 'Logic in Mathematics', in *Posthumous Writings*, ed. Hermes etc. Blackwell 1979

Garber, Daniel (2009) *Leibniz: Body, Substance, Monad*. OUP

Geach, Peter (1957) *Mental Acts*. RKP

_______ (1962) *Reference and Generality*. Cornell University Press

Gelman, Susan A. (2003) *The Essential Child*. OUP

Gill, Mary Louise (1989) *Aristotle on Substance*. Princeton

Gödel, Kurt (1944) 'Russell's Mathematical Logic', in Benacerraf and Putnam (eds)

_______ (1947) ''What is Cantor's Continuum Problem?', in Benacerraf and Putnam (eds)

Goodman, Nelson (1978) *Ways of Worldmaking*. Hackett

Gupta, Anil (2008) 'Definitions'. *Stanford Online Encyclopaedia of Philosophy*

Halbach, Volker (2011) *Axiomatic Theories of Truth*. OUP

Hanna, Robert (2006) *Rationality and Logic.* MIT

Harman, Gilbert (1974) 'Inference to the Best Explanation'. *Philosophical Review* 74

_______ (1995) 'Rationality', in *Reasoning Meaning and Mind*. OUP 1999

Harré, Rom (1993) *Laws of Nature*. Duckworth

Harré, Rom and Madden, E.H. (1975) *Causal Powers: A Theory of Natural Necessity*. Blackwell

Hart, W. D. (2010) *The Evolution of Logic*. CUP

Heck, Richard G. (2002) 'Cardinality, Counting and Equinumerosity'. *Notre Dame Journal of Formal Logic* 41/ 3

Heil, John (2003) *From an Ontological Point of View.* OUP

Hendry, Robin Findlay (2008) 'Chemistry', in *Routledge Companion to Philosophy of Science* ed. Psillos & Curd

Hobbes, Thomas (1651) *Leviathan*. Ed. MacPherson. Penguin 1968

_______ (1654) 'Of Liberty and Necessity', in *British Moralists 1650-1800*, ed. Raphael. Hackett 1991

Hodes, Harold (1984) 'Logicism and the Ontological Commitments of Arithmetic'. *Journal of Philosophy* 81 No 3

Hossack, Keith (2000) 'Plurals and Complexes', in *Journal of British Society for the Philosophy of Science* Vol 51

Hughes, R. I. G. (1993) *A Philosophical Companion to First-Order Logic*. Hackett

Hume, David (1739) *Treatise of Human Nature.* ed/tr. Selby-Bigge,L./Nidditch P. OUP 1978

Jacquette, Dale (ed.) (2002) *Philosophy of Mathematics: an anthology*. Blackwell

Jenkins, C. S. (2008) *Grounding Concepts*. OUP

Jeshion, Robin (2001) 'Frege's Notion of Self-Evidence', in *Mind* 110/440

Jubien, Michael (2009) *Possibility*. OUP

Kant, Immanuel (1781) *Critique of Pure Reason.* tr. Guyer, P. and Wood, A.. CUP

Koellner, Peter (2006) 'On the Question of Absolute Undecidability'. *Philosophia Mathematica* 14.2

Koslicki, Kathrin (1997) 'Isolation and Non-Arbitrary Division'. *Synthese* 112 no.3

_______ (2008) *The Structure of Objects*. OUP

_______ (2012) 'Varieties of Ontological Dependence', in Correia and Schnieder (eds)

Kreisel, Georg (1958) ''Hilbert's Programme', in Benacerraf, P. and Putnam, H. (eds)

Kripke, Saul (1980, first pub. 1972) *Naming and Necessity (2nd end)*. Blackwell.

Ladyman, James and Ross, Don (2007) *Everything Must Go: metaphysics naturalized.* OUP

Lavine, Shaughan (1994) *Understanding the Infinite*. Harvard

Leibniz, G.W. (1669-1716) *Philosophical Essays*. ed. Ariew and Garber. Hackett 1989

_______ (1686) *The Leibniz-Arnauld Correspondence*. ed. Mason and Parkinson. Manchester 1967

_______ (1710) *New Essays on Human Understanding*. ed. Remnant and Bennett. CUP 1996

Lewis, C. I. (1923) 'A Pragmatic Conception of the a priori', in *Pragmatism* (ed. H.S.Thayer). Hackett 1982

Lewis, David (1986a) *On the Plurality of Worlds*. Blackwell

_______ (1986b) *Philosophical Papers Volume 2*. OUP

_______ (1986c) 'Causal Explanation', in 1986b

_______ (1991) *Parts of Classes*. Blackwell

Lipton, Peter (1991) *Inference to the Best Explanation (2nd ed)*. Routledge 2004

Locke, John (1694) *An Essay Concerning Human Understanding (2nd ed)*. ed. Nidditch. OUP 1979

Lowe, E. J. (2008) 'Two Notions of Being', in *Being: Developments in Contemporary Metaphysics,* ed. Le Poidevin. CUP 2008

_______ (2013) 'What is the Source of our Knowledge of Modal Truths?'. *Mind* 121 (484)

Lycan, William (1995) *Consciousness*. MIT

Machamer,P, Darden,L and Craver,VF (2000) 'Thinking About Mechanisms'. *Philosophy of Science* Vol 67 No 1

Mackie, Penelope (2006) *How Things Might Have Been.* OUP

Maddy, Penelope (1981) 'Sets and Numbers', in Jacquette (2002)

_______ (1988) 'Believing the Axioms I'. *Journal of Symbolic Logic* Vol 53 No 2

_______ (2011) *Defending the Axioms*. OUP

Mancuso, Paolo (2008) 'Explanation in Mathematics', *Stanford online encyclopaedia of philosophy*

Marcus, Ruth Barcan (1971) 'Essential Attribution', in *Modalities*. OUP 1993

Margolis, E. and Laurence, S. (2009) 'Concepts', *Stanford online encyclopaedia of philosophy*

Martin, C.B. (2008) *The Mind in Nature.* OUP

Maudlin, Tim (2007) *The Metaphysics within Physics*. OUP

Mayberry, John (1994) 'What is Required of a Foundation for Mathematics?', in Jacquette (ed)

Merricks, Trenton (2003) *Objects and Persons*. OUP

Mill, John Stuart (1843) *System of Logic (9th edn).* Longmans, Green, Reader and Dyer 1875

Molnar, George (2003) *Powers*. ed. S.D. Mumford. OUP

Mumford, Stephen (1998) *Dispositions.* OUP

_______ (2004) *Laws in Nature.* Routledge

Mumford, Stephen/Anjum, Rani Lill (2011) *Getting Causes from Powers.* OUP

Nietzsche, Friedrich (1885-89) *Writings from the Late Notebooks*. ed/tr. Bittner,Rüdiger. CUP 2003

Oderberg, David S. (2007) *Real Essentialism*. Routledge

Pasnau, Robert (2011) *Metaphysical Themes 1274-1671*. OUP

Peacocke (1992) *A Study of Concepts*. OUP

Peirce, Charles Sanders (1877) 'The Fixation of Belief', in *Pragmatism* (ed. H.S. Thayer). Hackett 1982

Perry, John (1970) 'The Same F', in *Metaphysics: an anthology*, ed Kim,J and Sosa,S. Blackwell 1999

Politis, Vasilis (2005) *Aristotle and the Metaphysics*. Routledge

Portides, Demetris (2008) 'Models', in *Routledge Companion to Philosophy of Science* (ed. Psillos and Curd)

Prawitz, Dag (1971) 'Gentzen's Analysis of First-Order Proofs', in Hughes (ed)

Prior, A.N. (1960) 'The Runabout Inference-Ticket', in *Philosophical Logic* ed. P.F. Strawson. OUP 1967

Putnam, Hilary (1967) 'Mathematics without Foundations' in Benacerraf and Putnam (eds)

_______ (1981) 'Why There Isn't a Ready-Made World', in *Realism and Reason (Papers Vol. 3)*. CUP 1983

Quine, Willard (1950) 'Identity, Ostension, and Hypostasis', in Quine (1961)

_______ (1953a) 'Three Grades of Modal Involvement', in *The Ways of Paradox*. Harvard 1976

_______ (1953b) 'Reference and Modality', in Quine (1961)

_______ (1960) *Word and Object.* MIT

_______ (1961) *From a Logical Point of View (2nd edn).* Harper Torchbooks

_______ (1969) 'Natural Kinds', in *Ontological Relativity and Other Essays*. Columbia

_______ (1970) *Philosophy of Logic*. Prentice-Hall

Read, Stephen (1995) *Thinking About Logic*. OUP

Recanati, François (2012) *Mental Files*. OUP

Resnik, Michael D. (1997) *Mathematics as a Science of Pattern Recognition*. OUP

Ruben, David-Hillel (1990) *Explaining Explanation*. Routledge

Rumfitt, Ian (2001) 'Concepts and Counting'. *Aristotelian Society* Vol. 102 No 1

_______ (2010) 'Logical Necessity', in *Modality* ed Hale,B and Hoffmann,A. OUP

Russell, Bertrand (1903) *The Principles of Mathematics.* Routledge 1992

_______ (1907) 'The Regressive Method of Discovering Premises of Mathematics', in *Essays in Analysis* (ed. Lackey). Braziller 1973

_______ (1919) *Introduction to Mathematical Philosophy.* George Allen and Unwin 1975

_______ (1940) *An Inquiry into Meaning and Truth.* Penguin 1962

Salmon, Nathan U. (1980; 2nd edn 2005) *Reference and Essence*. Prometheus Books

Salmon, Wesley C. (1989) *Four Decades of Scientific Explanation*. Pittsburgh

Scerri, Eric R. (2007) *The Periodic Table*. OUP

Shapiro, Stewart (1997) *Philosophy of Mathematics*. OUP

_______ (2000) *Thinking About Mathematics*. OUP

Skolem, Thoralf (1922) 'Some Remarks on Axiomatised Set Theory', in *Frege to Gödel* (ed. van Heijenoort). Harvard 1967

Smart, J.J.C. (1990) 'Explanation: opening address', in *Explanation and Its Limits* (ed. Knowles,D). CUP 1990

Smith, Peter (2007) *An Introduction to Gödel's Theorems*. CUP

Spinoza, Benedict de (1677) *The Ethics, Correspondence etc.* trans. R.H.M. Elwes. Dover 1955

Steiner, Mark (1978) 'Mathematical Explanations', in Jacquette (2002)

Stevenson, Leslie (1972) 'Relative Identity and Leibniz's Law', in *Philosophical Quarterly* Vol 2 No 87

Strevens, Michael (2008) *Depth: an Account of Scientific Explanation*. Harvard

_______ (2011) 'No Understanding without Explanation', in *Studies in History and Philosophy of Science* 44

Tait, William W. (1996) 'Frege versus Cantor and Dedekind: On the Concept of Number', in Jacquette (2002)

Thagard, Paul (2000) *Coherence in Thought and Action*. MIT

_______ (2012) 'Coherence: the Price is Right', in *Southern Journal of Philosophy* Vol 50 No 1

Unger, Peter (1979) 'There Are No Ordinary Things'. *Synthese* Vol.41 no.2

Urmson, J.O. (1990) *The Greek Philosophical Vocabulary*. Duckworth

Van Inwagen (1990) *Material Beings*. Cornell

Wedin, Michael V. (2000) *Aristotle's Theory of Substance*. OUP

Weisberg/Needham/Hendry (2011) 'Philosophy of Chemistry'. *Stanford online encyclopaedia of philosophy*

Westerhoff, Jan (2005) *Ontological Categories.* OUP

Wiggins, David (1980) *Sameness and Substance*. Blackwell

_______ (1995) ‘Metaphysics: Substance’, in *Philosophy: a guide through the subject*, ed.Grayling OUP

_______ (2001) *Sameness and Substance Renewed*. CUP

Witt, Charlotte (1989) *Substance and Essence in Aristotle*. Cornell

Wittgenstein, Ludwig (1921) *Tractatus Logico-Philosophicus*. tr.Pears and McGuinness. RKP 1961

Woodward, James (2009) ‘Explanation’. *Stanford Online Encyclopaedia of Philosophy*

Yourgrau, Palle (1985) ‘Sets, Aggregates and Numbers’, in Jacquette (ed)